Data Article

An image dataset of cut-test-classified cocoa beans

F.A. Santos* , E.S. Palmeira, G.J. Jesus

Universidade Estadual de Santa Cruz, Brazil

A R T I C L E   I N F O

Article history:
Received 29 January 2019
Received in revised form 20 March 2019
Accepted 5 April 2019
Available online 13 April 2019

Keywords:
Cocoa beans
Cut test
Chocolate

A B S T R A C T

Classifying cocoa beans, considering some specific standards, is required and regulated around the world, if commercialization is intended, or it can be very useful, in order to select the best ones for a given proposal. This article presents an image dataset of cut test classified cocoa beans, provided by experts following the Brazilian regulative norm for that process. Fourteen classes, with 100 images for each, were registered in a device designed with standard constant artificial lighting and capturing settings and will be present in this paper. Images are hosted in the Núcleo de Bio-logia Computacional from Universidade Estadual de Santa Cruz, under the “tedais” name.

© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Data

An image dataset of cocoa beans classified by means of the Cut Test, along with their respective classifications, was built under standardized conditions and with the help of specialists from the Centro de Inovação do Cacau (CIC), institution which is hosted in the Universidade Estadual de Santa Cruz (UESC). The database has 100 images for each class, one for each bean, which each bean evaluated by two different specialists and, if the two agreed with each other, then the bean proceeded to the dataset

* Corresponding author.

E-mail addresses: fante.antunes@outlook.com (F.A. Santos), espalmeira@uesc.br (E.S. Palmeira), gildsonj@gmail.com (G.J. Jesus).

https://doi.org/10.1016/j.dib.2019.103916
2352-3409/© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
composition. The classes are separated as the following standards: Compartmentalized Brown, Compartmentalized White, Compartmentalized Partially Purple, Compartmentalized Purple, Compartmentalized Slatty, Plated Brown, Plated White, Plated Partially Purple, Plated Purple, Plated Slatty, Moldered, Flattened, Brittle and Agglutinated, totaling 14 classes. Fig. 1 presents a low resolution sample of two images from different classes of the dataset.

2. Experimental design, materials, and methods

This section brings a brief explanation about the Cut Test and a description of a prototype of a custom made device used to provide the same environmental conditions for the images capturing.

2.1. Cut test

With the advances in the production of chocolate around the world, the processes related to it also changed, which now includes, by regulative norms around the world, several classification and characterization tests to be made in order for the products to be legally commercialized. One of those tests, the Cut Test, is made by experts which, among other processes, cut the beans lengthwise to infer about its interiors coloration and compartmentalizing, which are characteristics related to the chemical composition and fermentation stage of the beans, both of which composes the flavor of the chocolate produced by said seeds [1–4]. The process of the Cut Test is manually made following specific definition (see Ref. [4]).

2.2. Prototype device

Several prototype for image capture were analyzed and a specific device was developed for this dataset construction. The measures of such device were obtained through several previous attempts and its constituted by two parts: (1) the internal device base, where the bean is placed in its center in
order to be observed, and (2) the external box composed by a set three dichroic LED lamps (see the specifications on Table 1), located through the box in order to obtain a harmonized internal illumination to mimic the environment that the Cut Test if performed, and a space in the center of top of the box to couple a digital camera. External box has dimensions 19.7 cm × 24.8 cm × 3.0cm and the distance from camera lens and the internal base is 8.5cm.

The lamps position configuration is shown in Fig. 2 (two of them where placed at the exact center of box’s sides and the other one at the top in the middle of the upper left corner). That quantity of light sources, and its placements, were configured in this way to prevent shadows from forming around the bean and to highlight the internal characteristics of the seed.

2.3. Camera and configurations

A digital camera (Sony Cyber-shot DSC-H55) with a 25 mm wide-angle lens was used for all image acquiring. The camera was in Scene Mode, with Snow scene selected, as it showed the best scene for the intense light environment of the box; with 14 megapixels image size (4320 × 3240 pixels) in 4:3 proportions, flash and face detection deactivated and in a constant 2.5 times zoom to better place the beans on the image and less waste of picture space with background. The images are in (RGB) color space, JPG format, with 24 bits per pixel resolution, 8 for each color layer, totaling 256 colors per RGB layer. Also, a few seconds, until the camera focused on the bean, were waited before each photo was taken.

2.4. Acquiring process

All images used in this dataset were Cut Test classified, by a specialist from the Centro de Inovação de Cacau (CIC), from several ongoing cut tests of many beans samples. The images were taken in less than an hour, for the most cases was just a feel minutes, after the bean was cut and their respective

| Lamp technical specifications |
|-------------------------------|
| Brand                         | Iluctron                   |
| Opening Angle                 | 120°                      |
| Potency                       | 2W                        |
| Tension                       | 110V and 220V             |
| Luminous Intensity            | 200lm                      |

Source: [http://www.iluctron.com.br/](http://www.iluctron.com.br/)
classification were registered. Fig. 3 shows the environment where the beans were classified and the setup were the images were acquired.

**Acknowledgments**

This research was made in partnership with the Centro de Inovação de Cacau (CIC) and Programa de Pós-Graduação em Modelagem Computacional em Ciência e Tecnologia (PPGMC). The financial support was provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB).

**Transparency document**

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.103916.
References

[1] Felipe A. Santos, Eduardo S. Palmeira, Gildson Q. Jesus, Color, structural and textural features for the classification of a cocoa beans image dataset using artificial neural network, in: XIV Workshop de Visão Computacional, 2018, pp. 80–84. Ilhéus. ANAIS WVC 2018, 2018.

[2] George Alan Roskruge Wood, R.A. Lass, Cocoa, John Wiley & Sons, 2008.

[3] Carla M.E. Catsberg, Gerda JM. Kempen-van Dommelen, Food Handbook, Springer Science & Business Media, 2013.

[4] Ministério da Agricultura, Brasil, P.e.A.: Regulamento técnico da amêndoa de cacau. Diário Oficial da União, Instrução Normativa, Brasília, 2008, p. 38.