



A system for classifying vegetative structures on coffee branches based on videos recorded in the field by a mobile device



J. Avendano^{a,*}, P.J. Ramos^{a,b}, F.A. Prieto^a

^a Universidad Nacional de Colombia - Carrera 30 No 45-03 Bogota, Colombia 111321

^b Centro Nacional de Investigaciones de Café (Cenicafé).- Kilómetro 4 Vía antigua Chinchiná - Manizales, Colombia 170009

ARTICLE INFO

Article history:

Received 6 February 2017

Revised 30 June 2017

Accepted 30 June 2017

Available online 5 July 2017

Keywords:

Vegetative structures

Structure from motion

Yield crop

Coffee branches

Classification

ABSTRACT

As a drink, coffee is one of the most in demand products worldwide; as an agricultural product, it requires non-destructive tools for its monitoring and control. In order to create a non-destructive method which can be used in the field, a system was developed to find and classify six types of vegetative structures on coffee branches: leaves, stems, flowers, unripe fruits, semi-ripe fruits, and ripe fruits. Videos were obtained from 12 coffee branches in field conditions, using the rear camera of a mobile device. Approximately 90 frames, those which had the most information from the scene, were selected from each video. Next, a three-dimensional (3D) reconstruction was generated using the Structure from Motion (SfM) and Patch-based Multi-view Stereo (PMVS) techniques for each branch. All acquired images were manually recorded, and a Ground Truth point cloud was generated for each branch. The generated point clouds were filtered using a statistical outliers filter, in order to eliminate noise generated in the 3D reconstruction process. The points that were located in the deepest part were considered to be the scene background, and were removed using a band-pass filter. Point clouds were sub-sampled using a VoxelGrid filter, to reduce the number of points to 50% and therefore reduce computation time of the processes that followed. Various two-dimensional (2D) and 3D features were taken from the point clouds: 11 based on RGB, Lab, Luv, YCbCr, and HSV color space, four based on curvatures, and the remaining two based on shape and curvedness indexes. A Support Vector Machine (SVM) was trained with the previously mentioned features by using eight branches for the training stage, and four branches for the validation stage. Experimental results showed a precision of 0.82 and a recall of 0.79, when classifying said vegetative structures. The proposed system is economical, as only a mobile device is needed to obtain information. Remaining system processes were performed offline. Additionally, the system developed was not affected by changes in lighting conditions, when recording videos on a coffee plantation.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The development of expert systems, as applied to agriculture, is a topic of great interest for numerous areas of computer sciences nowadays. Machine vision systems for application in agriculture have allowed for the development of non-invasive technologies which carry out agricultural tasks in fields, on farms, and in greenhouses, among other places. Examples of these tasks include: classification of soils to identify the most appropriate tillage technique to be implemented for crops (Ajdadi, Gilandeh, Mollazade, Hasanzadeh, 2016), creation of agricultural robots (Gollakota & Srinivas, 2011; Jiang, Koch, Scherer, & Zell, 2013), quality inspection

and food disease detection (Dutta, Singh, & Ghosal, 2015; Nansen, Singh, Mian, Allison, & Simmons, 2016; de Oliveira, Leme, Barbosa, Rodarte, & Pereira, 2016; Øystein Sture, Øye, Skavhaug, & Mathiassen, 2016; Zhang et al., 2015), identification of the type of fruit used, determination of liquor ageing time (Rodrigues et al., 2016), the automatic estimation of the weight of farm animals (Mortensen, Lisouski, & Ahrendt, 2016), measurement of food product volume, as in the case of dehydrated apples (Sampson, Chang, Rupasinghe, & Zaman, 2014), monitoring of coffee beans in the roasting stage (Virgen-Navarro, Herrera-López, Corona-González, Arriola-Guevara, & Guatemala-Morales, 2016), and automated coffee fruit selection (Herrera, Medina, Beleno, & Gualdrón, 2016).

The principal techniques used in the previously mentioned tasks related to machine vision, and they aim to make productive agricultural processes more efficient. Nowadays, there are also challenges in the digital image processing area. For example: detection and follow-up regarding organism structural ruptures

* Corresponding author.

E-mail addresses: joavendanope@unal.edu.co (J. Avendano), paula.ramos@cafedecolombia.com (P.J. Ramos), faprieto@unal.edu.co (F.A. Prieto).

and crop cells, automatic modeling of three-dimensional plant geometry, underground plant root segmentation and monitoring, crop growth tracking, and accurate harvest prediction (Minervini, Scharr, & Tsaftaris, 2015).

Several types of research, through use of image processing, have developed artificial vision systems for the classification of vegetative structures, for example, those developed for: vegetables (Moreira, Muñoz, Ruiz-Altisent, & Perdigones, 2012), grapes (Ivorra, Sánchez, Camarasa, Diago, & Tardaguila, 2015), date fruit (Ohali, 2011), tomatoes (El-Bendary, Hariri, Hassanien, & Badr, 2015), apples (Sampson et al., 2014; Zhang et al., 2015), and even coffee (Apaza, Portugal-Zambrano, Gutiérrez-Cáceres, & Beltrán-Castañón, 2014; Condori, Humari, Portugal-Zambrano, Gutiérrez-Cáceres, & Beltrán-Castañón, 2014; Sandoval, Prieto, & Betancur, 2010; Sanz-Uribe, Ramos-Giraldo, & Oliveros-Tascon, 2008). Previous investigations have been carried out under controlled lighting conditions. However, one of the main difficulties when developing artificial vision systems for application in the agricultural field is to create solutions which are not affected by lighting variations which may be found on site. Factors such as shade, noise, reflection, or excessive brightness are problems that result from illumination (Zhang et al., 2014). For instance, with images taken in the field, it is not possible to control lighting conditions, which vary, according to time of the day data are acquired. These variances generate difficulties when color and texture information is processed (2D) since there are outliers in features extracted from images. These do not allow for pattern creation, used in the posterior identification of structures present in the scene. Furthermore, aside from issues caused by lighting, there are also difficulties related to occlusion of vegetative structures present in the crop. These occlusions occur due to the nature of the crops and the way in which images are acquired (Djuricic, Weinmann, & Jutzi, 2014; Luo et al., 2016). Implementing techniques based on shape features in images taken in the field make occlusion related problems less acute (Patel, Jain, & Joshi, 2012; Verma, Rossant, Bloch, Orensanz, & Boisgontier, 2014).

Recently, a number of non-destructive techniques using computer vision have been implemented for the detection and identification of vegetative structures. This is achieved through use of images taken in field conditions. Such techniques have been applied to a great variety of crops including: corn (Guerrero et al., 2013; Kurtulmuş & İsmail Kavdir, 2014; Romeo et al., 2013), cereal (Hunt, Daughtry, Mirsky, & Hively, 2014), tomatoes (Verma et al., 2014), and oranges (Patel, Jain, & Joshi, 2011). They are based on the use of two-dimensional features. Three-dimensional techniques, have been implemented with peaches (Nielsen, Slaughter, & Gliever, 2012), grapes (Dey, Mummert, & Sukthankar, 2012), cotton (Latif, Filin, & Eizenberg, 2013), pineapples (Moonrinta, Chaivatratkul, Dailey, & Ekpanyapong, 2010), and some vegetable crops, such as a sprouts, cabbage, cauliflower, sunflowers, and beets (Jay, Rabatel, Hadoux, Moura, & Gorretta, 2015). Finally, some investigations use both two-dimensional and three-dimensional sensors, as is the case in the study presented by Gongal et al. (2016), where a system for automatic apple count was created. The tendency toward research and development of non-destructive systems, that work for the classification of vegetative structures in the field, is evident.

In contrast to the previously discussed investigations, the present study proposes the development of a system for classification of vegetative structures which are present on coffee branches. This system works with moving images taken in field conditions by a monocular system (the main camera of a mobile device) under those lighting conditions present in the environment at the time of image capture. Additionally, a number of branch structures of between one and four centimeters were used. Six classes of vegetative structures were classified: stems, leaves, flowers, and fruits, and in three different maturation stages: unripe, semi-ripe,

and ripe. The use of three-dimensional features is proposed, in order to mitigate the adverse effect that both lighting and occlusion have on the classification systems, which are solely based on two-dimensional features. Color information has a significant role when classifying vegetative structures present on coffee branches. For example, the shape and geometry of coffee beans in their different maturation stages can be quite similar. For this reason, it is necessary to integrate two-dimensional features which allow for the separation of classes. This work uses two-dimensional and three-dimensional features. Therefore, there is no need for implementation of multiple sensors. This creates an easy and accessible solution for coffee growers, since only a mobile device is necessary for data acquisition, employing the principal smartphone camera. The proposed system performed well with uncontrolled occlusion and lighting conditions. Although 82% precision was obtained, classes exist which due to their similarity in shape and color, present difficulties for correct classification.

This article is organized as follows: Section 2.2 presents the proposed materials and method used to classify leaf and stream structures, as well as unripe, semi-ripe, and ripe fruits present on coffee branches. Section 2.3 contains the results and discussion of the proposed method. Finally, Section 2.4 presents the conclusions.

2. Materials and method

The proposed system works in five phases: three-dimensional reconstruction, filtering, sub-sampling, feature extraction, and SVM classification. This system is illustrated in Fig. 1. Input data were sequences of images extracted from videos recorded on a coffee plantation with a mobile device. Information from 12 coffee branches was acquired. The video frames that offered the best information for each branch (approximately 90 per video) were selected. Each branch was reconstructed using the Structure from Motion (SfM) and Patch-based Multi-view Stereo (PMVS) techniques (Furukawa & Ponce, 2007; Snavely, Seitz, & Szeliski, 2006; 2008), thus obtaining a point cloud. Each point cloud was filtered, first removing outliers generated in the three-dimensional phase. Then, the points located in the deepest part, considered to be background, were eliminated, using a band-pass filter. Once the branch had been filtered, a sub-sampling process was carried out, using the voxelized grid approach. For each voxel, information was extracted from different color spaces (two-dimensional features). Additionally, three-dimensional features were obtained from the normal surface, calculated in the regions of each sub-sampled point cloud. Finally, with the two-dimensional and three-dimensional information obtained from eight branches, a SVM was trained to classify six vegetative structures: stems, leaves, and flowers, as well as unripe, semi-ripe, and ripe fruits. The four remaining branches were used to evaluate the proposed method's performance. Algorithm 1 shows the steps necessary in the proposed method.

The input data were acquired from a mobile device. However, the information processing was performed offline at a workstation, such as a computer or server. The computational cost of the developed system varies, in accordance with the number of frames processed, which depends on branch length. The time required by the system to classify a branch ranged between 40 and 60 min. Three-dimensional reconstruction processes and feature extraction required the most time to complete.

2.1. Data set creation

2.1.1. Data acquisition

For this study, Coffea Arabica Variedad Castillo® was used. The images used were acquired with the built-in rear camera of a Samsung S5 SM-G900M mobile device, placed at distances of between

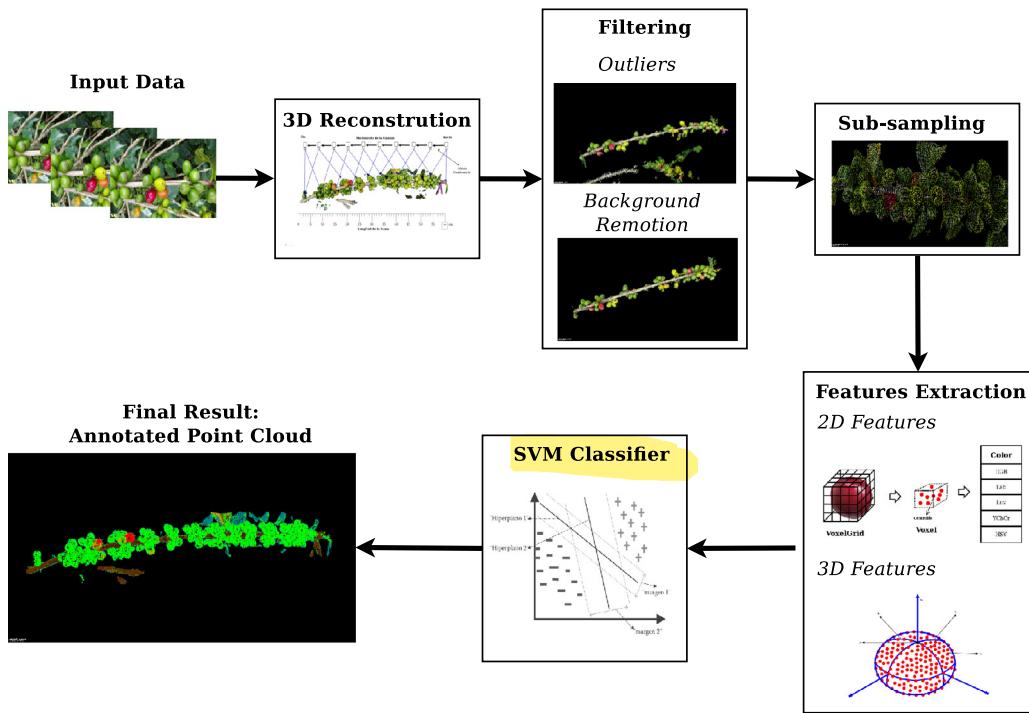


Fig. 1. Proposed system for classification of vegetative structures on coffee branches.

Algorithm 1 Algorithm proposed for classification of vegetative structures on coffee branches.

Input: Coffee branch image sequence.

r = Region Size for finding 3D Features (cm).

Output: Annotated Point Cloud.

Step 1 - 3D Reconstruction: A point cloud is generated from image sequence using SFM and PMVS algorithms.

Step 2 - Filtering: To eliminate atypical points in the point cloud generated in step 1, an statistical outlier filtering approach was performed. Then the scene's background was eliminated using a band-pass filter.

Step 3 - Sub-sampling: Each Point cloud was sub-sampled using a voxelized grid approach. The size of each voxel leaf (x, y, z) was selected as follows:

$$x = 0.2 * \text{Height of Point Cloud}$$

$$y = 0.2 * \text{Width of Point Cloud}$$

$$z = 0.2 * \text{Length of Point Cloud}$$

Step 4 - Feature Extraction: For each voxel in the previous step, the color space median (RGB, Lab, Luv, YCbCr, and HSV) was calculated. Each point had an associated descriptor, with the following features:

$$\text{2D Features} = [R\ G\ L\ a\ b\ u\ v\ Cb\ Cr\ H\ S]$$

For 3D Features, a region with (r) radius was found. In this region, six features were calculated: minimum curvature (k_1), maximum curvature (k_2), Gaussian curvature (K), Mean Curvature (H), Shape index (S_i), and Curvedness (C).

Step 5 - SVM Classification: Each point had 17 features:

$$\text{Features} = [R\ G\ L\ a\ b\ u\ v\ Cb\ Cr\ H\ S\ k_1\ k_2\ K\ H\ S_i\ C]$$

These features were the input for the SVM classifier. Each point was classified and annotated into one of the six classes.

eight and fifteen centimeters from the branch, using a handle, as illustrated in Fig. 2(a). The handle designed had a start/stop button for video recording [see Fig. 2(b)]. The camera was configured to record videos in Full-HD (1920 × 1080), at 30 fps, without activating the flash. Several videos were recorded along each branch, and the information was acquired under uncontrolled field conditions. The videos were recorded over several days, and at different times of a day in order to guarantee different lighting environments. The lighting was that of the environment at the moment of acquisition, and the image background were composed of soil, weeds, dry leaves, and branches from other trees.

The manual image annotation process was achieved with the tool developed by Ramos, Prieto, and Oliveros (2016), with which a Region of Interest (ROI) was automatically generated. This ROI was labeled with the corresponding vegetative structure. Such a tool is capable of quickly analyzing sequences of video with a minimum time for the annotation process. Fig. 3 shows two of the images used, with their respective annotated images. The vegetative structures present in the picture are labeled using different colors. For example, stems are shown in brown, leaves in blueish green, flowers in dark green, unripe fruits in light green, semi-ripe fruits in yellow, ripe fruits in red, and the background in light gray.

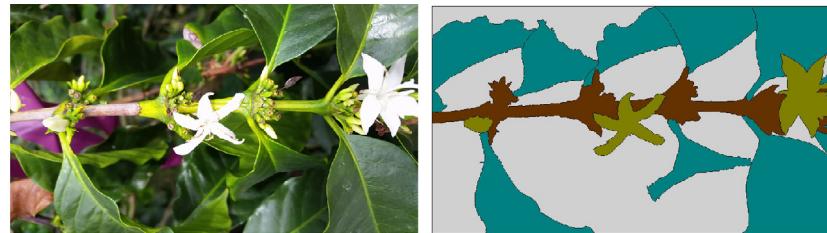
2.1.2. Generation of the Ground Truth point cloud

After obtaining the sub-sampled point cloud, the next step was to reconstruct a manually annotated (Ground Truth) point cloud, where each point corresponded to one of the six classes: stem, leaf, or flower, and unripe, semi-ripe, or ripe fruit. In order to generate the Ground Truth point cloud, for each point (q_i) in the Q sub-sampled point cloud, the two-dimensional images from which the point resulted were identified. Then, using Eq. (1), the position (u, v) of the p pixel was calculated. By means of p pixel location in manually annotated images, color information for each image was extracted and stored in an array called **Color**. The **Color** array contains source pixel color information (RGB), obtained from annotated images. Finally, the median value of each RGB color space from the **Color** array was assigned to color components in the q_i



(a) Support and handling of the mobile device for information video recording.
(b) Start/stop button for mobile device.

Fig. 2. Example of coffee branch image acquisition.



(a) Source Image 1.

(b) Annotated Image 1.



(c) Source Image 2.

(d) Annotated Image 2.

Fig. 3. Acquired images with their respective annotations.

point of the Ground Truth point cloud. The median was used in order to prevent generation of new color data, as generation of a Ground Truth point cloud that could maintain annotated images labels (colors) was expected.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K[R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (1)$$

where X , Y , and Z are the coordinates of q_i point, K is the matrix of intrinsic parameters found in the camera calibration phase, R the rotation matrix, and t the translation vector. R and t are calculated in the reconstruction phase.

Fig. 4 shows two of the generated Ground Truth point clouds. Ripe fruits are identified with red, stems with brown, leaves with blueish green, flowers with dark green, unripe fruits with light green, and semi-ripe fruits with yellow.

Table 1
Intrinsic camera parameters.

Focal distance		Center of the camera		Distortion coefficient				
f_x	f_y	C_x	C_y	k_1	k_2	k_3	k_4	k_5
698.1	698.1	319.5	179.5	711.2	87.8	0	0	-3.58

2.2. Three-dimensional reconstruction

The first step in the three-dimensional process consists of finding the camera's intrinsic parameters. In order to do this, the camera calibration algorithm developed by [Zhang \(2000\)](#) was used, where several images, which form a planar pattern, were acquired from different orientations. In this instance, the planar pattern is a chess board with squares measuring 1.7 by 1.7 cm. **Table 1** illustrates the intrinsic parameters found following the calibration phase.

Fig. 5 illustrates the process employed for SfM implementation. The mobile device's camera recorded videos in high resolution along each branch, beginning from the tip, and continuing

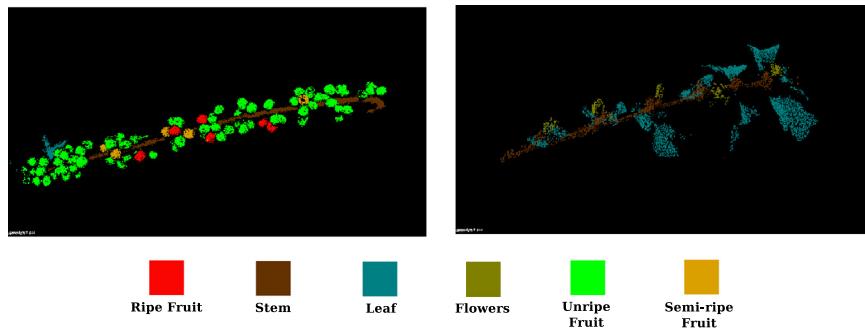


Fig. 4. Example of generated Ground Truth point clouds.

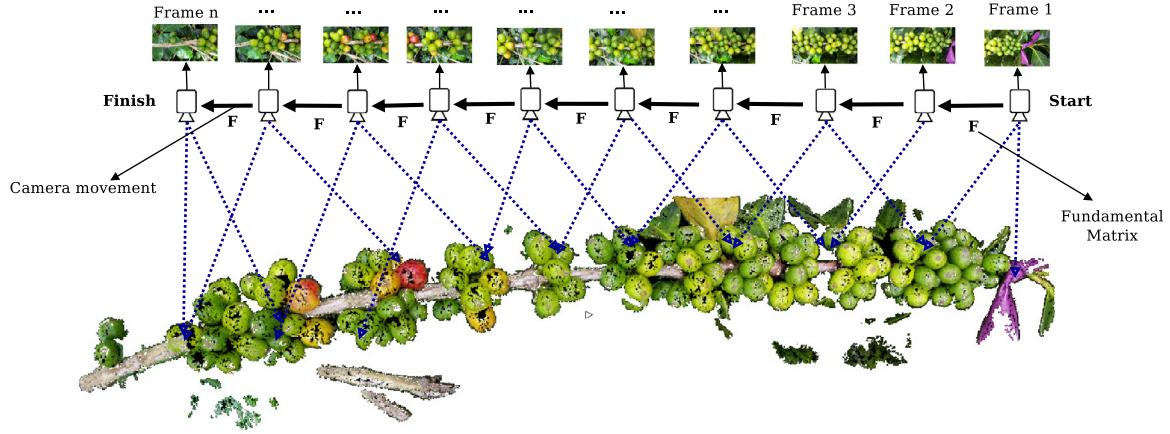


Fig. 5. Process used for SfM implementation.

until reaching the main trunk [see Fig. 2(b)]. The length per branch ranges between 40 and 60 cm, and the average video duration is 15 seconds. From each video, one in five frames is selected. Approximately 90 images are used to reconstruct each branch.

The three-dimensional reconstruction of each branch was performed by implementing the Bundler SfM and PMVS algorithms. Next, algorithms proposed by Snavely and Furukawa were used. They are described in detail in Furukawa and Ponce (2007), Snavely et al. (2006); 2008), and will be briefly explained here. The Bundler SfM is a machine vision system which implements SfM for the three-dimensional reconstruction of scenes which result from disorganized collections of images, such as, pictures from the internet. Firstly, Bundler SfM extracts keypoints from a set of images, using Scale-Invariant Feature Transform (SIFT) (Lowe, 2004). Next, it matches these points using the Approximate Nearest Neighbor (ANN) technique. With the intrinsic camera parameters, obtained through a calibration process, as well as matched points between images, camera movement and extrinsic parameters can be found using the fundamental matrix (Hartley & Zisserman, 2004). The Bundler SfM algorithm selects two of the frames with the largest number of matched keypoints. The points in these views are projected through a triangulation process, thus forming a basic structure. The points from the remaining views are added to the basic structure. Finally, a three-dimensional reconstruction of the scene is obtained, using the Bundler adjustment, which modifies the intrinsic camera parameters (focal point, distance, and distortion) to minimize error in the re-projection. The Bundler SfM produces disperse point clouds. In order to obtain dense point clouds, the PMVS is used. This is an algorithm for stereo vision which uses a set of images and the extrinsic camera parameters, to reconstruct the three-dimensional structure of an object or scene visible in the images. This algorithm reconstructs rigid structures only. The PMVS

generates a set of oriented points, instead of a polygonal model, where both the three-dimensional coordinates and the normal surface are calculated for each oriented point.

Results obtained in each phase of the three-dimensional reconstruction of one of the branches used are shown in Fig. 6. In Fig. 6(a) and (b) the key points found in two consecutive frames are shown. In Fig. 6(c) key point matching is shown. In Fig. 6(d) the disperse point cloud generated by the Bundler SfM algorithm is shown, and in Fig. 6(e) the dense point cloud generated by the PMVS is shown.

2.3. Filtering

The purpose of this phase is to eliminate any data that do not belong to the branch (atypical values, for example) and to separate vegetative structures present on the branch from the scene background. First, outliers are filtered, which eliminates atypical values in the point cloud. Next, by means of a band-pass filter, the scene's background is removed. The implementation of both filters was performed using the Point Cloud Library (Rusu & Cousins, 2011).

When the three-dimensional reconstruction is created using SfM and PMVS, atypical points are generated. These are errors that occur when triangulating image points. In order to eliminate these points, a statistical analysis is performed in the neighborhood area of each point, and afterwards, those points with a distance larger than the standard deviation are eliminated from the mean distance in the neighborhood. Fig. 7(a) shows a point cloud following the three-dimensional reconstruction process and, in Fig. 7(b), the result of the outlier filtering phase on this branch can be observed. The largest number of atypical values is present in the background of the scene. Each dense point cloud has on average 300,000 points. The neighborhood in which each point was

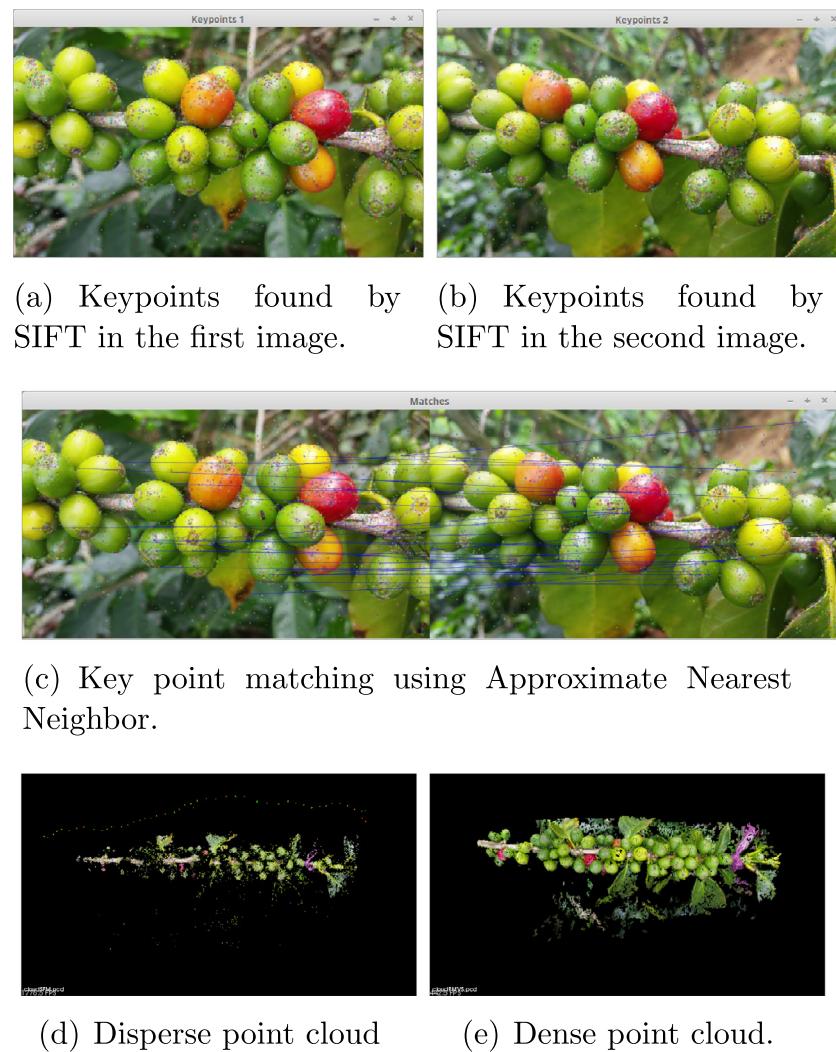


Fig. 6. Main steps of the three-dimensional reconstruction algorithm.

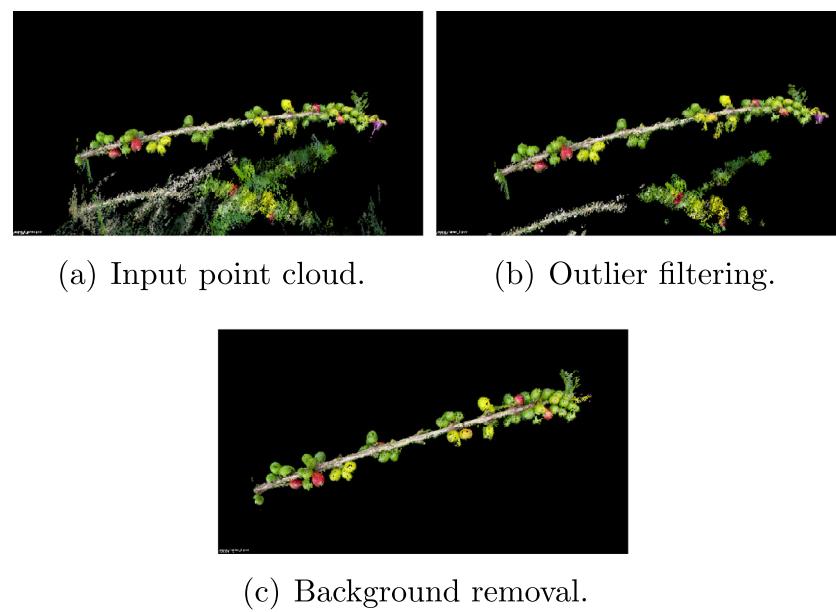


Fig. 7. Filtering phase.

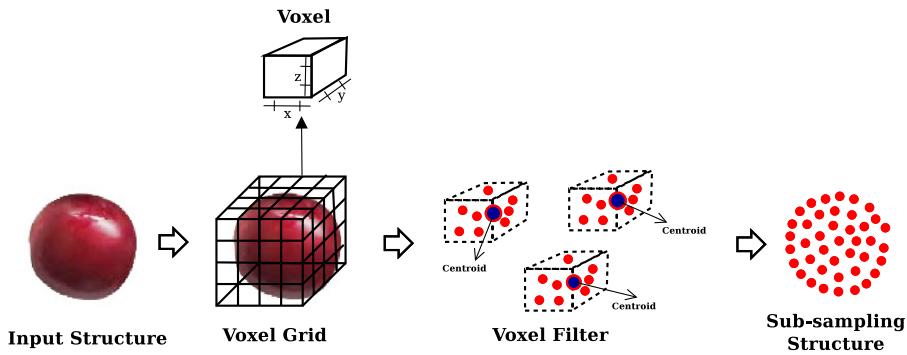


Fig. 8. Sub-sampling using the voxelized grid approach.

processed (to determine whether they corresponded to an outlier or not) corresponds to 0.1% of the total shape of the point cloud.

In order to remove the scene's background, a band-pass filter was used in the expected dimension (direction) where the points outside of a specific range were eliminated. The implemented filter worked on the z axis, since the background was in that direction. To remove a p point, one must evaluate whether the component in z of the p point is found within the previously defined Z_{\min} and Z_{\max} range, that is to say, if $Z_{\min} < p_z < Z_{\max}$. Fig. 7(c), details how the background of the scene is removed in this phase, leaving only branch structures in the point cloud.

When the filtering phase finished, it was possible to eliminate the atypical values from the reconstruction and as well as the background of the scene. Additionally, for each point cloud, a reduction of 50% of the original shape was obtained, which reduced the computational time in the following processes, due to a lesser use of data which contained the most relevant information about the scene.

2.4. Sub-sampling

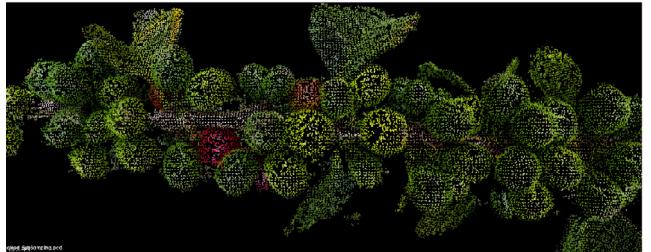
The filtered point clouds were sub-sampled in order to reduce computation time in the process of two-dimensional and three-dimensional feature extraction. The implemented method for this objective was a voxelized grid approach. Fig. 8 shows the sub-sampling process carried out on a coffee bean structure. Firstly, the structure's points are grouped in voxels; next, centroids are gathered in each voxel, which then become new information in the sub-sampled point cloud. The size of the voxels ((x, y, z)) is automatically selected, and depends on point cloud volume. At this point, some voxels measuring 20% of the total point cloud volume were used. At the end of the process, a new point cloud was generated, in which a lesser number of points was obtained, but the geometric features of the branches remained the same. In Fig. 9, the input point cloud [Fig. 9(a)] and the result of coffee branch sub-sampling [Fig. 9(b)] can be observed. The reduction of the number of points on each branch resulted in approximately 50% of the original size of the structure.

2.4.1. Three-dimensional features

The three-dimensional features extracted for each structure are based on the curvatures present in sub-regions of each sub-sampled point cloud. Overall, six features were found: four curvatures (maximum (k_2), minimum (k_1), Gaussian (H), mean (K)), shape index (S_i), and curvedness (C). The property of having a curved surface is called curvedness, where $C = 1$ indicates that the surface is completely curved, and $C = 0$ indicates a planar surface. These features were extracted using surface normals, which were calculated in regions of each point cloud. Each region was found using the nearest-neighbor search, in a data structure divided with



(a) Input point cloud.



(b) Sub-sampled point cloud.

Fig. 9. Sub-sampling phase.

kd-Trees (Bentley, 1975). A structure of kd-Tree data organizes a number of points in a space with k dimensions. For this investigation, three-dimensional trees were used ($k = 3$). Fig. 10 shows the process used to find three-dimensional features in a region, in this case, those of a coffee bean.

To calculate the surface normal, a breakdown into singular values was performed in each region. The eigenvector which corresponds to the smallest eigenvalue approximates to the \bar{N} normal of S surface on p point, whereas the k_1 minimum and k_2 maximum curvatures of S surface are calculated based on the following eigenvalues:

$$k_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \quad (2)$$

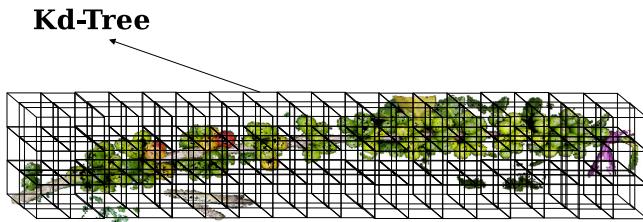
$$k_2 = \frac{\lambda_2}{\lambda_0 + \lambda_1 + \lambda_2}, \quad (3)$$

where $\lambda_0 < \lambda_1 < \lambda_2$.

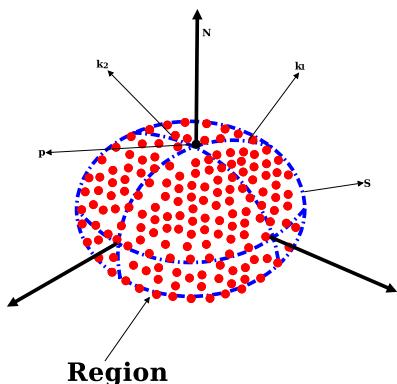
Remaining three-dimensional features, were calculated using the following equations:

- Gaussian curvature

$$K = k_1 k_2. \quad (4)$$



(a) Extraction of regions using kd-Tree.



(b) Calculation of curvatures n located regions.

Fig. 10. Region extraction process and curvature calculation.

• Mean curvature

$$H = \frac{1}{2}(k_1 + k_2). \quad (5)$$

• Shape index

$$S_i = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \left(\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \right). \quad (6)$$

• Curvedness

$$C = \sqrt{\frac{k_1^2(p) + k_2^2(p)}{2}}. \quad (7)$$

2.4.2. Two-dimensional features

The two-dimensional features were extracted when the point cloud sub-sampling was performed. The median of the various color spaces (RGB, Lab, Luv, YCbCr, and HSV) was calculated for each voxel point (see Fig. 11). Then, each centroid was matched

with a set of features which was composed of the following color components: R, B, L, a, b, u, v, Cb, Cr, H, and S. From the color space RGB, the R and G components were used. The B component was not used because, in the experimentation phase, it was shown that, system performance decreased when using the B color component. This occurred because of large variations in intensity in this color space, due to the different lighting conditions in information acquisition. In the case of Lab space, all components were used. In the remaining color spaces, the lighting component was not considered because was already present in the L component of the Lab space.

In the images captured, different intensities can be observed for a single point, as they were acquired under uncontrolled lighting conditions. For this reason, just as for the generation of the Ground Truth point cloud, working with the median rather than the mean or standard deviation was preferred. This prevented the generation of new data in the features associated with each point.

2.5. Classification

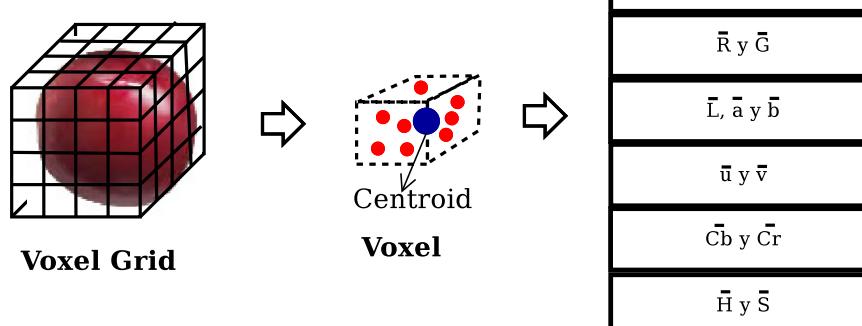
For the classification of vegetative structures, a SVM was used, employing the LIBSVM library, developed by Chang and Lin (2011). The SVM was trained with the seventeen features, eleven of them corresponding to color spaces (R, G, L, a, b, u, v, Cb, Cr, H, and S), four of them to curvature (maximum, minimum, Gaussian, and mean), and the remaining two to shape index and curvedness. For system training, eight of the twelve branches were used. Remaining branches were used to evaluate the performance of the proposed method. Overall, six semantic classes were classified: stems, leaves, and flowers, as well as unripe, semi-ripe, and ripe fruit. The classifier output is a point cloud, where each point is assigned to one of the six classes. The standard SVM model (Chang & Lin, 2011) is defined as:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (8)$$

$$\text{s.t. } y_i(\langle \mathbf{W}, \phi(\mathbf{x}_i) \rangle + b) + \xi_i \geq 1, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, l$$

Where \mathbf{x}_i are the training vectors, subjects to $\mathbf{x}_i \in \mathbb{R}^n$, $y_i \in \mathbb{R}^l$ such that $y_i \in \{0, 1, 2, 3, 4, 5\}$, and l is the number of training samples used. $\phi(\mathbf{x}_i)$ maps \mathbf{x}_i into a higher-dimensional space and $C > 0$ is the regularization parameter. The dual model of Eq. (8) is defined as follows:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^l \alpha_i \quad (9)$$

**Fig. 11.** Extraction of two-dimensional features.

Best $\log_2(C) = 4.0$ $\log_2(\gamma) = 3.0$ accuracy = 78.11 %
 $C = 16.0$ $\gamma = 8.0$

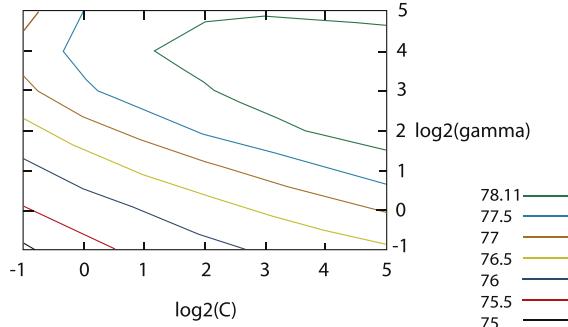


Fig. 12. Cross-validation process using a region size = 2.5.

$$\text{s.t. } \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l$$

$k(x_i, x_j)$ is the kernel function, in the case that we use the Gaussian kernel defined in Eq. (10).

$$k(x_i, x_j) = \exp(-\sigma \|x_i - x_j\|^2), \quad (10)$$

where $\sigma > 0$. To tune the C and σ parameters from the SVM, a cross-validation process was used. Fig. 12 shows the results obtained following said cross-validation process. In this case, the region size used for 3D feature extraction was 2.5 cm and the optimal values for C and σ were 16 and 8 respectively.

SVM was extended to multi-label classification using the one versus rest method. Thus, k binary classifiers were constructed, $f_i (i = 1, \dots, k)$ that separated class i from other $k - 1$ classes. The decision rule for the k -class problem is:

$$y(x) = \arg \max_{1 \leq i \leq k} f_i(x) \quad (11)$$

3. Results and discussion

In this study, pieces of information from 12 coffee branches were acquired. For each branch, videos were recorded, and the

Table 2
Percentage of data used by class in the training and validation phases.

Class	Training	Validation
Ripe fruit	4.52%	2.99%
Stems	14.03%	14.36%
Leaves	12.59%	27.88%
Flowers	2.58%	1.42%
Unripe fruit	64.30%	51.76%
Semi-ripe fruit	1.98%	1.59%

following processes were carried out: three-dimensional processing, filtering, sub-sampling, feature extraction, and generation of Ground Truth information, all of which are detailed in Section 2.2. For SVM training, eight branches were used. The four remaining branches were used to validate the classifier's performance. In Fig. 13, the four branches used for validation of the proposed method are shown. These branches were selected in a way that would ensure data obtention for each of the six vegetative classes or structures of interest. Table 2 shows the percentage of data used per class for the training and validation phases. The highest percentage of data was found in the unripe fruit class, with a respective percentage of 64.3% and 51.76% in the training and validation phases. This is due to a higher amount of information about unripe fruits than any of the other classes on coffee branches. In order to attend to the unbalanced data set, shown in Table 2, an SVM extension, known as Weighted SVM (WSVM) (Chang & Lin, 2011; Luo & Chen, 2013; Osuna, Freund, & Girosi, 1997), was used. Here a different penalty is applied for each class, using higher penalties for the class with the greatest number of samples. In WSVM, the models presented in Eqs. (8) and (9) respectively become:

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \mu_i \xi_i \quad (12)$$

$$\text{s.t. } y_i (\langle \mathbf{W}, \phi(x_i) \rangle + b) + \xi_i \geq 1, \quad i = 1, 2, \dots, l, \\ \xi_i \geq 0, \quad i = 1, 2, \dots, l$$



(a) Branch No. 1.



(b) Branch No. 2.



(c) Branch No. 3.



(d) Branch No. 4.

Fig. 13. Branches used for system validation.

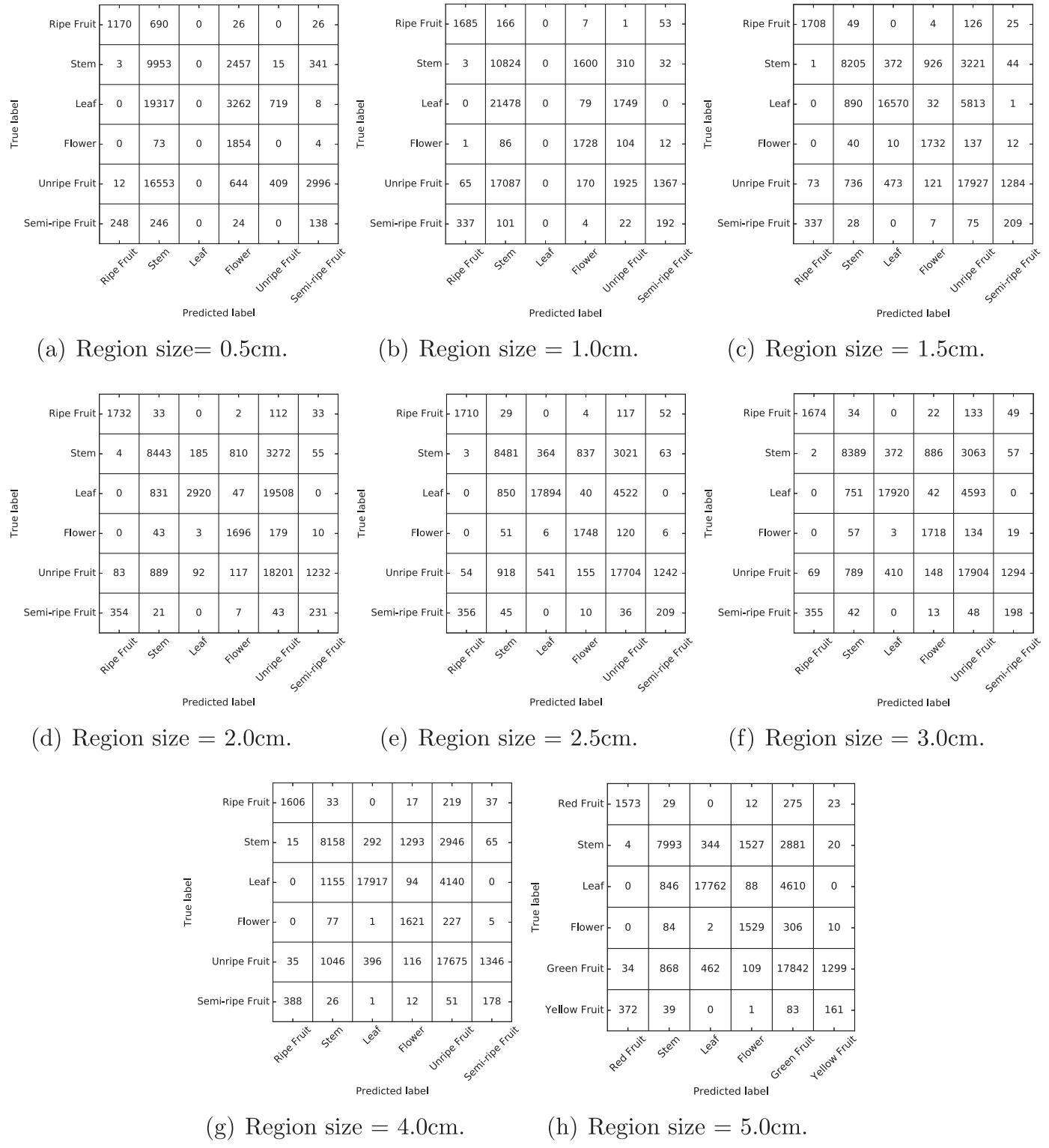


Fig. 14. Confusion matrices according to region size.

and

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^l \alpha_i \quad (13)$$

$$s.t. \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq \mu_i C, \quad i = 1, 2, \dots, l,$$

where $\mu_i (i = 1, \dots, l)$ indicates the weight of the class x_i .

Although there is an imbalance in the data according to the class used, WSVM allowed system performance to remain unaffected by the class imbalance. For example, the ripe fruit class has 4.52% training data. However, Table 3 shows that one of the best system performances was obtained for the ripe fruit class, with an F1-Score between 0.7 and 0.83. This shows that the proposed system does not depend directly on the amount of data used per class for training.

Table 3

F1-Score for class vs. region size.

Class	Region size (cm)							
	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5.0
Ripe fruit	0.70	0.84	0.85	0.85	0.85	0.83	0.81	0.79
Stems	0.33	0.35	0.72	0.73	0.73	0.73	0.73	0.73
Leaves	0.00	0.00	0.81	0.22	0.85	0.85	0.85	0.85
Flowers	0.36	0.63	0.73	0.74	0.74	0.72	0.64	0.59
Unripe fruit	0.04	0.16	0.75	0.59	0.77	0.77	0.77	0.77
Semi-ripe fruit	0.07	0.17	0.19	0.21	0.19	0.17	0.16	0.15

3.1. Evaluation indexes

In order to evaluate the performance of the proposed system, a confusion matrix was used, as well as metrics for total accuracy, precision, recall or specificity, and the F1-Score (Eqs. (14), (15), (16) and (17)). Accuracy relates correctly classified data as well as all evaluated data. Precision measures the relationship that exists between the quantity of correctly classified data in a given class and the number of datum classified for said class. Recall measures the relationship that exists between correctly classified data and the real number of datum that belong to a class. Finally, and in order to avoid a biased evaluation of the system, the F1-Score index was calculated. This offers a balance between precision and recall. Said metrics were defined as follows:

$$\text{Accuracy} = \frac{\sum_i^n cm_{ii}}{\sum_i^n \sum_j^n cm_{ij}}. \quad (14)$$

$$\text{Precision} = \frac{cm_{ii}}{\sum_j^n cm_{ji}}. \quad (15)$$

$$\text{Recall} = \frac{cm_{ii}}{\sum_j^n cm_{ij}}. \quad (16)$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (17)$$

where cm is the constructed confusion matrix with classification results and Ground Truth data, and n is the number of classes.

3.2. Vegetative structure classification

In Section 2.6, the feature extraction process was explained. In the case of three-dimensional features, it was necessary to determine the size of the regions from which they were extracted. Table 3 shows, by class, the F1-Score index obtained by the system, when the size of a region varied. The changes in region size were between 0.5 and 5 cm. In total, for all branches, 61,188 points were classified. Fig. 14 shows the confusion matrices obtained for each region size. When examining the system performance by class, the following was observed:

- For the unripe and ripe fruit classes, the best performance was achieved by using a 2.5 cm region size, where an F1-Score of 0.85 was obtained for the ripe fruit class and 0.77 for the unripe fruit class (see Table 3). These results are reasonable, taking into account that the range within which the size of a coffee bean may vary in these maturation stages is between 1 and 2.5 cm. As region size increases, system performance decreases, since the regions found by the system do not fit the geometry of the fruit.
- In the stems class, the system-obtained F1-Score was 0.72, from a size of 1.5 cm. From this point onward, system performance remains the same. This happens because the length of the stem within each fruit cluster, ranges between 1.5 and 5 cm.



(a) Annotation of semi-ripe fruit.



(b) Source image

Fig. 15. Annotated semi-ripe fruit result analysis.

- The leaf class, shows a low performance for 0.5, 1, and 2 cm regions. For region sizes between 0.5 and 1 cm, a 0.0 F1-Score is obtained because the system confuses some unripe fruit with leaves [see Fig. 14(a) and (b)]. This happens due to the similarity in region shape of unripe fruit and leaves. Something similar happens in the case of 2 cm, where a 0.22 F1-Score is obtained. However, in this case, leaf regions are those which resemble unripe fruit regions [Fig. 14(d)].
- The flowers present on processed coffee branches are in their open flower stage, when each flower has a radius measuring between 2 and 3 cm. In the flower class, the best performance was obtained between 2 and 2.5 cm, with a 0.74 F1-Score. It can also be shown that the larger the size of the region, the lower the performance, in the flower class. This is true because, with larger regions, the system does not adjust to the geometry of the flowers.
- In the leaves class, the unripe fruit's lowest performance was obtained with the 0.5, 1, and 2 cm region sizes, with a 0.04, 0.16, and 0.59 F1-Scores respectively, because the system confuses these two classes. Unripe fruits can measure between 1 and 2.5 cm, depending on their ripening time. On the branches used in the experiments, this radius ranged between 1.5 and 2.5 cm. Table 3 shows that the best performance for the unripe fruit class was obtained with a 1.5 cm radius, where a 0.77 F1-Score was obtained.
- The semi-ripe fruits generally present in zones in which color tonality resembles that of ripe and unripe fruits, as shown in Fig. 15(b), where a semi-ripe fruit is circled in white. Because of this, the class with the lowest performance is that of semi-ripe fruit, where the highest F1-Score is 0.21, and this was obtained through use of a 2 cm region. This is reasonable in terms of reality, as the size of a semi-ripe fruit must range

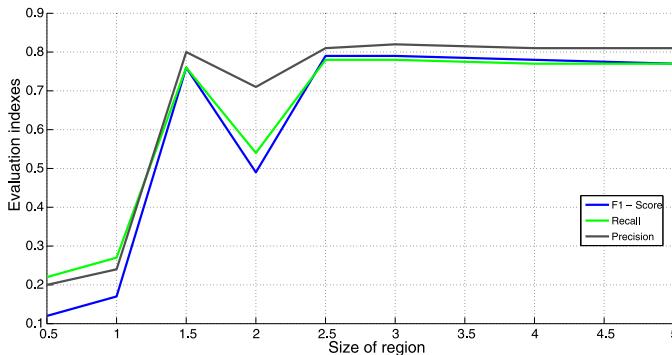


Fig. 16. Region size vs. evaluation indexes.

between that of an unripe and ripe fruit. Fig. 14 shows that semi-ripe fruits are confused with the ripe fruits and, as the size of the region grows, they are also confused with unripe fruits. This happens because semi-ripe fruits have color features and shapes which resemble both the ripe and unripe fruit classes. In Fig. 15(a), the system classification for a semi-ripe fruit, which is circled in white, together with one of the source images from which the three-dimensional reconstruction was obtained can be observed. The system separates the fruit from other structures present in the scene. This is achieved thanks to the three-dimensional features used, because the curvatures in other structures such as stems, leaves, and flowers are different than those of fruits. When visually evaluating results obtained in this structure (Fig. 15(a)), it can be observed that the system did separate the fruit, but classified zones as unripe, semi-ripe and ripe fruit.

- It is important to analyze why system performance remains constant for certain classes, even when the region size is enlarged to extract three-dimensional features. In the case of the stem, leaf, and unripe fruit classes, system performance converges and remains constant when the region size is enlarged. This does not occur in the remaining classes: (flowers, ripe and semi-ripe fruits) if region size is enlarged, system performance decreases. In Table 2, where the data percentage used per class was compared, it can be seen that those classes in which performance decreases when the size of the region is enlarged are those with the least amount of data. Because the objective of the proposed system is to classify vegetative structures present on coffee branches taken from videos in field conditions by a mobile device, the branches used were randomly chosen from the crop. Most coffee branches used had a larger proportion of unripe fruit than fruit in any other stage. The branches containing flowers did not have fruit structures, but rather included more leaves and stem structures. When the region size, from which features are extracted is enlarged, classes with the least amount of data perform poorly. The regions processed do not fit into known classes, and the system starts to confuse these regions.

The six vegetative structures present on the coffee branches have an average size of 2.5 cm. For example, flowers present on processed branches range from 1 to 3 cm. Coffee fruits can measure between 1 and 2 cm, depending on the maturation stage. As for the stem fragments, they have a length which varies between 1.5 and 4 cm, for each fruit cluster. Finally, leaf fragments present in the scene measure between 2 and 5 cm. When Fig. 16 was analyzed, where the average evaluation indexes for all classes are shown, with the region size from which three-dimensional features

were extracted, the best performance was obtained using a 2.5 cm radius, with a 0.79 F1-Score, a 0.81 precision score, and 0.78 recall were obtained.

The results of using a 2.5 cm long region are shown in Fig. 17. On the left side, Ground Truth point clouds, and on the right side, point clouds annotated by the system can be observed. The classification made by the system is good, because the different vegetative structures on the branches can be visually identified. However, there are zones that are not correctly classified. For example: the leaf on branch No. 1 [Fig. 17(b)], the stems near the flowers on branch No. 4 [Fig. 17(h)], and the semi-ripe fruit on branches 1, 2, and 3 [Fig. 17(b), (c), and (d)].

Another experiment attempted to combine the unripe, semi-ripe, and ripe fruit classes into a single group that was called "fruits". It attempted to reevaluate system performance by using the same 2.5 cm long region size. The results of this experiment are shown in Fig. 18's confusion matrix, where 94% accuracy was obtained in fruit classification.. This is useful for applications where a fruit count is to be performed, for example, in coffee mass estimation.

Table 4 shows the results of various developments and existing computer vision applications, which function in field conditions, and in some cases, under uncontrolled lighting conditions, as is the case in this study. It can be shown that, when working only with two-dimensional features, lighting conditions difficulties, occlusion, and noise occur (Patel et al., 2011; 2012; Verma et al., 2014). On the other hand, when using three-dimensional information, these difficulties can be partly overcome. Nevertheless, in some cases, it is necessary to use two-dimensional information, because geometric data (3D) are not sufficiently relevant for the classification of vegetative structures. An example of this is the case of coffee, where fruits in their different maturation stages can have the same geometric shape, and yet have different colors.

Some of the investigations listed in Table 4 have been applied to different crops, such as, tomatoes (Verma et al., 2014), pineapples (Moonrinta et al., 2010), grapes (Dey et al., 2012), and apples (Gongal et al., 2016). The previous studies have been applied to crops where the disposition of vegetative structures allows for improved separation of classes than those of coffee crops. On coffee branches, fruits are found in clusters, and their colors are very similar. For instance, semi-ripe fruits are similar in shape and color to ripe and unripe fruits. Furthermore, leaf color intensities are also similar to that of unripe fruit. The work presented by Ramos, Prieto, Montoya, and Oliveros (2017) counts fruits present on coffee branches from images captured in field conditions by a mobile device. The proposed system performed well, with an coefficient of determination (R^2) of 0.82. This (R^2) is the correlation square between predicted and actual scores. However, one of the disadvantages of the work performed by Ramos et al. (2017) was the fact that it was impossible to correctly separate the branch from the background scene.

Related investigations in Table 4 tend to classify only fruits, whereas in this study, structures are separated, for example: fruits in their different maturation stages, leaves, flowers, and stems present on each branch. It is also important to highlight the difference in the data acquisition process in the field. In this work, only a mobile device was used, and some moving images were acquired with variant uncontrolled lighting conditions. With a 0.79 F1-Score, the proposed system can be considered adequate for the vegetative structures classification of forms present on coffee branches, because it integrates two-dimensional and three-dimensional information that works in different noise and lighting conditions, as well as in outdoor environments. Also, it is accessible to coffee growers, as one requires nothing more than a mobile device for information acquisition.

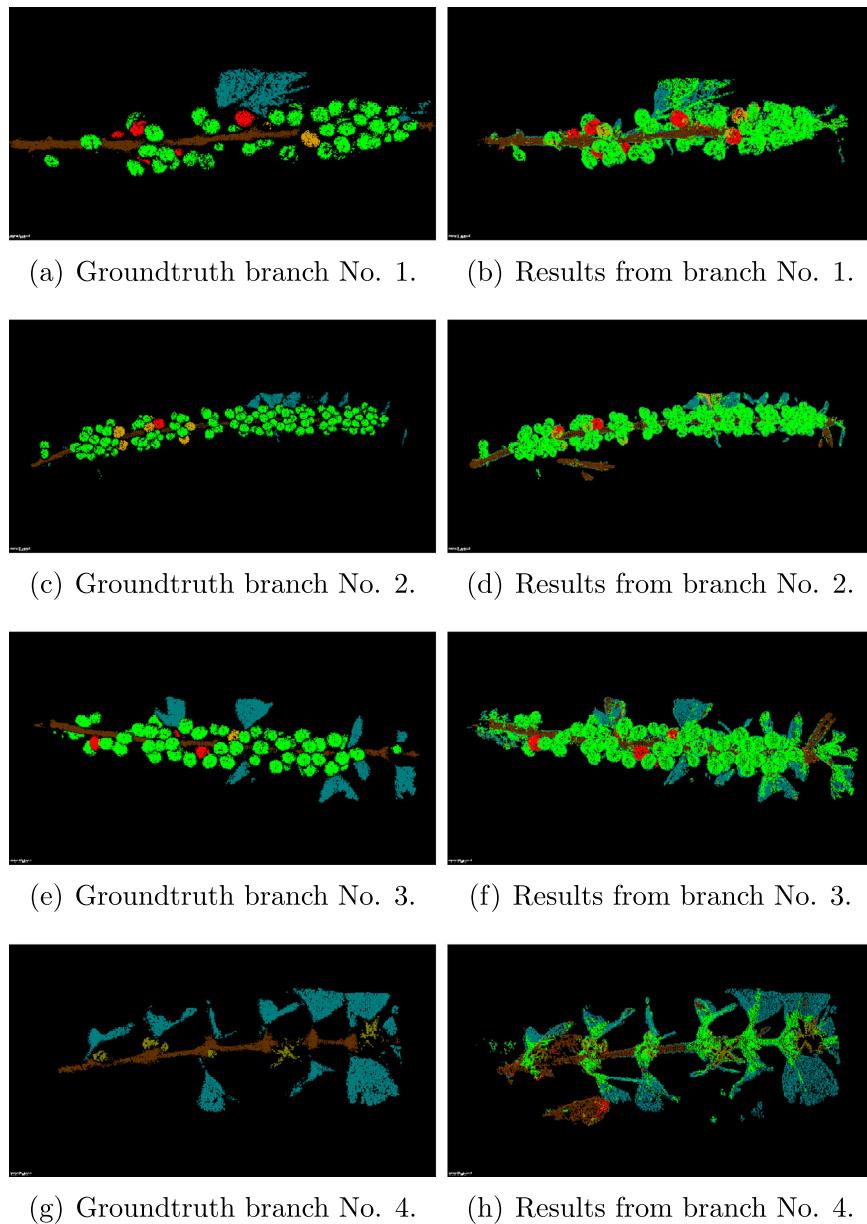


Fig. 17. GroundTruth point clouds vs. proposed system results.

4. Conclusions

This work presented the development of a system for the classification of vegetative structures: flowers, stems, and leaves, as well as unripe, semi-ripe, and ripe fruits present on coffee branches, using sequences of images acquired on a coffee plantation, with a mobile device's camera. The use of the SfM and PMVS techniques allowed for the obtention of an economical solution to acquisition of three-dimensional information in field conditions. This is because, for the three-dimensional reconstruction of branches, only those videos acquired on the mobile device were necessary; in this case, a mobile phone moved along the branches to record the videos. Generally, in images acquired from coffee branches, there are withered leaves, weeds, and branches from other trees that comprise part of the scene's background. This background makes it difficult to classify branch vegetative structures, because non-homogeneous structures, which confuse the automatic classification systems, appear. Using three-dimensional information, it was

possible to remove the scene's background, and classify only those structures present on the branch.

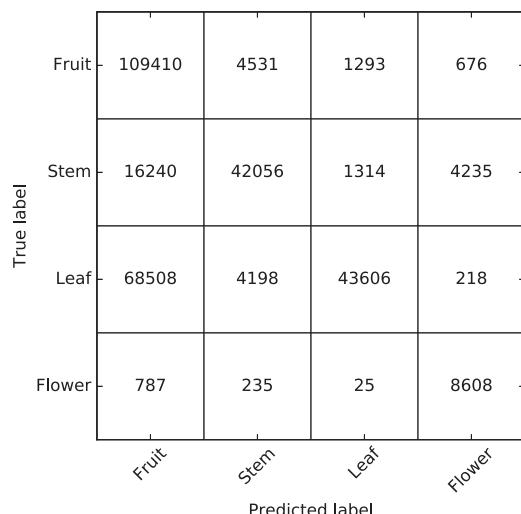
Due to lighting conditions difficulties, and occlusions which present when working with images acquired in field conditions, this work made use of color features, curvatures, and shape and curvedness indexes for the classification of the different vegetative structures present in the scene. Although acquired images had different noise and lighting conditions, the proposed system showed no alteration in the face of these occlusion and lighting conditions. Experimental results showed 81% global precision, 78% global accuracy, and a 0.79 F1-Score in the validation data.

When analyzing system performance for each class, it was shown that the size of the region from which three-dimensional features are calculated, directly affects the classification result. System performance proved to be related to structure size. For example: in order to separate fruits from other structures, a region size between 1.5 and 2.5 cm must be used. This was the real size range for a coffee bean. The average real size of vegetative

Table 4

Comparative table of related investigations with respect to the proposed method.

Reference	Crop	Features used	Issues	Number of structures	Avg. radio (cm)	Performance
Moonrinta et al. (2010)	Pineapples	SIFT	SfM ambiguity	(1) Fruits	10	Error ≈ 29.45%
Patel et al. (2011)	Cherries, grapes, apples, pears, strawberries, oranges, peaches, etc.	Color and shape	Brightness	(1) Fruits	3	Accuracy = 98%
Hunt et al. (2014) Dey et al. (2012)	Cereals Grapes	Infrared color 2D: Color and curvatures	Computational cost Noise induced by the independent classification of each point	(1) Crop. (3) Fruits, leaves and stems	Does not report 3	$r = 0.73$ Accuracy = 84.8%
Verma et al. (2014) Jay et al. (2015)	Tomatoes Vegetables	Shape Height and vegetation index	Occlusion Structure movement due to wind	(1) Fruits (1) Leaves	5 20–60	Error ≤ 10% $r = 0.94$
Gongal et al. (2016)	Apples	Fruit color and spatial location	Does not report	(1) Fruits	2–3	Accuracy = 79.8%
Ramos et al. (2017)	Coffee	Color and shape	Background remotion	(4) Fruits	1–2	$R^2 = 0.82$
Proposed method	Coffee	Color, curvature, curvedness and shape	Similarity of curvature and color spaces in some branch areas	(6) Unripe fruit, semi-ripe fruit, ripe fruit, leaves, stems and flowers	3–4	F1-Score = 0.79%

**Fig. 18.** Confusion matrix combining the classes: red fruit, green fruit, and semi-ripe fruit.

structures (leaves, flowers, stems, as well as unripe, semi-ripe, and unripe fruits) was 2.5 cm long. The best performance was obtained when working with a 2.5 cm, where 0.81 precision and a 0.78 recall were obtained. This indicates that the proposed method performs well when classifying vegetative structures present on the coffee branches. Nonetheless, there are structures with shapes and colors that do not allow for an accurate classification, due to their similarity to other structures. For instance, there are zones on unripe fruits with shapes and colors similar to those of leaves. As for semi-ripe fruits, in some cases, their color is similar to that of unripe fruits and, in other instances, to ripe fruits. In order to improve system performance, the inclusion of different features based on textures that complement those used in the present work and which allow for a better classification of the semi-ripe fruit class is proposed.

Acknowledgements

This work was partially sponsored by the *Departamento Administrativo de Ciencia, Tecnología e Innovación—Colciencias*, under project code 225166945211.

References

- Ajdadi, F. R., Gilandeh, Y. A., Mollazade, K., & Hasanzadeh, R. P. (2016). Application of machine vision for classification of soil aggregate size. *Soil and Tillage Research*, 162, 8–17. <http://dx.doi.org/10.1016/j.still.2016.04.012>.
- Apaza, R. G., Portugal-Zambrano, C. E., Gutiérrez-Cáceres, J. C., & Beltrán-Castañón, C. A. (2014). An approach to improve the recognition of defects in coffee beans using retinex algorithms. In *Computing conference (CLEI), 2014 XL Latin American* (pp. 1–9). doi:[10.1109/CLEI2014.6965102](https://doi.org/10.1109/CLEI2014.6965102).
- Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18, 509–517. doi:[10.1145/361002.361007](https://doi.org/10.1145/361002.361007).
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 27:1–27:27. software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Condori, R. H. M., Humari, J. H. C., Portugal-Zambrano, C. E., Gutiérrez-Cáceres, J. C., & Beltrán-Castañón, C. A. (2014). Automatic classification of physical defects in green coffee beans using CGLCM and SVM. In *Computing conference (CLEI), 2014 XL Latin American* (pp. 1–9). doi:[10.1109/CLEI2014.6965169](https://doi.org/10.1109/CLEI2014.6965169).
- Dey, D., Mummert, L., & Sukthankar, R. (2012). Classification of plant structures from uncalibrated image sequences. In *2012 IEEE workshop on applications of computer vision (WACV)* (pp. 329–336). doi:[10.1109/WACV.2012.6163017](https://doi.org/10.1109/WACV.2012.6163017).
- Djuricic, A., Weinmann, M., & Jutzi, B. (2014). Potentials of small, lightweight and low cost multi-echo laser scanners for detecting grape berries. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5, 211–216. doi:[10.5194/isprsarchives-XL-5-211-2014](https://doi.org/10.5194/isprsarchives-XL-5-211-2014).
- Dutta, M. K., Singh, A., & Ghosal, S. (2015). A computer vision based technique for identification of acrylamide in potato chips. *Computers and Electronics in Agriculture*, 119, 40–50. <http://dx.doi.org/10.1016/j.compag.2015.10.007>.
- El-Bendary, N., Hariri, E. E., Hassanien, A. E., & Badr, A. (2015). Using machine learning techniques for evaluating tomato ripeness. *Expert Systems with Applications*, 42(4), 1892–1905. <http://dx.doi.org/10.1016/j.eswa.2014.09.057>.
- Furukawa, Y., & Ponce, J. (2007). Accurate, dense, and robust multi-view stereopsis. In *IEEE conference on computer vision and pattern recognition 2007. CVPR '07*. (pp. 1–8). doi:[10.1109/CVPR.2007.383246](https://doi.org/10.1109/CVPR.2007.383246).
- Gollakota, A., & Srinivas, M. B. (2011). Agribot—A multipurpose agricultural robot. In *2011 annual IEEE India conference* (pp. 1–4). doi:[10.1109/INDCON.2011.6139624](https://doi.org/10.1109/INDCON.2011.6139624).
- Gongal, A., Silwal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2016). Apple crop-load estimation with over-the-row machine vision system. *Computers and Electronics in Agriculture*, 120, 26–35. <http://dx.doi.org/10.1016/j.compag.2015.10.022>.
- Guerrero, J., Guijarro, M., Montalvo, M., Romeo, J., Emmi, L., Ribeiro, A., et al. (2013). Automatic expert system based on images for accuracy crop row detection in maize fields. *Expert Systems with Applications*, 40(2), 656–664. <http://dx.doi.org/10.1016/j.eswa.2012.07.073>.
- Hartley, R. I., & Zisserman, A. (2004). *Multiple view geometry in computer vision* (2nd ed.). Cambridge University Press. ISBN: 0521540518.
- Herrera, J., Medina, S., Beleno, K., & Gualdrón, O. (2016). Design of an automated coffee selection system by means of computer vision techniques. *UIS Ingenierías*, 15(1), 7–14.
- Hunt, E. R., Daughtry, C. S. T., Mirsky, S. B., & Hively, W. D. (2014). Remote sensing with simulated unmanned aircraft imagery for precision agriculture applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(11), 4566–4571. doi:[10.1109/JSTARS.2014.2317876](https://doi.org/10.1109/JSTARS.2014.2317876).
- Ivorra, E., Sánchez, A., Camarasa, J., Diago, M., & Tardagüila, J. (2015). Assessment of grape cluster yield components based on 3d descriptors using stereo vision. *Food Control*, 50, 273–282. <http://dx.doi.org/10.1016/j.foodcont.2014.09.004>.

- Jay, S., Rabatel, G., Hadoux, X., Moura, D., & Gorretta, N. (2015). In-field crop row phenotyping from 3d modeling performed using structure from motion. *Computers and Electronics in Agriculture*, 110, 70–77. <http://dx.doi.org/10.1016/j.compag.2014.09.021>.
- Jiang, L., Koch, A., Scherer, S. A., & Zell, A. (2013). Multi-class fruit classification using RGB-D data for indoor robots. In *2013 IEEE international conference on robotics and biomimetics (ROBIO)* (pp. 587–592). doi:[10.1109/ROBIO.2013.6739523](https://doi.org/10.1109/ROBIO.2013.6739523).
- Kurtulmuş, F., & İsmail Kavdir (2014). Detecting corn tassels using computer vision and support vector machines. *Expert Systems with Applications*, 41(16), 7390–7397. <http://dx.doi.org/10.1016/j.eswa.2014.06.013>.
- Latif, R. N., Filin, S., & Eizenberg, H. (2013). Estimating plant growth parameters using an energy minimization-based stereovision model. *Computers and Electronics in Agriculture*, 98, 260–271. <http://dx.doi.org/10.1016/j.compag.2013.07.012>.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110. doi:[10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).
- Luo, L., & Chen, X. (2013). Integrating piecewise linear representation and weighted support vector machine for stock trading signal prediction. *Applied Soft Computing*, 13(2), 806–816. <https://doi.org/10.1016/j.asoc.2012.10.026>.
- Luo, L., Tang, Y., Zou, X., Ye, M., Feng, W., & Li, G. (2016). Vision-based extraction of spatial information in grape clusters for harvesting robots. *Biosystems Engineering*, 151, 90–104. <http://dx.doi.org/10.1016/j.biosystemeng.2016.08.026>.
- Minervini, M., Scharr, H., & Tsafaris, S. A. (2015). Image analysis: The new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Processing Magazine*, 32(4), 126–131. doi:[10.1109/MSP.2015.2405111](https://doi.org/10.1109/MSP.2015.2405111).
- Moonrinta, J., Chaivivatrakul, S., Dailey, M. N., & Ekpanyapong, M. (2010). Fruit detection, tracking, and 3d reconstruction for crop mapping and yield estimation. In *ICARCV* (pp. 1181–1186). IEEE.
- Moreda, G., Muñoz, M., Ruiz-Altisent, M., & Perdigones, A. (2012). Shape determination of horticultural produce using two-dimensional computer vision—A review. *Journal of Food Engineering*, 108(2), 245–261. <http://dx.doi.org/10.1016/j.jfoodeng.2011.08.011>.
- Mortensen, A. K., Lisouski, P., & Ahrendt, P. (2016). Weight prediction of broiler chickens using 3d computer vision. *Computers and Electronics in Agriculture*, 123, 319–326. <http://dx.doi.org/10.1016/j.compag.2016.03.011>.
- Nansen, C., Singh, K., Mian, A., Allison, B. J., & Simmons, C. W. (2016). Using hyperspectral imaging to characterize consistency of coffee brands and their respective roasting classes. *Journal of Food Engineering*, 190, 34–39. <http://dx.doi.org/10.1016/j.jfoodeng.2016.06.010>.
- Nielsen, M., Slaughter, D. C., & Gliever, C. (2012). Vision-based 3d peach tree reconstruction for automated blossom thinning. *IEEE Transactions on Industrial Informatics*, 8(1), 188–196. doi:[10.1109/TII.2011.2166780](https://doi.org/10.1109/TII.2011.2166780).
- Ohali, Y. A. (2011). Computer vision based date fruit grading system: Design and implementation. *Journal of King Saud University—Computer and Information Sciences*, 23(1), 29–36. <http://dx.doi.org/10.1016/j.jksuci.2010.03.003>.
- de Oliveira, E. M., Leme, D. S., Barbosa, B. H. G., Rodarte, M. P., & Pereira, R. G. F. A. (2016). A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food Engineering*, 171, 22–27. <http://dx.doi.org/10.1016/j.jfoodeng.2015.10.009>.
- Osuna, E., Freund, R., & Girosi, F. (1997). Support vector machines: Training and applications. *Technical Report*. Cambridge, MA, USA.
- Patel, H. N., Jain, R. K., & Joshi, M. V. (2012). Automatic segmentation and yield measurement of fruit using shape analysis. *International Journal of Computer Applications*, 45(7), 19–24.
- Patel, H. N., Jain, R. K., & Joshi, M. V. (2011). Fruit detection using improved multiple features based algorithm. *International Journal of Computer Applications*, 13(2), 1–5.
- Ramos, P., Prieto, F., Montoya, E., & Oliveros, C. (2017). Automatic fruit count on coffee branches using computer vision. *Computers and Electronics in Agriculture*, 137, 9–22. <http://doi.org/10.1016/j.compag.2017.03.010>.
- Ramos, P., Prieto, F., & Oliveros, C. (2016). Semiautomatic annotation system to generate ground truth in image sequence of agricultural sceneries. *6th international workshop applications of computer image analysis and spectroscopy in agriculture*, Aarhus, Denmark.
- Rodrigues, B., Soares, A., Costa, R., Baalen, J. V., Salvini, R., Silva, F., Caliari, M., Cardoso, K., Ribeiro, T., Delbem, A., Federson, F., Coelho, C., Laureano, G., & Lima, T. (2016). A feasibility cachaca type recognition using computer vision and pattern recognition. *Computers and Electronics in Agriculture*, 123, 410–414. <http://dx.doi.org/10.1016/j.compag.2016.03.020>.
- Romeo, J., Pajares, G., Montalvo, M., Guerrero, J., Guijarro, M., & de la Cruz, J. (2013). A new expert system for greenness identification in agricultural images. *Expert Systems with Applications*, 40(6), 2275–2286. <http://dx.doi.org/10.1016/j.eswa.2012.10.033>.
- Rusu, R. B., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). *IEEE international conference on robotics and automation (ICRA), Shanghai, China*.
- Sampson, D. J., Chang, Y. K., Rupasinghe, H. V., & Zaman, Q. U. (2014). A dual-view computer-vision system for volume and image texture analysis in multiple apple slices drying. *Journal of Food Engineering*, 127, 49–57. <http://dx.doi.org/10.1016/j.jfoodeng.2013.11.016>.
- Sandoval, Z., Prieto, F., & Betancur, J. (2010). Digital image processing for classification of coffee cherries. In *Electronics, robotics and automotive mechanics conference (CERMA)*, 2010 (pp. 417–421). doi:[10.1109/CERMA.2010.54](https://doi.org/10.1109/CERMA.2010.54).
- Sanz-Uribe, J. R., Ramos-Giraldo, P. J., & Oliveros-Tascon, C. E. (2008). Algorithm to identify maturation stages of coffee fruits. In *Advances in electrical and electronics engineering—IAENG special edition of the world congress on engineering and computer science 2008, WCECS '08* (pp. 167–174). doi:[10.1109/WCECS.2008.28](https://doi.org/10.1109/WCECS.2008.28).
- Snavely, N., Seitz, S. M., & Szeliski, R. (2006). Photo tourism: Exploring photo collections in 3d. *ACM Transactions on Graphics*, 25(3), 835–846. doi:[10.1145/1141911.1141964](https://doi.org/10.1145/1141911.1141964).
- Snavely, N., Seitz, S. M., & Szeliski, R. (2008). Modeling the world from internet photo collections. *International Journal of Computer Vision*, 80(2), 189–210. doi:[10.1007/s11263-007-0107-3](https://doi.org/10.1007/s11263-007-0107-3).
- Øystein Sture, Øye, E. R., Skavhaug, A., & Mathiassen, J. R. (2016). A 3d machine vision system for quality grading of atlantic salmon. *Computers and Electronics in Agriculture*, 123, 142–148. <http://dx.doi.org/10.1016/j.compag.2016.02.020>.
- Verma, U., Rossant, F., Bloch, I., Orensan, J., & Boisgontier, D. (2014). Shape-based segmentation of tomatoes for agriculture monitoring. In *International conference on pattern recognition applications and methods (ICPRAM)* (pp. 402–411).
- Virgen-Navarro, L., Herrera-López, E. J., Corona-González, R. I., Arriola-Guevara, E., & Guatemala-Morales, G. M. (2016). Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. *Expert Systems with Applications*, 54, 162–169. <http://dx.doi.org/10.1016/j.eswa.2016.01.027>.
- Zhang, B., Huang, W., Gong, L., Li, J., Zhao, C., Liu, C., et al. (2015). Computer vision detection of defective apples using automatic lightness correction and weighted (RVM) classifier. *Journal of Food Engineering*, 146, 143–151. <http://dx.doi.org/10.1016/j.jfoodeng.2014.08.024>.
- Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., et al. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review. *Food Research International*, 62, 326–343. <http://dx.doi.org/10.1016/j.foodres.2014.03.012>.
- Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334. doi:[10.1109/34.888718](https://doi.org/10.1109/34.888718).