



Research Article

On-the-Go Grapevine Yield Estimation Using Image Analysis and Boolean Model

Borja Millan ^{1,2}, **Santiago Velasco-Forero**², **Arturo Aquino**¹ and **Javier Tardaguila** ¹

¹*Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, Gobierno de La Rioja), Finca La Grajera, 26007 Logroño, La Rioja, Spain*

²*MINES ParisTech, PSL Research University, CMM-Center of Mathematical Morphology, 77300 Fontainebleau, France*

Correspondence should be addressed to Javier Tardaguila; javier.tardaguila@unirioja.es

Received 31 March 2018; Revised 3 September 2018; Accepted 9 October 2018; Published 16 December 2018

Guest Editor: Edward Sazonov

Copyright © 2018 Borja Millan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper describes a new methodology for noninvasive, objective, and automated assessment of yield in vineyards using image analysis and Boolean model. Image analysis, as an inexpensive and noninvasive procedure, has been studied for this purpose, but the effect of occlusions from the cluster or other organs of the vine has an impact that diminishes the quality of the results. To reduce the influence of the occlusions in the estimation, the number of berries was assessed using the Boolean model. To evaluate the methodology, three different datasets were studied: cluster images, manually acquired vine images, and vine images captured on-the-go using a quad. The proposed algorithm estimated the number of berries in cluster images with a root mean square error (RMSE) of 20 and a coefficient of determination (R^2) of 0.80. Vine images manually taken were evaluated, providing 310 grams of mean error and $R^2 = 0.81$. Finally, images captured using a quad equipped with artificial light and automatic camera triggering were also analysed. The estimation obtained applying the Boolean model had 610 grams of mean error per segment (three vines) and $R^2 = 0.78$. The reliability against occlusions and segmentation errors of the Boolean model makes it ideal for vineyard yield estimation. Its application greatly improved the results when compared to a simpler estimator based on the relationship between cluster area and weight.

1. Introduction

Sustainable viticulture requires continuous monitoring of the vineyard to assist the decision-making procedure and to optimize cultural practices like pruning, irrigation, and disease management. The use of noninvasive proximal sensors reduces the time and labour resources, favouring objective data acquisition. Image analysis techniques allow for fast and reliable measurements, and recent studies have aimed its use in viticulture. Application examples include canopy status assessment [1, 2] and, more recently, pruning mass determination [3]. As a noninvasive, reliable, and low-cost technology, image analysis is also a candidate for its integration in fully automated systems for vineyard monitoring [4]. These tools are key devices for the future viticulture, as they will reduce

management costs and will allow the application of more sustainable practices.

Grapevine yield estimation is encouraged by its economical relevance [5–7] and can help to optimize plant growth and to improve fruit quality [8]. Early yield estimation can be generated from the flower number per inflorescence assessed using computer vision [9]. Estimations representing final yield variability can be acquired nearby to harvest time using cluster images [10]. To improve the image quality and ease the segmentation process, some authors captured the images under controlled conditions, in the laboratory or using a specially developed chamber [11, 12]. Due to the destructive, slow, and labour-demanding nature of this process, it is hard to scale it to increase the sample points. Another approach would be the manual acquisition of images on the field [13–15], but although this method

requires less workforce, a more automatized procedure is desirable for an industrial application. Finally, modified agricultural vehicles can be used to automate the image capture of large datasets [16, 17]. This approach has to face the limitation introduced by the lack of supervision during the capture, which greatly affects image quality. The segmentation process of images acquired on the field is challenging, because of the uncontrolled scenario characteristics and the unevenness in the berry surface caused by the pruina [11]. Also, it must be noted that not all the berries in a cluster are visible due to occlusions from other berries or vegetal material from the vine. A method that has resistance to these problems (occlusions and segmentation errors) will greatly improve the prediction reliability.

The Boolean model and random set theory was developed by Matheron [18] and Serra [19]. From an image-processing viewpoint, the practical advantage of this model relies in its capability to estimate the number of particles present in an image, even when errors in the segmentation or occlusions are present. It has been mainly used for modeling material structure characteristics [18–20], for estimating the spatial distribution of bacterial colonies in cheese [21] or the number of cells in a cluster [22]. However, to the best of our knowledge, it has not been used in agriculture for berry and yield estimation.

This study aims at grapevine yield assessment using image analysis and the Boolean model. This solution was tested under three different scenarios: cluster images, manually acquired vine images and on-the-go captured vine images using a quad at a speed comparable to other agricultural equipment used in vineyard management.

2. Materials and Methods

2.1. Image Acquisition. The experiments were conducted in September 2014 and 2015 in a commercial vineyard located in Falces (latitude $42^{\circ}27'45.96''$, longitude $1^{\circ}48'13.42''$, and altitude 325 m; Navarra, Spain). The vines were growing in a vertical shoot-positioning system, with north-south row orientation at 2×1 m disposition. Five different grapevine (*Vitis vinifera* L.) varieties were used. The choice of a multi-varietal experiment was made to increase the variability in yield components (number of berries per cluster, mean berry weight, and mean cluster weight). The six first basal leaves of the selected vines were manually removed after berry set.

Three different sets of images were captured:

(i) *Manually Acquired Cluster Images.* A set of 45 cluster images from four different grapevine varieties (Cabernet Sauvignon, Garnacha, Syrah, and Tempranillo) was captured in the field on the 4th of September 2014 and harvested the next day. The images were taken using a Nikon D5300 digital reflex camera (Nikon Corp., Tokyo, Japan) equipped with a Sigma 50 mm F2.8 macro (Sigma Corp., Kanagawa, Japan). RGB images were captured with uncontrolled illumination using an orange cardboard as background and saved at a resolution of 24 Mpx (6000×4000 pixels), 8 bits per channel.

(ii) *Manually Acquired Vine Images.* A set of 45 images from four different grapevine varieties (Cabernet Sauvignon, Garnacha, Syrah, and Touriga Nacional) were taken in the field at the same date as the cluster images. RGB images were captured using a Nikon D5300 digital reflex camera equipped with a Nikon AF-S DX 10 NIKKOR 18–55 mm f/3.5–5.6G VR lens. The acquisition was realized under uncontrolled illumination using a white panel as background and a tripod to maintain a capturing distance around 120 cm. The images were saved at a resolution of 24 Mpx (6000×4000 pixels), 8 bits per channel.

(iii) *On-the-Go Acquired Vine Images.* 64 images from three different grapevine varieties (Cabernet Sauvignon, Syrah, and Tempranillo) were captured at night time on the 9th of September of 2015 using a quad (Trail Boss 330, Polarix Industries, Minnesota, USA) at a speed around 7 km/h. Clusters were harvested and weighted the next day. The vehicle was equipped with a Sony alpha 7-II digital mirrorless camera (Sony Corp., Tokyo, Japan). The camera had a Vario-Tessar FE 24–70 mm lens. RGB images were saved at a resolution of 24 Mpx (6000×3376 pixels), 8 bits per channel, and manually combined to obtain 28 sections composed of three vines. A 900 LED Bestlight panel and two Travor spash IS-L8 LED lights were used for scene illumination. The quad was fitted with an adjustable mechanical structure that allowed for different height and depth fixation to adapt to the vine configuration (Figure 1(a)). The structure also provided protection against branch impact and allowed the attachment of the illumination equipment. The camera was triggered by a custom-built controller using an Arduino MEGA (Arduino LLC, Italy). The controller generated the shooting signal based on the information received from an inductive sensor attached to the rear axle. This sensor produced 3 pulses per rear-axle revolution, thus allowing the camera to obtain images with an approximate 40% of superposition rate.

2.2. Boolean Model for Berry Number Estimation. Boolean random closed sets [18] have been widely used for particle number estimation in images [23]. The main strength of this model is its robustness against partly covered objects and errors in segmentation.

The model can be applied if the structure is Boolean [19] but is not limited to this case due to the central limit theorem [22]. To estimate the number of objects in a region Z , the following formulation can be used:

$$\text{Number of objects} = -\frac{a_z}{a'} \log q, \quad (1)$$

where a_z is the area under study (ROI), a' is the mean area of the object, and q is the ROI porosity:

$$q = \frac{\# \text{pixels of the ROI} - \# \text{pixels of the segmented area}}{\# \text{pixels of the ROI}}. \quad (2)$$



(a)



(b)

FIGURE 1: On-the-go capturing system: (a) modified quad with automatic camera triggering, LED illumination, and structure for easy position adjustment; (b) example of an on-the-go captured vine image.

The Boolean model can be directly used for berry number estimation, but the ROI must be defined so that the concentration of particles is similar on it. In the case of vine images, particle (berries) concentration is limited to portions in the image (clusters), so a ROI not corresponding to all the image areas must be selected for proper porosity calculation. The ROI was automatically obtained by applying a morphological opening [24] (morphological erosion followed by dilation) for all the segmented clusters using a circular kernel of the same radius of the mean berry size.

To evaluate the prediction capabilities of the Boolean model, four tests were conducted (each one composed of 100 simulations). The tests were performed by using MATLAB (R2010b, MathWorks, Natick, MA, USA) to generate synthetic images containing randomly placed particles. First, the test compared the error of the Boolean model for 50 randomly positioned particles of a radius equal to 5 in an image composed of 100×100 pixels. Next, random variation on the radius of each particle (up to 30%) was used to generate a new set of simulations. The same tests were also

performed for 500 particles in an identical area for fixed and variable radii.

For comparison purposes, a naïve estimator was also defined as follows:

$$\text{Number of objects} = \frac{\text{\#pixels segmented as cluster}}{\text{\#pixels corresponding to mean object area}}. \quad (3)$$

This estimator only takes into account the relationship between the total area of the particles (cluster/s) and the mean particle area (berry). This formulation is similar to other approaches used in the bibliography [13, 14].

2.3. Image Analysis Algorithm for Berry Number Estimation.

The three previously described sets of images (manually acquired cluster images, vine images acquired manually, and vine images captured on-the-go) were analysed using similar approaches: first the clusters were segmented, then the Boolean model was applied to estimate berry number.

The Boolean model used for berry number estimation only requires as inputs an average radius of the particle (berry) and the area of the segmented regions or, more specifically for this application, the segmented cluster (the procedure is described in Section 2.4). To determine the mean berry radius, different approaches were used depending on the type of the images to be analysed:

(i) *Cluster Images.* The berry radius was manually extracted (an operator selected two points at the equatorial line of a berry). This process was repeated for every image because of the measurement variation depending on the distance between the camera and the cluster.

(ii) *Vine Images.* An average radius was set (manually extracted in one image as in the cluster dataset) and applied to all images from the same grapevine variety.

The algorithm for image analysis was implemented in MATLAB and process batches of images in a fully automated way. The cluster segmentation procedure was based on a Mahalanobis distance classifier and is defined in the following section. For the on-the-go images, misclassification between the pixels corresponding to clusters and the metal wires used for the vine support were observed. An additional filtering step is described in Section 2.5 and the benchmarking and validation process of the classification in Section 2.6.

2.4. Cluster Segmentation. In our approach, cluster segmentation was the first step that must be implemented to obtain the yield estimation. Every pixel in the image was characterized as a six-dimension vector denoted by \vec{u} , using two different colour models: red-green-blue (RGB) and the hue-saturation-value (HSV) representation. HSV and RGB are different colour spaces, with RGB being closer to physical image acquisition and HSV having the advantage of separating the colour and illumination information (*croma* and *luma*, respectively), thus making colour information invariant to nonuniform illumination. We note that the hue

component in HSV colour space is an angular variable with values between 0° to 360° . In this case, the “beginning” coincides with the “end,” i.e., 0° has the same meaning as 360° , and methods to measure distances between any two points should take careful note of that. Taking advantage of the blue colour (with the hue value centred at 240°) is the dominant coloration for the clusters; the H component of the vector definition of the pixel was calculated using a modification of the standard definition of the HSV to RGB conversion, assigning the blue colour to the centre of the interval (128).

Colour-based segmentation was performed using the Mahalanobis distance [25] on each pixel. The Mahalanobis distance between two vectors (\vec{u} , \vec{v}) with the same distribution and covariance matrix Σ is defined as

$$d_m(\vec{u}, \vec{v}) = \sqrt{(\vec{u} - \vec{v})^T \Sigma^{-1} (\vec{u} - \vec{v})}. \quad (4)$$

In this application, \vec{u} is an image pixel, \vec{v} represents the reference pixels (seeds) for each class to be identified, and the covariance matrix (Σ) is calculated as follows:

$$\Sigma = \begin{bmatrix} \sigma_{RR} & \sigma_{RG} & \sigma_{RB} & \sigma_{RH} & \sigma_{RS} & \sigma_{RV} \\ \sigma_{GR} & \sigma_{GG} & \sigma_{GB} & \sigma_{GH} & \sigma_{GS} & \sigma_{GV} \\ \sigma_{BR} & \sigma_{BG} & \sigma_{BB} & \sigma_{BH} & \sigma_{BS} & \sigma_{BV} \\ \sigma_{HR} & \sigma_{HG} & \sigma_{HB} & \sigma_{HH} & \sigma_{HS} & \sigma_{HV} \\ \sigma_{SR} & \sigma_{SG} & \sigma_{SB} & \sigma_{SH} & \sigma_{SS} & \sigma_{SV} \\ \sigma_{VR} & \sigma_{VG} & \sigma_{VB} & \sigma_{VH} & \sigma_{VS} & \sigma_{VV} \end{bmatrix}, \quad (5)$$

and the covariance matrix elements can be calculated as

$$\sigma_{RG} = \sigma_{GR} = \frac{1}{n-1} \sum_{i=0}^n (Ri - \bar{R})(Gi - \bar{G}), \quad (6)$$

where $Ri, Gi \dots Si, Vi$ are the values of the i th match, and $\bar{R}, \bar{G} \dots \bar{S}, \bar{V}$ are the mean values in the image to be processed.

The seeds used as reference for each set were manually selected from a different image for each variety (as there exists differences in the cluster colorations between them). The number of classes depends on the type of images: three for cluster images (grape, rachis, and background) and six for vine images, including manual and on-the-go (leaf, background, trunk, shoot, cable, and cluster) corresponding to the different elements present in the scene.

The Mahalanobis distance considers not only the distance to the centroid of the sample pixels but also the fact that the variances in each direction are different, as well as the covariance between variables [13]. The use of Mahalanobis distance in colour images standardizes the influence of the distribution of each feature, taking into account the correlation between each pair of terms [26].

After the distance was calculated for each pixel, it was converted to an occurrence probability to obtain a membership probability map (MPM) [27] using the Boltzmann

distribution [28]. The Boltzmann distribution is a probability distribution that gives the probability for a system to be in a certain state as a function of that state's energy and temperature. For this application, the Mahalanobis distance is used as the energy of the system. The formula that describes the probability for a given pixel in the coordinates (x, y) for a class i is

$$\text{MPMcolour}_{x,y,i} = \frac{e^{(-d_i(x,y)/kT)}}{\sum_{j=1}^M e^{(-d_j(x,y)/kT)}}, \quad (7)$$

where $d_k(x, y)$ corresponds to the Mahalanobis distance of the pixel located at the (x, y) coordinates and its reference value for the class i . kT is a constant that in the original formulation of the Boltzmann distribution corresponds to the multiplication of the Boltzmann constant and the thermodynamic temperature; for this application, it was set to 10. The denominator guarantees that all the probabilities are normalized, and the sum of the M class probabilities is equal to 1 for every pixel of the MPM.

Cluster segmentation in both the cluster and manually acquired vine images was performed using the maximum pertinence to cluster class from MPMcolour. Additional MPMs were used for on-the-go acquired vine images as described in the following section.

2.5. Additional Filters for Cluster Segmentation for On-the-Go Captured Images. The MPMcolour information can be combined with other MPMs generated using morphological data to aid in the segmentation process. Hence, three additional MPMs were defined to improve the cluster segmentation for the on-the-go images:

(i) *Cluster Proximity MPM (MPMcluster_proximity)*. As a preprocess, a pyramidal decomposition with step values similar to berry size (5 by 5 pixels) was conducted on the pixels that had the maximum likelihood of being part of the cluster class (from MPMcolour). Next, a Gaussian filter with a standard deviation set to 3 times the average grape radius was used to expand the cluster pertinence probabilities. By doing this, pixels in the neighbourhood of the previously filtered cluster candidates increase their possibility of pertinence to the cluster class. Also, isolated pixels that were not close to clusters will decrease its cluster class membership probabilities.

(ii) *Shape-Angle MPM (MPMshape)*. Due to the misclassifications between the cluster and cable class, and taking advantage of the well-defined shape characteristics of the cable, a filter was defined. As a first step, all the connected components (CCs) corresponding to the cable and cluster class (from MPMcolour) were extracted, and all the CCs whose areas were lower than the size of the mean berry were eliminated, which is to say

$$\text{Area}(\text{CC}_i) > r_{\text{berry}}^2 * \pi, \quad (8)$$

where $\text{Area}(\text{CC}_i)$ corresponds to the number of pixels of the i th CC and r_{berry} is the mean berry radius.

Then, the length and orientation of the major and minor axes for every remaining CC were determined. The shape relation was calculated as the division of the major by the minor axis length:

$$\text{Shape relation} = \frac{\text{major axis length}}{\text{minor axis length}}. \quad (9)$$

Combining these two descriptors, a new MPM was generated as follows:

$$\begin{aligned} \text{MPMshape} &= (1 - \text{Shape relation}) \\ &* \left(1 - \text{abs} \left(\frac{\text{Major axis orientation}}{90} \right) \right). \end{aligned} \quad (10)$$

(iii) *Linear Occurrence Zone (MPMlinear_occurrence_zone)*. As the cables along the vines were usually placed at fixed heights, there were horizontal sections in the images where the probability of a pixel to belong to the cable class was higher. To determine these zones independently from the camera or cable position in the image, an automatic detector was built. The CCs most likely to correspond to the cable class were used. For this purpose, all the CCs with an orientation around $\pm 30^\circ$ and with a shape relation lower than 0.5 were chosen to generate a binary image (Cable_b). From this, an accumulator for each row based on the sum of the number of pixels selected as the cable class was generated using the following expression:

$$\text{Accumulator}_y = \sum_{x=1}^{\text{number of columns}} \text{Cable}_b(x, y), \quad (11)$$

being

$$\text{Cable}_b(x, y) = \begin{cases} 1 & \text{if } I(x, y) \in \text{Filtered CCcable} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

for every column x and row y in the image I .

This accumulator holds the number of pixels of the filtered cable candidates for each row; as an example, the accumulator of Figure 2(a) is shown in Figure 2(b). The next step is to apply a Gaussian filtering, thus allowing for some flexibility in the angle of the cables and not limiting it to the horizontal case. The result of the smoothing is presented in Figure 2(c). The final MPM of the linear occurrence zone is obtained by expanding the smoothed accumulator to all the rows of the image. Figure 2(d) shows the MPM (in grayscale) along with the filtered CCs that were overprinted in red colour for illustration purposes.

The final MPM used to classify the pixels as clusters for the on-the-go images was obtained by the element-wise multiplication of the four previously calculated MPMs: MPMcolour, MPMcluster_proximity, MPMshape, and MP

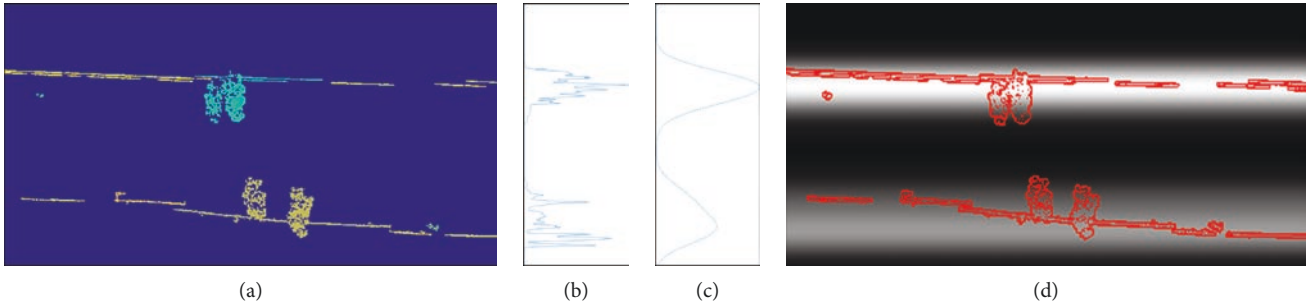


FIGURE 2: Steps for the generation of the `MPMlinear_occurrence_zone` aimed for reduction of misclassification between the cluster and cable class during segmentation: (a) objects segmented as cable candidates from automatically taken images using a quad; (b) accumulator of the number of pixels of cable candidates for each row; (c) smoothed accumulator; (d) membership probability map (MPM) for cable occurrence based on the position of the cable (in grayscale) with the original candidates overlaid in red.

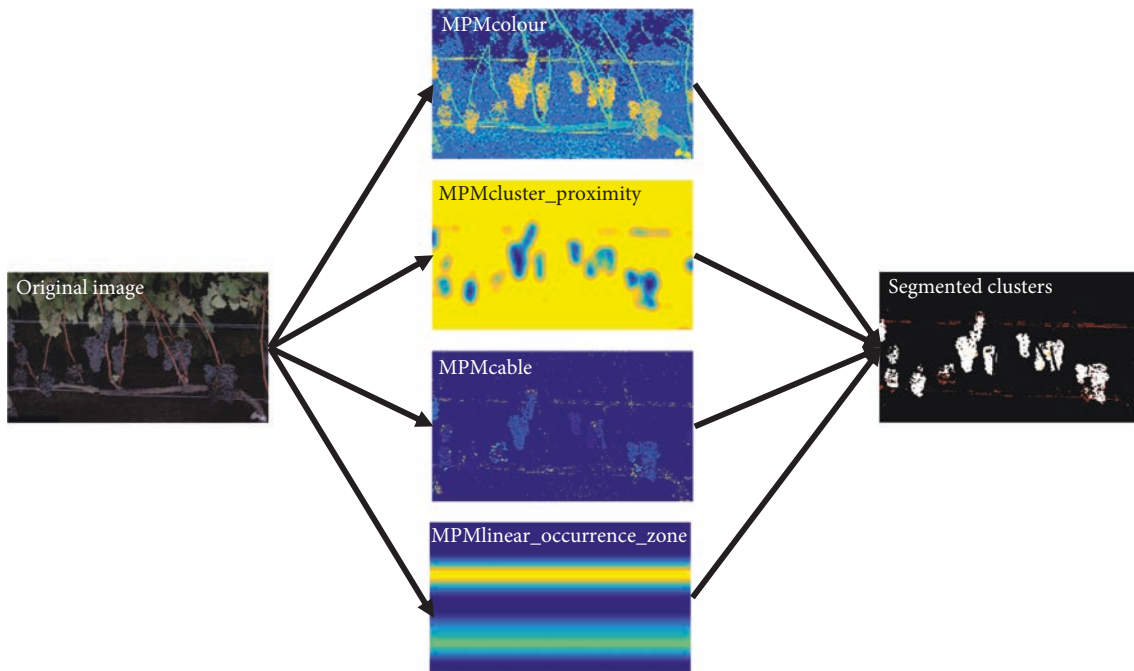


FIGURE 3: Cluster segmentation process for on-the-go captured image. The original image is used to obtain four MPMs (membership probability maps): `MPMcolour`, `MPMcluster_proximity`, `MPM cable`, and `MPMlinear_occurrence_zone`. These MPMs were combined to classify the pixels corresponding to clusters for the on-the-go captured images.

`Mlinear_occurrence_zone`. The process is represented in Figure 3.

2.6. Validation. To evaluate the developed algorithms, yield estimation has to be compared with real data. Also, due to the especial characteristics of the on-the-go images, the segmentation was ranked before and after the filtering MPMs were applied.

The ground truth for every data set was obtained as follows:

(i) *Manually Acquired Cluster Images.* All the photographed clusters were picked and introduced into pretagged plastic bags to allow their conservation during their transport to the laboratory. Then, they were destemmed, and the berries

were detached, counted, and weighted. The number of berries per cluster and their weight was used to obtain the average berry weight.

(ii) *Manually Acquired Vine Images.* After the image capturing process, all the vines were harvested, and the clusters were weighted together to obtain the final yield per vine.

(iii) *On-the-Go Acquired Vine Images.* After image acquisition, the sections composed of three vines were harvested and the clusters weighted together to obtain the final yield per section.

To evaluate the segmentation process of the on-the-go images and the improvements of the multi-MPM filtering, it is necessary to obtain a ground truth. An application

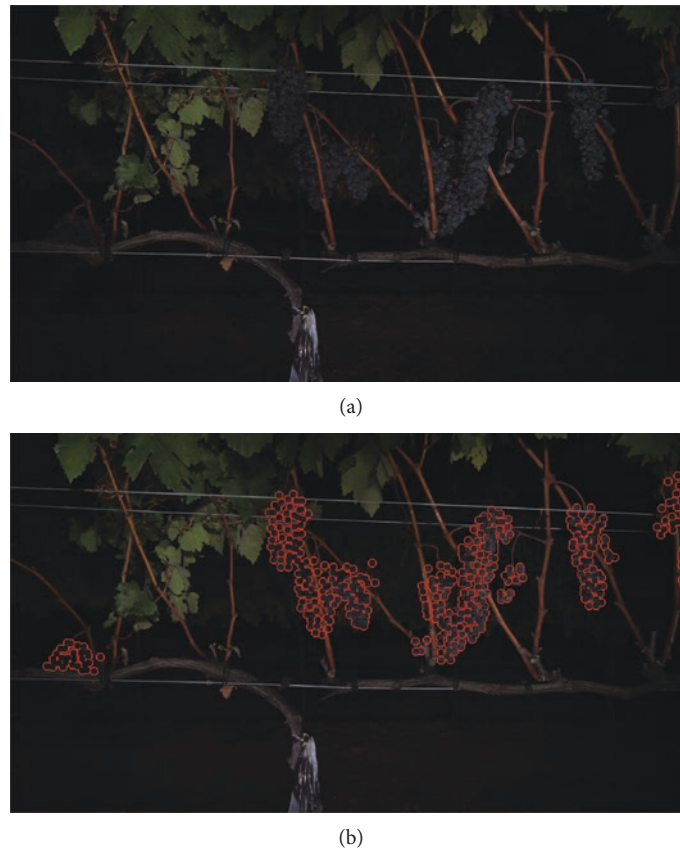


FIGURE 4: Ground truth generation for segmentation performance benchmarking: (a) example image of a vine captured on-the-go of cv. Tempranillo; (b) ground truth mask of the clusters. The berries were manually selected using a custom-built application.

allowing to manually select the berry centres was built to generate a mask representing the area occupied by the clusters in the image. An example of an on-the-go automatically captured photograph is shown in Figure 4(a), and the manually selected pixel classification for benchmarking this image is shown in Figure 4(b).

The mask generated using this application was used to obtain the following metrics:

(i) *True Positive (TP)*. A pixel classified as corresponding to a cluster that actually matches a cluster pixel in the manually selected mask.

(ii) *False Positive (FP)*. A pixel classified as corresponding to a cluster that does not match a cluster pixel in the manually selected mask.

(iii) *False Negative (FN)*. A pixel that was automatically classified as not corresponding to a cluster but actually corresponding to a cluster in the mask.

Finally, the Recall and Precision metrics were used for evaluating the quality of each analysed image as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (13)$$

where Recall provides the percentage of actual cluster pixels detected;

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (14)$$

where Precision indicates the percentage of pixels correctly assessed.

3. Results and Discussion

3.1. Evaluation of the Occlusion Robustness of the Boolean Model. As described in Section 2.2, four tests were performed to evaluate the occlusion robustness of the Boolean model and to compare its results to those generated by the naïve estimator. Figures 5(a) and 5(b) show the simulations corresponding to 50 particles of fixed and variable radii, respectively. As can be checked in Table 1, the error rates for both estimators were low and similar but with slight improvement for the case of the naïve estimator. For the third and fourth experiments, the number of particles was 10-fold higher, making particle occlusion more likely to occur under these conditions (Figures 5(c) and 5(d)). The Boolean model estimates the number of particles with an error rate similar to the low occlusion case. On the contrary, the error yielded by the naïve estimator rose to 25% for fixed and variable

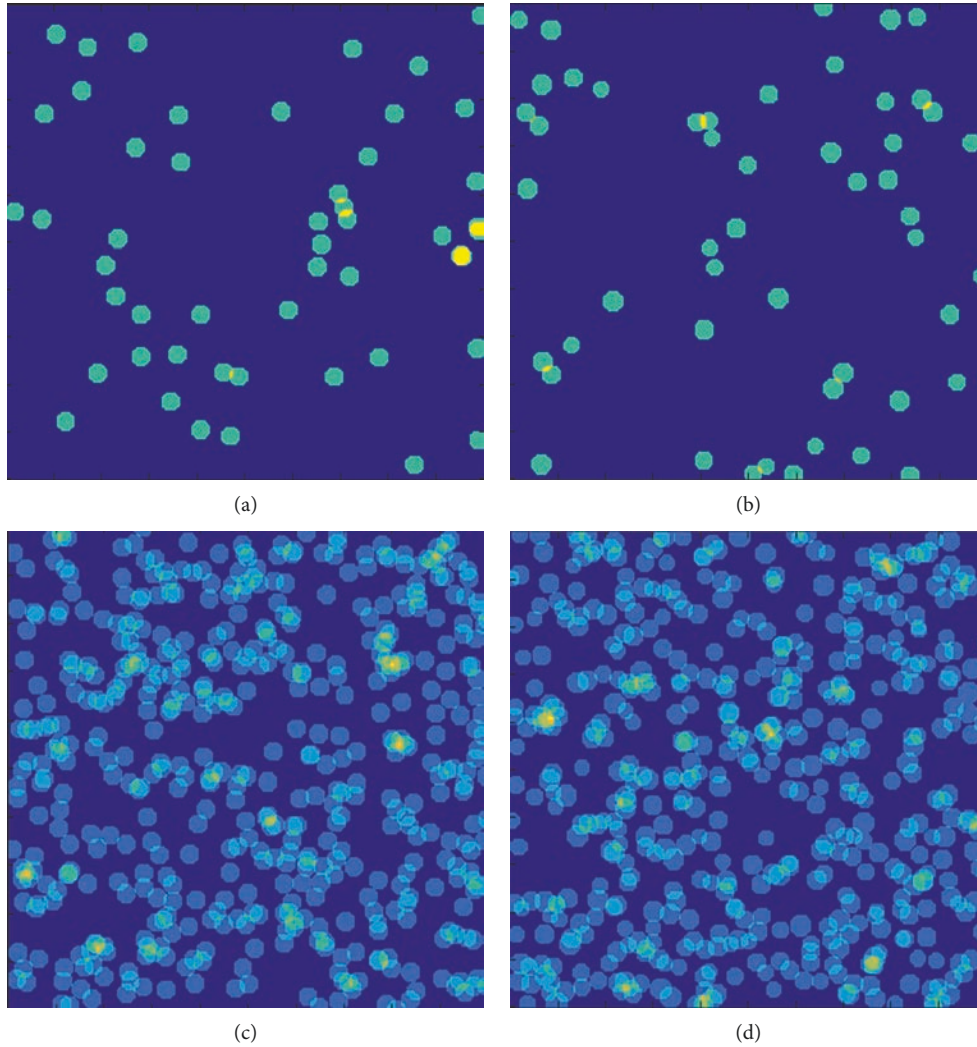


FIGURE 5: Simulation example of a random distribution of particles in a 100×100 pixel area: (a) 50 particles of radius = 5, (b) 50 particles with a random variation in the radius up to 30%, (c) 500 particles with radius = 5, and (d) 500 particles with a random variation in the radius of the particle up to 30%.

TABLE 1: Results for the estimation error of the number of particles for randomly generated simulations of 50 and 500 particles, with and without variation in radius, for the naïve estimator and the Boolean model.

	50 particles		500 particles	
	Rad = 5	Rad = $5 \pm 15\%$	Rad = 5	Rad = $5 \pm 15\%$
Naïve estimator*	1.7%	2.3%	24.9%	25.1%
Boolean model*	2.4%	2.7%	2.1%	2.1%

The presented error rates were standardized to the total particle number.
*Results after 100 iterations.

radii. These findings are coincident to the ones obtained by Angulo [22] for the number of cell cluster estimation in fluorescence marked cell images, where the number of nuclei obtained by the Boolean model is more robust than a simple ratio of surfaces (equivalent to the naïve estimator). Some

approaches had been studied for evaluating berry occlusions within a cluster. Nuske et al. [16] tested the relationship between total berry count, visible berry count, and 3D models from 2D images, but the results showed no improvement on partially occluded berry assessment. As showed in the simulations, the use of the Boolean model would improve the berry number estimation robustness.

3.2. Evaluation of the Berry Number per Cluster Estimation. An example image of a cluster, corresponding to the Cabernet Sauvignon variety, is shown in Figure 6(a). The uncontrolled conditions during the capturing process explains the excess of illumination in the berries that are placed at the right side of the image, which received direct sun illumination, in contrast to the rest of the cluster that had indirect lighting. Due to the image characteristics, segmentation errors occurred affecting the area finally segmented (Figure 6(b)). Results obtained after applying the estimation models (Boolean and naïve) are shown in

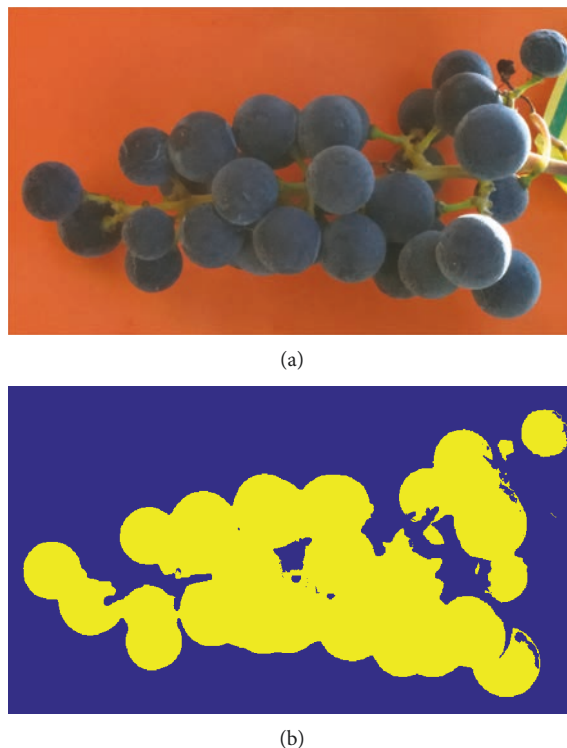


FIGURE 6: Segmentation of manually taken cluster images: (a) Example image of a cluster of cv. Cabernet Sauvignon captured under field conditions with an orange cardboard as background; (b) segmented image of the cluster using the Mahalanobis distance on six dimensions (i.e., using RGB and HSV representations).

TABLE 2: Results obtained for the estimation of the berry number per cluster using the naïve estimator and the Boolean model in manually acquired cluster images of four different grapevine varieties.

Grapevine variety	Manual counting		Naïve estimator			Boolean model		
	Mean berry number	Number of clusters	Mean berry number	RMSE	R^2	Mean berry number	RMSE	R^2
Cabernet Sauvignon	68.4	15	26.7	48.8	0.34*	79.1	14.8	0.86***
Grenache	53.8	12	18.2	37.5	0.72***	44.1	23.3	0.54**
Syrah	79.0	11	32.3	50.7	0.72***	92.5	21.2	0.69**
Tempranillo	141.4	7	43.9	107.6	0.81**	136.6	22.0	0.66*
Total	85.7	45	30.3	60.0	0.71***	88.1	20.1	0.80***

Manual counting refers to the berry number obtained by manually destemming the cluster in the laboratory. The naïve and Boolean estimation was generated based on the analysis of the cluster images manually taken under field conditions. Asterisks represent statistical significance: * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

Table 2 and in Figure 7. Table 2 describes the results obtained per variety using the two models: the naïve estimator and the Boolean model, including the ground truth generated by manually destemming the clusters. Figure 7 compares the results analysing all the images together ($n = 45$), including the 4 varieties, using the naïve estimator and the Boolean model. The naïve estimator failed to provide a good prediction, with a global RMSE = 60 (Table 2) vs RMSE = 20.1 obtained using the Boolean model. It can be understood by observing Figure 7, as the naïve estimation slope was not close to 1 and its prediction interval of 95% (represented in dotted lines) does not surround the 1 : 1 line, greatly affecting to the estimation precision. This contrasted with the results

obtained from the Boolean model, whose slope was 0.93, and the prediction lines are almost in parallel with the 1 : 1 line, demonstrating its prediction capabilities.

Table 2 shows that the results obtained using the naïve estimator were very variable upon the grapevine variety. This was caused by the occlusions (more likely to occur in more compact varieties) and errors in the segmentation. On the other hand, the results obtained with the Boolean model were more homogenous, minimizing differences between varieties and improving the results when all of them were examined together. This homogeneity suggests that this method is more generalizable, although more extensive studies must be conducted to prove this premise.

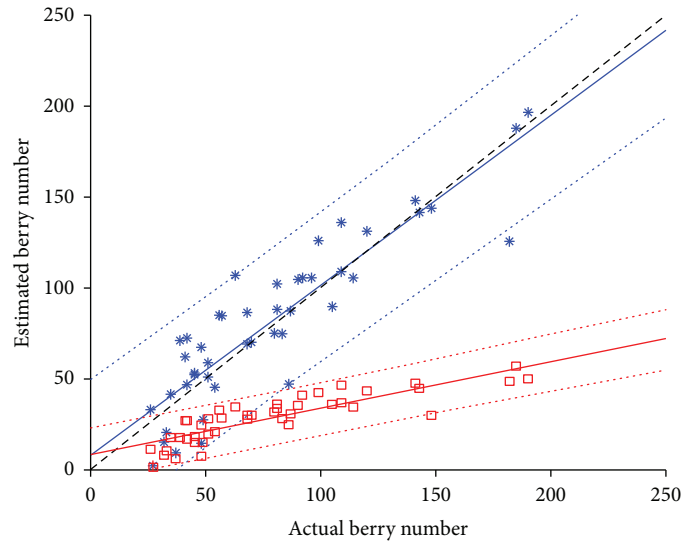


FIGURE 7: Berry number per cluster estimation using manually captured images ($n = 45$) and a naïve estimator represented in red squares ($y = 0.25x + 8.4$) and a Boolean model represented in blue stars ($y = 0.93x + 8.1$). The dashed line corresponds to 1:1, and dotted lines relate to 95% prediction intervals.

TABLE 3: Comparison of the measured coefficient of determination (R^2) for the estimation of berry number per cluster using image analysis for different varieties in other published studies (under different capturing conditions) and in this work using the Boolean model.

	Tempranillo	Grenache	Syrah	Cabernet Sauvignon	Cabernet Sauvignon and Syrah	Capturing conditions
Diago et al. [11]	0.84	0.69	—	0.62	—	Laboratory
Herrero-Huerta et al. [15]	0.78	—	—	—	—	Field
Liu et al. [12]	—	—	—	—	0.85	Laboratory
The present work	0.66	0.54	0.69	0.86	0.79	Field

The results obtained are comparable to others in the bibliography. The outcomes obtained by Diago et al. [11] are similar (Table 3), but it must be noted that their methodology is not applicable under field conditions. The procedure requires collecting the clusters and taking the images in a chamber with controlled lighting and background. Apart from that, the algorithm is more complex, requiring the segmentation of the image, edge detection, circle detection, and filtering. On the other hand, the presented method only requires the segmentation and mean berry radius for the berry number estimation. Herrero-Huerta et al. [15] developed a system for berry number assessment from images taken in the field. This procedure relies on a 3D structure reconstruction from at least 5 images with high overlapping (80–90%). Their findings (Table 3) are very similar to the ones detailed in this publication but without the need of multiple image acquisitions per cluster. Finally, Liu et al. [12] proposed a similar methodology using 3D models extracted from images captured in a laboratory under controlled conditions. They presented their results combining Cabernet Sauvignon and Syrah clusters; these figures are also included in Table 3. Results are similar to those obtained by the Boolean model but without the constraint of taking the images in the laboratory. It must be noted that the experiments conducted under laboratory

conditions are destructive and labour demanding, and thus, it is not easy to expand the sampling rate for an industrial application.

3.3. Evaluation of the Yield Estimation from Manually Captured Vine Images. An example image of a vine of cv. Cabernet Sauvignon can be observed in Figure 8(a). The image segmentation was carried out using the described Mahalanobis classifier (Section 2.4), and the result of the segmentation can be observed in Figure 8(b). Even when the overall classification quality was good, some errors were observed, especially with parts of the trunk being classified as clusters. This greatly affected the performance of the naïve estimator (Table 4), providing a RMSE of 777.2 g when all the varieties were studied together ($n = 45$). On the contrary, the Boolean model offered more robustness against errors in the segmentation and occlusions. Indeed, the RMSE for yield estimation was 310.2 g when all the varieties were studied as a whole, and also, performance for each grapevine variety was higher for the Boolean model than for the naïve estimator. R^2 values showed less difference between the two models. However, looking at Figure 9, it is clear that the naïve model did not offer a correct estimation (the slope is far from 1 and the prediction interval does not surround the 1:1 line), even when providing appropriate R^2 values.



FIGURE 8: Cluster segmentation on manually captured vine images: (a) image of a vine cv. Cabernet Sauvignon captured under uncontrolled illumination conditions with a digital camera fixed on a tripod and using a white panel as background; (b) segmentation result using the Mahalanobis distance classifier on six dimensions (i.e., using RGB and HSV representations).

TABLE 4: Results obtained for the yield estimation per vine based on manually captured grapevine images using the naïve estimator and the Boolean model.

Grapevine variety	Manual harvest		Naïve estimator			Boolean model		
	Mean yield (g)	Number of vines	Mean yield (g)	R^2	RMSE (g)	Mean yield (g)	R^2	RMSE (g)
Cabernet Sauvignon	1311	12	736	0.82***	661.0	1386	0.85***	214.2
Grenache	1750	12	938	0.87***	993.9	1816	0.89***	320.6
Syrah	1231	11	509	0.93***	673.8	966	0.88***	263.7
Touriga Nacional	1249	10	622	0.55*	713.7	1202	0.45*	421.9
Total	1431	45	728	0.82***	777.2	1389	0.81***	310.2

Manual harvest refers to the weight of all the clusters corresponding to each vine. Asterisks represent statistical significance: * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

Dunn and Martin [14] analysed the prediction potential of the segmentation of Cabernet Sauvignon grapevines. They obtained a $R^2 = 0.85$ for the relation of normalized cluster area on a section of 1 m by 1 m. It should be pointed out that their measured R^2 did not correspond to the validation of a model, but it was calculated on the calibration set. Nevertheless, this value is similar to those obtained for the validation of the models presented in this study without the need of a hanging frame that was used to extract the ROI. Diago et al. [13] used the number of pixels segmented as the cluster class to generate a linear model for yield estimation. The prediction produced $R^2 = 0.73$ and $RMSE = 749$ g. This approach is similar to the use of the naïve model, and the obtained results are equivalent to the ones produced by this estimator but sensibly surpassed by the performance of the Boolean model ($R^2 = 0.81$ and $RMSE = 310$ g).

3.4. Evaluation of the Yield Estimation from On-the-Go Captured Vine Images. The images were captured using the modified quad shown in Figure 1(a). The setup allowed for image capture at a speed of 7 km/h, being similar to the operation speed of other agricultural vehicles. The continuous movement of the vehicle, the vibrations induced by the rough terrain, and the explosion motor did not produce motion

blur in the images due to camera automatic stabilization and precise camera parametrization (Figure 1(b)). Errors were encountered in the classification process, with cross interference between the cluster and the cable class (representing the metal wire used for trellising the vines to a vertical shoot positioning system). To evaluate the convenience of the multi-MPM-filtering approach (described in Section 2.5), the segmentation was quantified using manually classified images as ground truth. The differences in the results when multiple MPMs were applied are not remarkable in terms of Recall but are notable for the Precision (Table 5). This demonstrated that false positives were correctly eliminated during the filtering, with little loss of true positives. The relative low values of Recall (Table 5) can be explained by the difficulty in pixel discrimination because of the lack of uniformity in the illumination. Figure 4(b) shows the regions manually segmented as clusters. As can be confirmed, these regions were hardly distinguishable even by manual evaluation. An illumination improvement might enhance the segmentation process and thus Recall.

The problems during the segmentation clearly affected the performance of the naïve estimator (Table 6), whose $RMSE = 2472$ g, when all the varieties were studied together ($n = 28$), resulted in a lack of its practical application, even

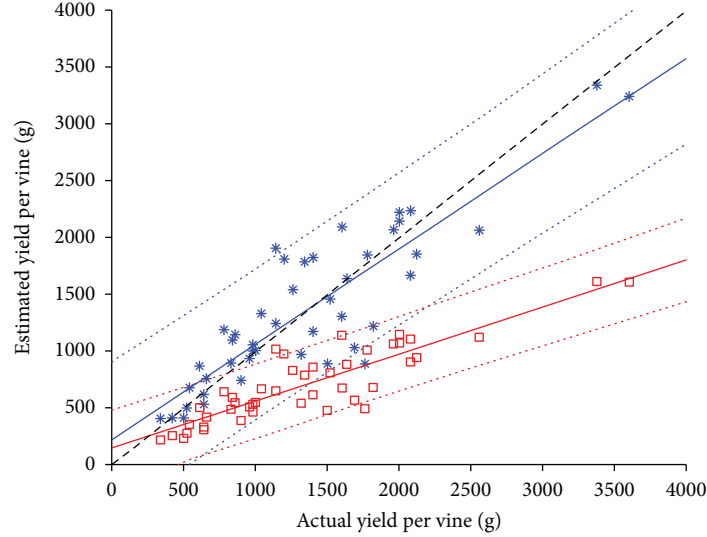


FIGURE 9: Yield per vine estimation using manually captured images ($n = 45$) and a naïve estimator represented in red squares ($y = 0.42x + 133.0$) and a Boolean model represented in blue stars ($y = 0.86x + 195.2$). The dashed line corresponds to 1:1, and dotted lines relate to 95% prediction intervals.

TABLE 5: Benchmark of the segmentation of clusters in images taken automatically on-the-go with and without applying filtering (cluster proximity, shape-angle, and linear occurrence zone).

Grapevine variety	Without filtering		With filtering	
	Recall	Precision	Recall	Precision
Cabernet Sauvignon	0.54	0.73	0.56	0.82
Syrah	0.64	0.68	0.57	0.80
Tempranillo	0.58	0.69	0.55	0.73
Total	0.58	0.71	0.56	0.79

The performance of the segmentation was tested against manually segmented images.

when the coefficient of determination was acceptable ($R^2 = 0.71$). This represents the same scenario as in the cluster and manually taken images: the naïve estimator did not compensate for the occlusions and errors in the segmentation, and the prediction interval does not surround the 1:1 line in all the intervals (Figure 10). On the other hand, the Boolean model was capable of correctly estimating the yield, offering $RMSE = 610.1$ g. It must be noted that the estimation refers to segments composed by three vines, so this value represents an improvement if it is compared to the manually captured images that yielded a $RMSE = 374.2$ g for one isolated vine.

Similar to this work, Font et al. [17] used a quad equipped with cameras and artificial illumination to capture 25 cluster images (not the entire grapevine) at night time. Then, they estimated cluster weight from its segmented area. The prediction had 16% of error when all the varieties were analysed together. In comparison, the results obtained for the Boolean model had 15.6% of error using images framing three vines instead of cluster images (the mean cluster number per section was 47). In another recent article, Nuske et al. [16]

also used a quad with artificial lighting for image capturing of grapevines. The collected images were analysed to identify visible berries to estimate yield. This setup allowed assessing yield with a $R^2 = 0.73$ for the best datasheet, being comparable to the results given by the naïve estimator ($R^2 = 0.71$), which also bases its estimation on the visible berries. They also tried to boost the yield estimation thru an evaluation of the self-occlusion of berries using 3D models of berries (ellipsoid 3D model) and clusters (convex hull 3D model). The results showed that the proposed correction models did not improve the overall estimation. In contrast to this, the Boolean estimator, which also compensates for partially occluded berries, generated better results ($R^2 = 0.78$).

4. Conclusions

This work presented a new method for accurate, nondestructive, and in-field grapevine yield estimation by using computer vision. Yield information is very valuable for viticulturists and grape growers, allowing them to take decisions prior to harvest based on objective measurements. A novel use of Boolean models has been assessed over three different data sets: images of isolated clusters, manually captured images of grapevines, and on-the-go captured images of grapevines using a modified quad at night time.

The use of Boolean models allowed to overcome two of the major difficulties in visual yield estimation: this technique is robust against segmentation errors and partial occlusions, situations that are usual in the case of images taken under field conditions. It provided more precision, using not only a model that is simpler than other previous proposals but also less complex image analysis techniques. The capacity to estimate the visible berry number and the

TABLE 6: Results obtained for the estimation of the yield per segment (composed of three vines) based on images captured with an “on-the-go” platform.

Grapevine variety	Manual harvest		Naïve estimator			Boolean model		
	Mean yield (g)	Number of segments	Mean yield (g)	R^2	RMSE (g)	Mean yield (g)	R^2	RMSE (g)
Cabernet Sauvignon	3406	8	1321	0.64*	1863.7	3235	0.50*	574.5
Syrah	3322	10	1598	0.87***	2031.9	3115	0.86***	390.6
Tempranillo	5031	10	2175	0.59**	3194.1	4660	0.60**	790.8
Total	3920	28	1698	0.71***	2472.0	3670	0.78***	610.1

These measurements were obtained when using a naïve estimator and the Boolean model on images segmented using a Mahalanobis distance classifier and a three-step filtering process. Asterisks represent statistical significance: * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$.

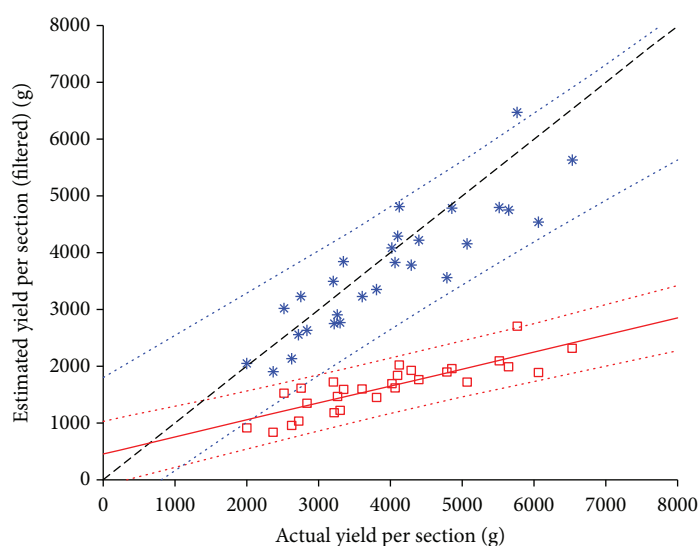


FIGURE 10: Yield per section (composed by 3 vines) using images (64 images combined to generate 28 segments) captured on-the-go and a naïve estimator represented in red squares ($y = 0.30x + 460.6$) and a Boolean model represented in blue stars ($y = 0.80x + 552.7$). A multistep filtering process was applied to improve the cluster segmentation (described in Section 2.5). The dashed line corresponds to 1:1, and dotted lines relate to 95% prediction intervals.

partially hidden ones was confirmed by the comparison between the results obtained with the Boolean model and the naïve estimator.

The simplicity and precision of the Boolean model formulation makes it ideal for its application on grapevine yield estimation, allowing its implementation in a fully automated system. The images were captured around 7 km/h, comparable to other agricultural equipment used in vineyard management, establishing this procedure close to industrial application. This methodology can also be used to generate maps that represent the spatial variability of the vineyards, allowing for grapevine zoning, segmented harvest, and thus an increase in quality.

Data Availability

The images and on-the-field recorded data used to support the findings of this study have not been made available because our institution is currently defining a protocol for data sharing.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

We would like to thank Vitis Navarra for their help with the field measurements. We also thank Dr. Jesús Angulo for his comments and suggestions. Borja Millan would like to thank the Center of Mathematical Morphology for the support, guidance, and laboratory facilities during his 6-month research stay conducting to this publication. Borja Millán was funded by the FPI grant [536/2014] by the University of La Rioja.

References

- [1] M. P. Diago, M. Krasnow, M. Bubola, B. Millan, and J. Tardaguila, “Assessment of vineyard canopy porosity using machine vision,” *American Journal of Enology and Viticulture*, vol. 67, no. 2, pp. 229–238, 2016.

- [2] M. Gatti, P. Dosso, M. Maurino et al., “MECS-VINE®: a new proximal sensor for segmented mapping of vigor and yield parameters on vineyard rows,” *Sensors*, vol. 16, no. 12, p. 2009, 2016.
- [3] A. Kicherer, M. Klodt, S. Sharifzadeh, D. Cremers, R. Töpfer, and K. Herzog, “Automatic image-based determination of pruning mass as a determinant for yield potential in grapevine management and breeding,” *Australian Journal of Grape and Wine Research*, vol. 23, no. 1, pp. 120–124, 2017.
- [4] A. Kicherer, K. Herzog, M. Pflanz et al., “An automated field phenotyping pipeline for application in grapevine research,” *Sensors*, vol. 15, no. 3, pp. 4823–4836, 2015.
- [5] J. A. Wolpert and E. P. Vilas, “Estimating vineyard yields: introduction to a simple, two-step method,” *American Journal of Enology and Viticulture*, vol. 43, no. 4, pp. 384–388, 1992.
- [6] S. R. Martin, G. M. Dunn, T. Hoogenraad, M. P. Krstic, P. R. Clingeffer, and W. J. Ashcroft, “Crop forecasting in cool climate vineyards,” in *Proceedings for the 5th International Symposium on Cool Climate Viticulture and Enology*, Melbourne, Australia, 2002.
- [7] G. Dunn, *Yield Forecasting*, Grape and Wine Research and Development Corporation, 2010.
- [8] G. M. Dunn and S. R. Martin, “The current status of crop forecasting in the Australian wine industry,” in *Proceedings of the ASVO Seminar Series: Grapegrowing at the Edge*, pp. 4–8, Tanunda, Barossa Valley, South Australia, 2003.
- [9] B. Millan, A. Aquino, M. P. Diago, and J. Tardaguila, “Image analysis-based modelling for flower number estimation in grapevine,” *Journal of the Science of Food and Agriculture*, vol. 97, no. 3, pp. 784–792, 2017.
- [10] S. Nuske, S. Achar, T. Bates, S. Narasimhan, and S. Singh, “Yield estimation in vineyards by visual grape detection,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2352–2358, San Francisco, CA, USA, September 2011.
- [11] M. P. Diago, J. Tardaguila, N. Aleixos et al., “Assessment of cluster yield components by image analysis,” *Journal of the Science of Food and Agriculture*, vol. 95, no. 6, pp. 1274–1282, 2015.
- [12] S. Liu, M. Whitty, and S. Cossell, “A lightweight method for grape berry counting based on automated 3D bunch reconstruction from a single image,” in *ICRA, International Conference on Robotics and Automation (IEEE), Workshop on Robotics in Agriculture*, p. 4, Seattle, WA, USA, 2015.
- [13] M. P. Diago, C. Correa, B. Millán, P. Barreiro, C. Valero, and J. Tardaguila, “Grapevine yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions,” *Sensors*, vol. 12, no. 12, pp. 16988–17006, 2012.
- [14] G. M. Dunn and S. R. Martin, “Yield prediction from digital image analysis: a technique with potential for vineyard assessments prior to harvest,” *Australian Journal of Grape and Wine Research*, vol. 10, no. 3, pp. 196–198, 2004.
- [15] M. Herrero-Huerta, D. González-Aguilera, P. Rodríguez-González, and D. Hernández-López, “Vineyard yield estimation by automatic 3D bunch modelling in field conditions,” *Computers and Electronics in Agriculture*, vol. 110, pp. 17–26, 2015.
- [16] S. Nuske, K. Wilshusen, S. Achar, L. Yoder, S. Narasimhan, and S. Singh, “Automated visual yield estimation in vineyards,” *Journal of Field Robotics*, vol. 31, no. 5, pp. 837–860, 2014.
- [17] D. Font, M. Tresanchez, D. Martínez, J. Moreno, E. Clotet, and J. Palacín, “Vineyard yield estimation based on the analysis of high resolution images obtained with artificial illumination at night,” *Sensors*, vol. 15, no. 4, pp. 8284–8301, 2015.
- [18] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, NY, USA, 1975.
- [19] J. Serra, “The Boolean model and random sets,” *Computer Graphics and Image Processing*, vol. 12, no. 2, pp. 99–126, 1980.
- [20] D. Jeulin, “Random texture models for material structures,” *Statistics and Computing*, vol. 10, no. 2, pp. 121–132, 2000.
- [21] S. Jeanson, J. Chadœuf, M. N. Madec et al., “Spatial distribution of bacterial colonies in a model cheese,” *Applied and Environmental Microbiology*, vol. 77, no. 4, pp. 1493–1500, 2011.
- [22] J. Angulo, “Nucleus modelling and segmentation in cell clusters,” in *Progress in Industrial Mathematics at ECMI 2008. Mathematics in Industry, Vol 15*, A. Fitt, J. Norbury, H. Ockendon, and E. Wilson, Eds., pp. 217–222, Springer, Berlin, Heidelberg, 1st edition, 2010.
- [23] J. P. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Orlando, FL, USA, 1982.
- [24] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, Berlin, Heidelberg, 2nd edition, 2004.
- [25] G. J. McLachlan, “Mahalanobis distance,” *Resonance*, vol. 4, no. 6, pp. 20–26, 1999.
- [26] H. M. Al-Otum, “Morphological operators for color image processing based on Mahalanobis distance measure,” *Optical Engineering*, vol. 42, no. 9, p. 2595, 2003.
- [27] J. Angulo and S. Velasco-forero, “Semi-supervised hyperspectral image segmentation using regionalized stochastic watershed,” in *Proceedings Volume 7695, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI; 76951F*, Orlando, FL, USA, May 2010.
- [28] D. A. McQuarrie, *Statistical Mechanics*, University Science Books, 1976.