Title: Grapevine Flower Estimation by Applying Artificial Vision Techniques on Images with Uncontrolled Scene and Multi-Model Analysis

Authors: Arturo Aquino, Borja Millán and Javier Tardáguila

Instituto de Ciencias de la Vid y del Vino (University of La Rioja, CSIC, Gobierno de La Rioja), 26006, Logroño, La Rioja, Spain

Corresponding author: Arturo Aquino (arturo.aquino@unirioja.es, (0034) 941299737)

Abstract

New technologies in precision viticulture are increasingly being used to improve grape quality. One of the main challenges being faced by the scientific community in viticulture is early yield prediction. Within this framework, flowering as well as fruit set assessment is of special interest since these two physiological processes highly influence grapevine yield. In addition, an accurate fruit set evaluation can only be performed by means of flower counting. Herein a new methodology for segmenting inflorescence grapevine flowers in digital images is presented. This approach, based on mathematical morphology and pyramidal decomposition, constitutes an outstanding advance with respect to other previous approaches since it can be applied on images with uncontrolled background. The algorithm was tested on 40 images of 4 different *Vitis vinifera* L. varieties, and resulted in high performance. Specifically, values for *Precision* and *Recall* were 83.38% and 85.01%, respectively. Additionally, this paper also proposes a comprehensive study on models for estimating actual flower number per inflorescence. Results and conclusions that are developed in the literature and treated herewith are also clarified. Furthermore, the use of non-linear models as a promising alternative to previously-proposed linear models is likewise suggested in this study.

Keywords: grapevine flower segmentation; flower estimation; yield prediction; precision viticulture; image analysis.

1 Introduction

The progress of technology has produced an increased interest in the development of novel techniques in the field of viticulture. Objective and automated vineyard assessment is of special interest nowadays. In this respect, yield prediction in vineyards is probably the most challenging goal from a technical point of view, and is experimenting much interest by the scientific community (Nuske *et al.*, 2011; Nuske *et al.,* 2014; Font *et al.*, 2014; Diago *et al.*, 2012; Roscher *et al.*, 2014; Dunn *et al*., 2004). Yield predictions are key tools for managing vines to optimize growth and then, for improving fruit quality.

Grapevine yield is predominantly determined by two physiological processes: flowering and fruit set (May, 2004). Fruit set presents a well-known variability among varieties and clones (May, 2004; Dry, 2010; Galet, 1983), and can also be affected by physiological, environmental and pathological factors (Carbonneau, 2007). Furthermore, fruit set also shows a great inter- and intra-vine variability (May, 2004). Therefore, a count of the flower number per inflorescence is essential for its accurate estimation. Moreover, performing this task in a non-destructive manner is of vital importance for the goals of precision viticulture.

For reasons mentioned above, some methods for flower number estimation have been presented. On the one hand, May (2000) and Keller *et al.* (2001) proposed a method based on wrapping sample inflorescences with a fine mesh from the beginning of anthesis until fruit set completion. Then, the collected flower caps in the mesh were manually counted in order to estimate the number of flowers per cluster. This method, in spite of being valid, is time consuming and labour demanding. On the other hand, Poni *et al.* (2006) proposed the use of digital photography for flower number estimation. First, the authors photographed each sample inflorescence in a study set against a dark background. Then, the number of flowers present in each image was manually counted. Finally, the real number of flowers per inflorescence was estimated using a linear model. This model performed a linear regression between actual flower number and the flower number manually counted on photos. The model was calibrated using data from twenty inflorescences taken from extra vines. The work by Poni *et al.* represented a conceptual advance in the estimation of flowers per inflorescence, since its automation would extremely decrease the workload from previous approaches. To this extent, Diago *et al.* (2014) developed an automated methodology for counting flowers in inflorescences by means of image analysis. Images were taken placing a dark background cardboard behind inflorescences for facilitating the calculation of a region of interest (ROI). After ROI extraction using colour discrimination, flowers were detected by recognising the reflection pattern produced by the light on the surface of flowers. Finally, the authors studied correlation between the number of flowers present in an image and the real number of flowers in its corresponding inflorescence. As a result, acceptable correlations per variety were found, whereas correlation of the defined variables was poorer considering varieties as a whole. This result led authors to discuss the suitability of using individual linear models per variety instead of a general one. To the best of the authors' knowledge, despite artificial vision is increasingly being applied to viticulture, the work by Diago *et al.* (2014) is unique for the automated estimation of flower number per inflorescence.

The present paper proposes a new methodology for the automated segmentation of flowers in inflorescence images under field conditions by means of morphological image processing and pyramidal decomposition. The algorithm is capable of working without the need of placing a dark background cardboard behind the inflorescences. This feature eases the use of the algorithm in field, since the cardboard is uneasily placed in specific situations, and likewise gets wet or dirty, or even torn. Moreover, the process of placing the dark cardboard and taking the photo at the same time is hardly performable by a person alone. Additionally, a rigorous study on models for the estimation of real number of flowers per inflorescence from flowers counted on images is presented. Conclusions from results of this study do not completely match with those previously developed by other authors. Therefore, authors find necessary a more in depth study and discussion over the results appearing hereafter.

2 Material and methods

2.1 Image acquisition

For developing and testing the segmentation algorithm, 40 inflorescence RGB images of *Vitis vinifera* L. cvs Airen, Albariño, Tempranillo and Verdejo were acquired, 10 per variety, in a commercial vineyard located in Vergalijo (Navarra, Spain), during May 2014 season. Phenological stage of varieties was 18, according to the scale proposed by Coombe et al (1995) (flower caps still in place, but cap colour fading from green). RGB images were captured at 6000 × 4000 pixels in size (24 Mpx), 8 bits per channel, using a Nikon D5300 reflex camera (Nikon corp., Tokyo, Japan); no tripod was used. The lens used was a Sigma (Sigma corp., Kanagawa, Japan) 50mm F2.8 macro. With respect to camera configuration, the settings for the main parameters were:

- Diaphragm opening: to obtain the maximum field depth provided by the lens, the minimum value (f/36) was used.
- ISO sensitivity: it was set to values providing proper image illumination.
- Shutter speed: this parameter was automatically set by the camera.

The distance between the camera lens and the inflorescence was not pre-established, but this was considered to be around 30–50 cm. No artificial lighting system or background homogenisation were used in order to mimic the variable outdoor conditions.

The 40 inflorescences photographed for creating the described set were not cut, since they were monitored until harvest. As a consequence, the total number of flowers, indispensable data for developing the estimation models study, could not be counted. This is why, with this purpose, a new set of 48 images of the same varieties (12 per variety) were taken under the same conditions than those previously detailed.

2.2 Methodology for flower segmentation in inflorescence digital images

The methodology proposed for flower segmentation was divided into two main phases: the ROI extraction (section 2.2.1) and flower segmentation (section 2.2.2). As a preliminary step, images were scaled down to a resolution of 1500 x 1000 pixels in size (0.25 times the original size) for reducing computational workload. Another important decision was the selection of the image colour space used. Images were taken according to the RGB colour scheme (this is determined by the constructive features of the camera sensor); however, this scheme did not properly represent image information in this study. Conversely, the HSV colour space represents structured image information into three noteworthy axes: hue, saturation and value. The hue channel condenses information on the colour shade; the saturation expresses its pureness; and the value its lightness. Therefore, RBG images were converted to HSV colour space prior to being processed.

Much of the processing carried out in this paper is based on mathematical morphology. For completeness purposes, a brief description of this image processing technique along with mathematical definition of used operators is offered in Appendix A.

2.2.1 *ROI* extraction

Inflorescences appear in images as a greenish look on heterogeneous, variable and unknown background (Fig. 1). Dealing with this situation is probably the most challenging task, since this stochastic variable may be source of numerous unexpected detection errors. With the aim of avoiding this background effect, a ROI is extracted from the image.

Let *H* be the 8-bit image (pixel values in the interval [0, …, 255]) of the hue channel from the original HSV image, a first ROI approach is the extraction of all green objects using colour information contained in channel *H:*

$$ROI_G(x,y) = \begin{cases} 0 \ if \ 40 \le H(x,y) \le 76 \\ 1 \ \text{otherwise} \end{cases}$$

__Fig. 1.__ Examples of scene variability in inflorescence images.

Since inflorescence occupies an outstanding area in image, a cleaner version of ROI_G can be safely created by discarding small connected components:

$$ROI_{G'} = \{CC_i \subseteq ROI_G | \#(CC_i) < (\#(H) * 0.02)\}$$

where # represents the cardinal operator and CC_i a connected component (defined by 8-connectivity). Therefore, those connected components in ROI_G not containing at least 2% of total pixels in the image are discarded. The $ROI_{G'}$ mask maps green objects in H but is still insufficient for the described purposes, since other green objects apart from the inflorescence may be present in the scene (see Fig. 2). This is why further processing is needed to obtain a more accurate ROI.

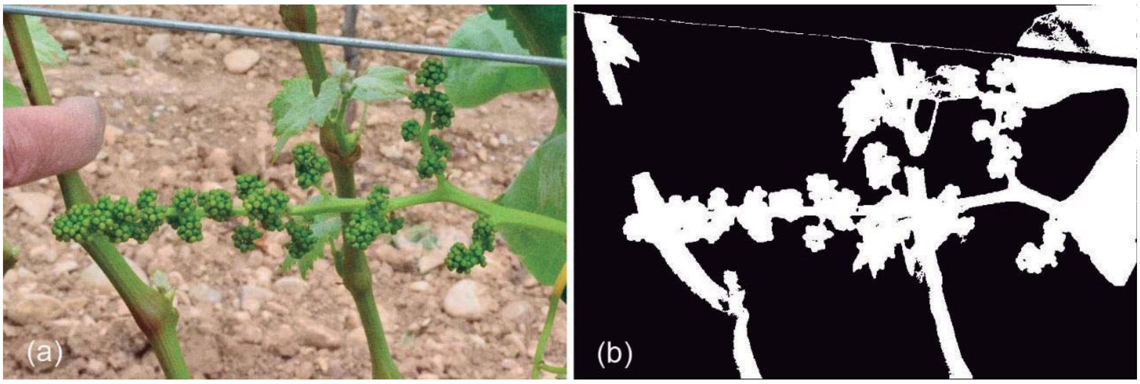

__Fig. 2.__ Extraction of a ROI based on pixel colour: (a) original RGB image, (b) ROI calculated by binarization on the H channel of the HSV colour space for green objects detection.

Inflorescences are constituted by compact sets of flowers having these last circular shape. Therefore, the flowers are joined together configuring those sets. As a consequence, the incidence of light creates a honeycomb-like connected structure of shadows going along the interior borders (Fig. 3). This feature is exploited to extract the inflorescence from the rest of the scene.

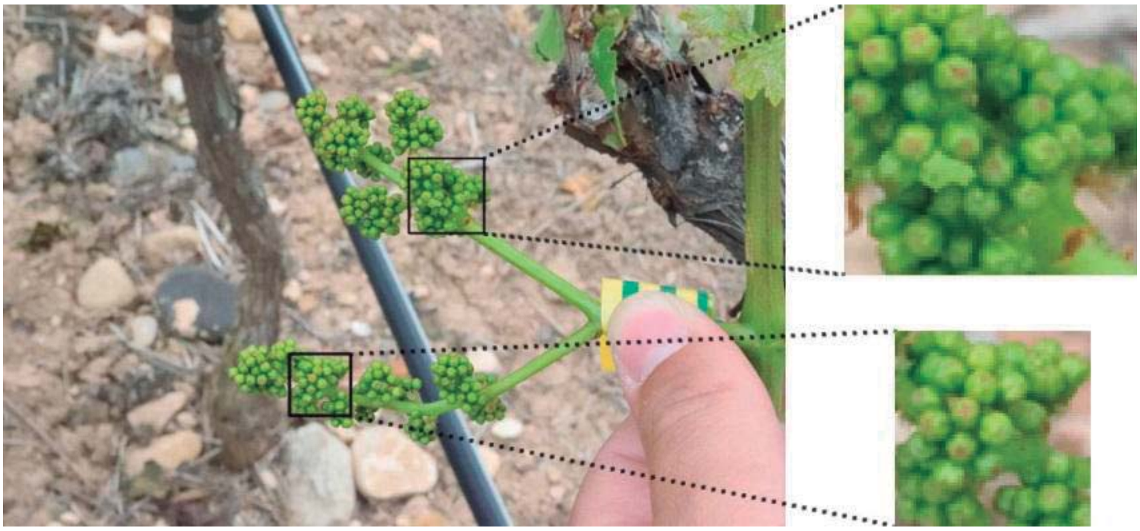

Fig. 3. *Illustration of the interior structure of shadows produced by flowers.*

Consider *S* to be the image from the saturation channel. As shown in Fig. 4-(c), flower shadows appear brighter than its surroundings in this image modality. In addition, it can be assumed that they are piecewise linear. With these considerations, the first step is blurring linear bright objects in the image thinner than flower shadows. Opening the image with a linear structuring element (SE) of width 1 and length *L*, all those bright structures which cannot contain it are removed whereas those which can contain the SE are preserved. Performing multiple openings with such an SE at different rotations and taking the infimum of all those results, linear objects with different orientations are evaluated. Mathematically:

$$I_L = inf_{i=1,\ldots,12}\{\gamma_{B_i}(S)\}$$

where B_i represents SE *B* at rotation *i*; 12 rotations of *B* taken each 15º apart were used. The length *L* of *B* should be chosen so as to ensure that *B* can be contained by shadow segments at all rotations. A value of 5 pixels was selected according to performed tests, although other values may also be valid. Results of this operation can be examined in Fig. 4-(d).

As was previously mentioned, shadows in *S* are brighter than their surroundings. In other words, they constitute a frontier separating dark regions representing flowers. The next

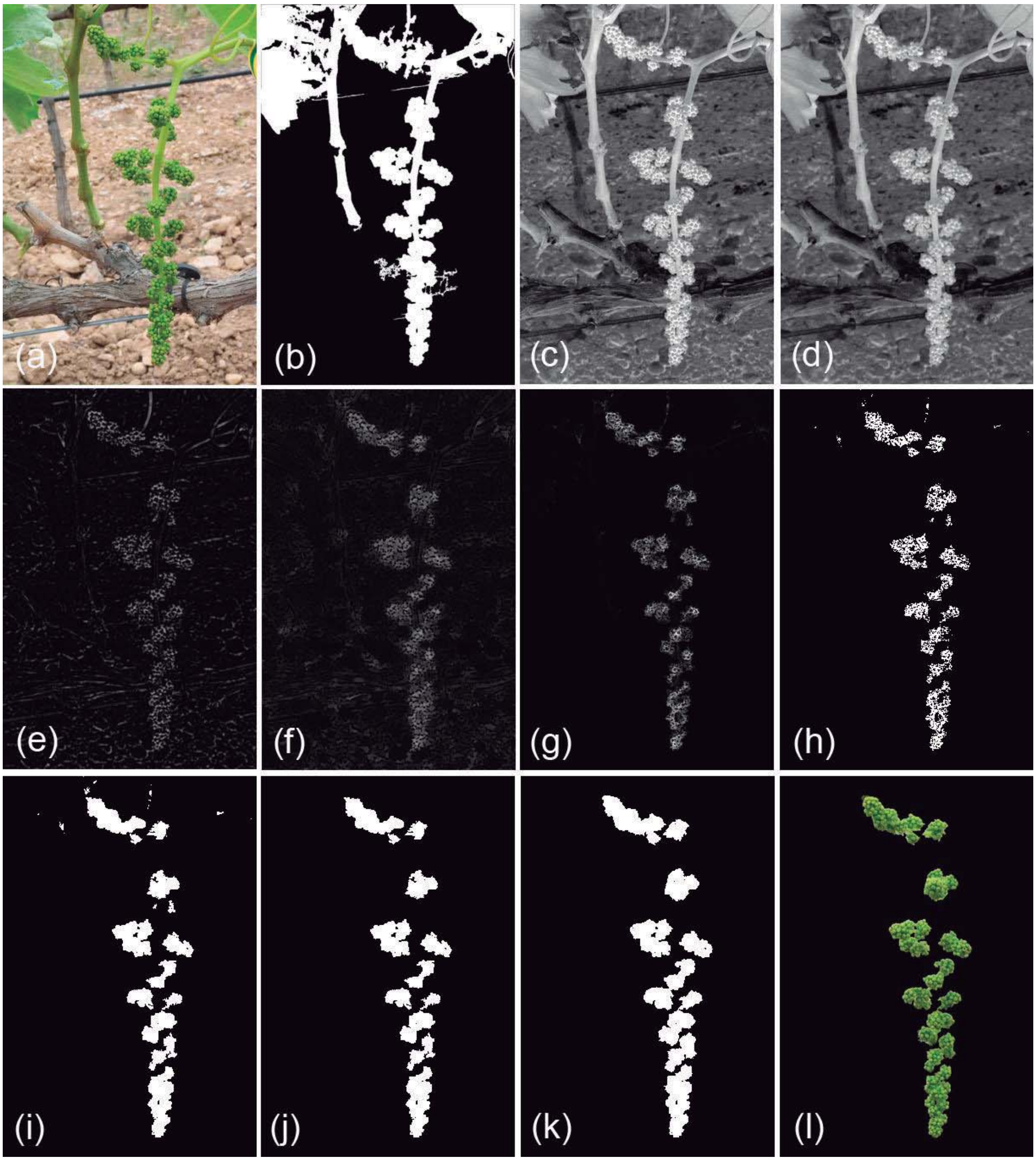

Fig. 4. *Illustration of processing for ROI calculation: (a) original RGB image; (b) ROI based on colour; (c) saturation channel of the HSV colour space; (d) elimination of thin bright objects; (e) top-hat image; (f) morphological analysis of granulometry; (g) product of images (e) and (f) filtered with ROI in (b); (h) binarization of (g); (i) result of iterative and controlled region growing on (h); (j) elimination of false positives; (k) definitive ROI resulting from partial reconstruction of (b) from marker (j); and (l) obtained ROI illustrated on original image (a).*

processing is aimed at extracting these so-called borders while discarding those in contrast, that is, the dark borders. To this effect, a morphological top-hat transformation is applied using a SE with diamond shape:

$$I_{TH} = TH_B(I_L)$$

where *B* is the SE. A diamond shape was selected since it presents good fit in confluence points. The value of its diagonal length was chosen so as to prioritize the extraction of wider borders. This value is not critical since several ones are valid; the value 7 was set in our case (results of this processing can be seen in Fig. 4-(e).

Image I_{TH} contains the original structure of shadows along with other bright borders, which may have been retained. The following approach is to strengthen only flower shadows as much as possible. To this extent, a texture analysis based on morphological granulometry assessment is firstly performed in order to emphasize dark circular patterns in I_{TH} (note that flowers have these features in I_{TH}). Closing the image with a circular SE *B* of radius *R*, all those flowers, with a radius less than or equal to R, are recognized. By taking the supremum of multiple closings with a proper range of *R* values, flowers with different sizes are detected. Mathematically this is defined as

$$I_F = sup_{i=1,\dots 12}\{\varphi_{B_i}\}$$

where radius values from 1 to 12 were considered (outcomes of this processing can be seen in Fig. 4-(f)). A great part of flower shadows in I_{TH} belongs to borders of detected flowers in I_F. Thus, by means of calculating the product of both images, flower shadows are definitely intensified with respect to other objects (Fig. 4-(g)):

$$I_S = I_{TH} \times I_f$$

Operator $\times$ stands for the element-by-element product. It should be stressed that the result of this operation has to be scaled down to range values [0, …, 255]. Moreover, mask $ROI_{G'}$ firstly calculated is used at this point to discard those objects in I_S not corresponding to green objects in the original image:

$$I_{S'} = I_S \times ROI_{G'}$$

Once the object of interest has been strengthen from the background, the following step is its segmentation. This process is performed by means of binarization. Due to the huge image variability conditions, binarization threshold has to be automatically calculated for every image from its features. To this effect, the Otsu thresholding method (Otsu, 1975) automatically calculates a threshold for a grey-level image by taking the assumption that it is composed of two sets, the background and the foreground. Then, the method obtains the optimum threshold T_{otsu} by maximizing the between-class variance. This threshold is used to perform a 2-level-based binarization followed by morphological reconstruction. Two binary images I_{otsu} and $I_{otsu'}$ are obtained by using values T_{otsu} and T_{otsu}*0.65, respectively:

$$I_{otsu}(x,y) = \begin{cases} 0 \; if \; I_{S'} \leq T_{otsu} \\ 255 \; otherwise \end{cases}; \; I_{otsu'}(x,y) = \begin{cases} 0 \; if \; I_{S'} \leq T_{otsu} * 0.65 \\ 255 \; otherwise \end{cases}$$

Binary image I_{otsu} contains borders corresponding to the strongest shadows. $I_{otsu'}$ contains those in I_{otsu} along with others from weaker flower shadows. Since these make a connected structure, a morphological reconstruction using $I_{otsu'}$ as mask and I_{otsu} as marker will extract it and discard other objects (see Fig. 4-(h) and Fig. 5):

$$I_{BIN1} = R^{\infty}_{I_{otsu'}}(I_{otsu})$$

Once inflorescence borders have been extracted in I_{BIN1}, they are used to create a first version of the definitive ROI. Binary image I_{BIN1} contains borders from flower shadows and may also contain some noise. Shadow borders are circular or at least have an arc shape (due to segmentation discontinuities). Contrary, noisy borders tend to be more variable in shape. In order to extract the inflorescence discarding other objects, an iterative and selective filling algorithm starting from I_{BIN1} was designed. Basically, this algorithm dilates objects in an increasing magnitude. At each expanding step, holes are filled. Those areas that have been expanded by filling are retained and the rest are restored to their initial size. Mathematically:

$$ROI_1 = (R^{\infty}_{I_i}(I_i - I_{i-1}) \vee I_{i-1})$$

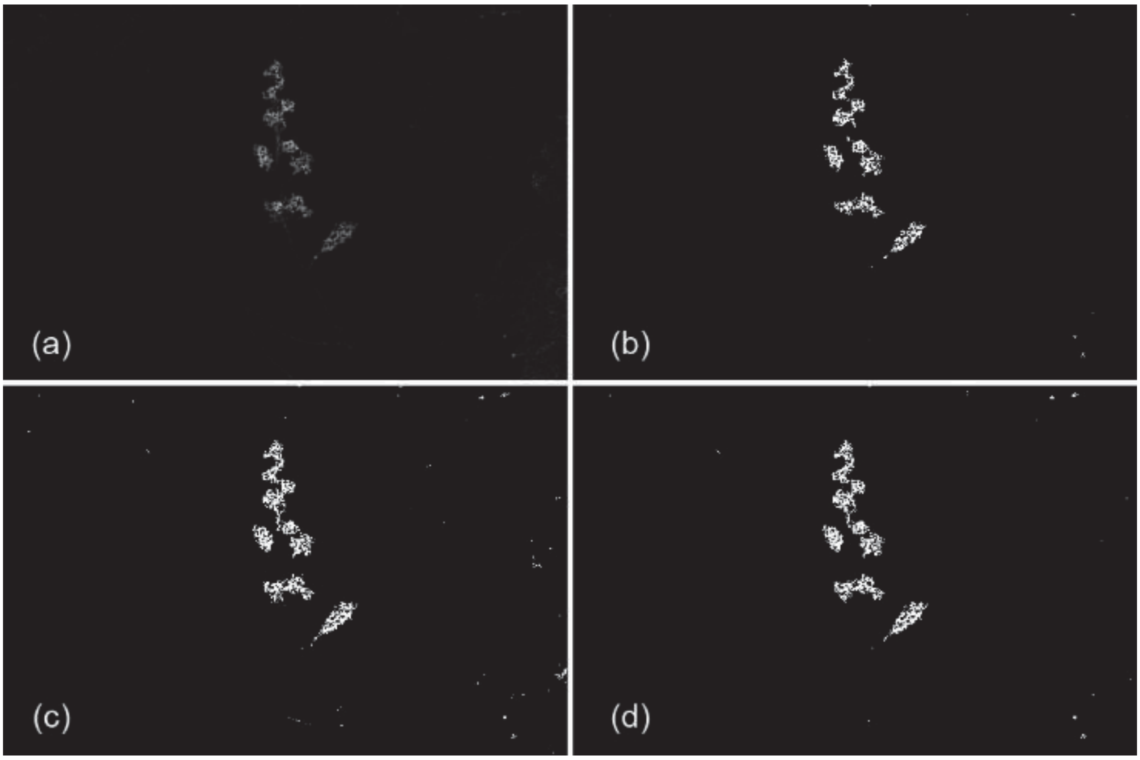

*__Fig. 5.__ Illustration of the 2-level-based binarization process followed by morphological reconstruction: (a) image to be binarized, the goal is to segment the inflorescence from the rest of the scene; (b) binary image resulting from applying the threshold given by the Otsu method (Totsu); (c) binary image obtained with threshold Totsu*0.65; and (d) image resulting from the morphological reconstruction of (c) from marker (b).*

where *i* is such that

$$R_{I_i}^{\infty}(I_i - I_{i-1}) = R_{I_{i-1}}^{\infty}(I_{i-1} - I_{i-2})$$

and

$$I_i = \psi\left(\delta_{B_i}(I_{BIN1})\right) \wedge ROI_{G'};\ i = 1, \dots, n;\ I_0 = I_{BIN1}$$

Indeed, the designed algorithm fills circular patterns even when they are incomplete and keep invariable other irregular shapes. Results of this processing are shown in Fig. 4-(i); a step by step illustration is given in Fig. 6-(a)-(c). Assuming that wider connected components in ROI_1 belong to the inflorescence, noisy objects are eliminated from ROI_1 by performing:

$$ROI_2 = I_i$$

where

$$I_i = R_{I_{i-1}}^{\infty}\left(\gamma_{B_i}(ROI_1) \wedge I_{i-1}\right)$$

i is that fulfilling

$$I_i = I_{i-1}; i = 1, \dots, n$$

and being

$$I_0 = ROI_1$$

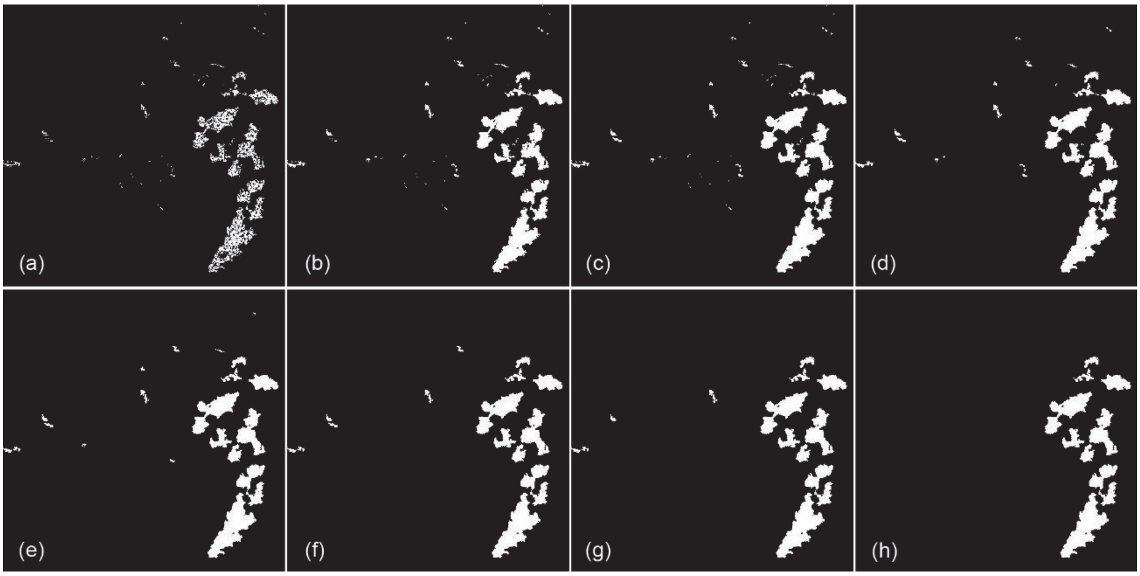

Fig. 6. *Process of controlled region growing and false positives elimination: (a) starting image; (b) first step of growing; and (c) second and definitive step of growing, since the idempotence is reached. (d)-(h) False positives are increasingly eliminated until idempotence.*

Certainly, noisy objects are increasingly removed from ROI_1 until stability (the result can be observed in Fig. 4-(j); Fig. 6-(d)-(h) illustrates the whole process). Finally, since the inflorescence has been extracted from the interior flower shadows, peripheral flowers may not have been included in ROI_2. The definitive ROI, ROI_{def}, is the partial reconstruction of $ROI_{G'}$ from marker ROI_2:

$$ROI_{def} = R^n_{ROI_{G'}}(ROI_2)$$

For establishing a precise value for the number of partial reconstruction steps, parameters such as flower size and exact distance to object, among others, should be known. Since they are unknown in this work, a value based on testing was selected; this value was 8. An illustration on ROI calculation can be examined in Fig. 4, images (k) and (l).

2.2.2 Flower segmentation

From a geometric point of view, flowers are small quasi-spheres. When light reaches a spherical object, a point of maximum reflection is produced on its surface. The magnitude of this reflection progressively decreases around this location according to a circular pattern. Ideally, the point of maximum reflection could be considered as the centre of a family of circumferences of increasing radius. Intensity of reflection decreases as the radius increases. This phenomenon is exploited for detecting individual flowers in inflorescences.

Let *V* be the value channel of the HSV colour space representing image illumination information (Fig. 7-(a)). Firstly, flower frontiers are improved by means of opening the image with a rotating SE of width 1 and length *L*. Since flower borders are dark in image *V*, they are strengthened by performing this set of openings and taking the infimum of all the results:

$$I_E = inf_{i=1,\dots,12}\{\gamma_{B_i}(V)\}$$

where B_i stands for SE *B* at rotation *i*; 12 rotations of B were used. The value of *L* was chosen to ensure that *B* could be contained by border segments at least at one rotation. A proper *L* value was 5 pixels, although slightly higher and lower values may also be useful.

As a second step, irrelevant peaks are selectively removed from the image. Concretely, peaks with a height of only one grey level are considered as noise and consequently discarded. This is achieved by means of morphologically reconstructing I_E from marker I_{E-1}:

$$I_C = R^{\infty}_{I_E}(I_E - h), h = 1$$

Indeed, morphological reconstruction of I_E from I_{E-1} restores all pixel values that were not originally regional maxima of height 1. This operation is known as h-maxima transformation (being *h*=1 in this case).

When flowers are close to open, their surface gets wrinkled around the emerging or blooming point creating lobs. These lobs may produce more than a maximum light reflection, thus creating redundant detection points (false positives). In order to avoid such output as much as possible, a gradual detection scheme is applied by making use of Gaussian pyramidal decomposition (Burt, 1981). This technique consists in creating a set of images from the original one by means of smoothing and down-sampling it. Taking advantage of the spatial proximity of redundant maxima, the main idea is to detect maximum reflections and spatially

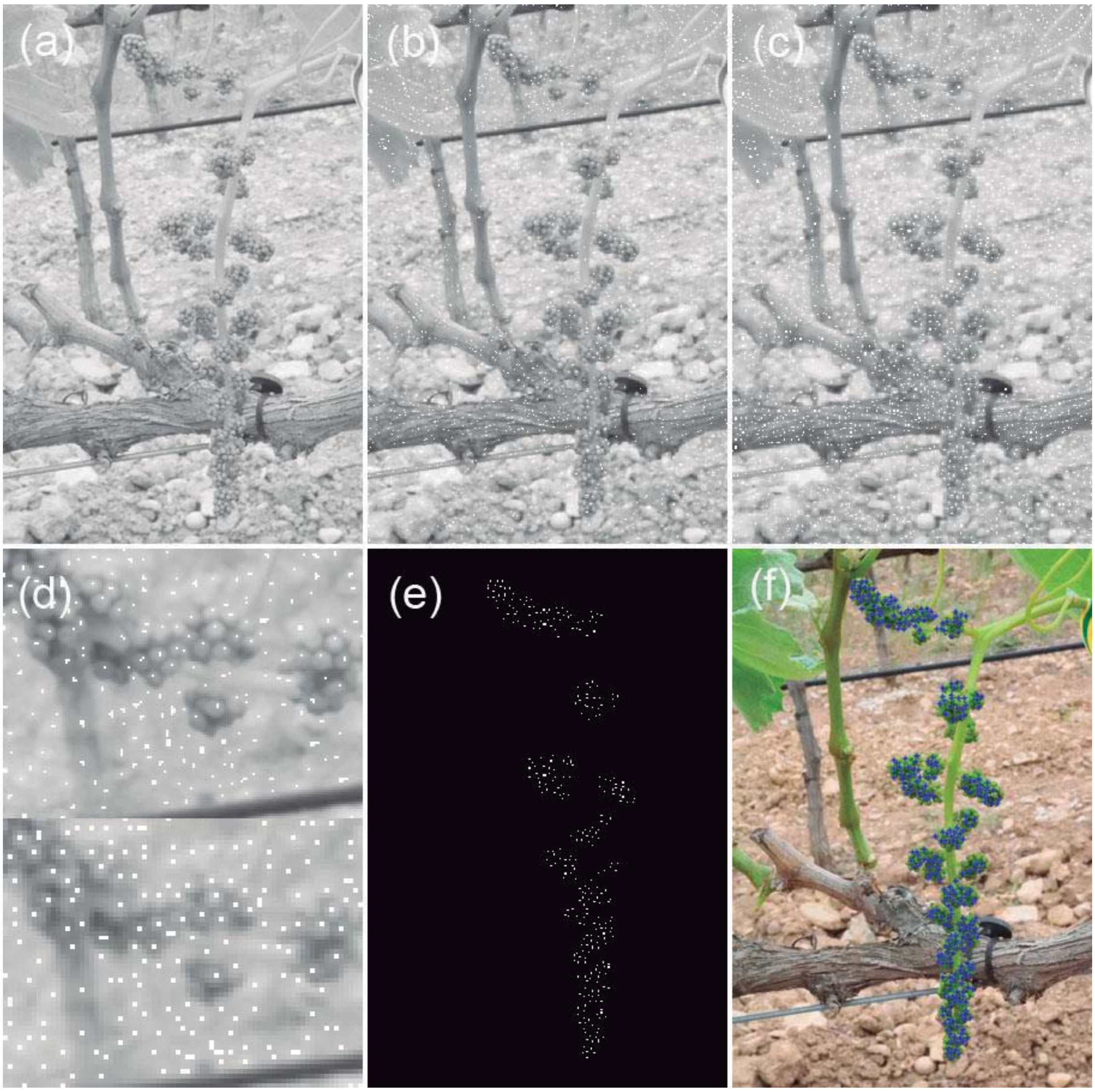

Fig. 7. *Illustration of processing for flower segmentation on the same image than in Fig. 4: (a) Value channel of the original image; (b) image from the first step of pyramidal decomposition, white dots are detected regional maxima; (c) image from the second step of pyramidal decomposition, where white dots are regional maxima; (d) two zoomed sub-images, the upper one from (b) and the lower one from (c); (e) representation of regional maxima found within the FOV; and (f) blue spots are the centroids of regional maxima in (e) and represent finally detected flowers.*

fusing them in a certain level of its pyramidal decomposition. To this effect, the first level of pyramidal decomposition of I_C is computed:

$$I_C^{PD_1} = PD(I_C, w)$$

where $PD(I_C, w)$ denotes a step of pyramidal decomposition of image I_C using the generating kernel pattern w. This kernel was defined by Burt (1981) as:

$$w = \left[\frac{1}{4} - \frac{a}{2}, \frac{1}{4}, a, \frac{1}{4} - \frac{a}{2}, \frac{1}{4}\right], a = 0.375$$

Note that the size of $I_C^{PD_1}$is half the value of size of I_C. Then, regional maxima in $I_C^{PD_1}$ are found by using the h-maxima transformation

$$I_{h-max}^{PD_1} = R_{I_C^{PD_1}}^{\infty}\left(I_C^{PD_1} - h\right), h = 1$$

and subtracting the result from the original image:

$$I_{R-max}^{PD_1} = I_C^{PD_1} - I_{h-max}^{PD_1}$$

All pixels in $I_{R-max}^{PD_1}$ with a value strictly higher than 0 belong to regional maxima in $I_C^{PD_1}$. In order to preserve these maxima in the following step of decomposition (remember that the aim of this processing is spatially fusing close maxima and not eliminating any of them), value of these pixels is set to the maximum (Fig. 7-(b)):

$$I'^{PD_1}_C(x,y) = \begin{cases} I_C^{PD_1}(x,y) \; if \; I_{R-max}^{PD_1}(x,y) = 0 \\ 255 \; otherwise \end{cases}$$

Next, a following step of pyramidal decomposition is performed and regional maxima are calculated:

$$I_C^{PD_2} = PD(I'^{PD_1}_C, w)$$

$$I_{h-max}^{PD_2} = R_{I_C^{PD_2}}^{\infty}\left(I_C^{PD_2} - h\right), h = 1$$

$$I_{R-max}^{PD_2} = I_C^{PD_2} - I_{h-max}^{PD_2}$$

Detected regional maxima, represented on $I_C^{PD_2}$, can be observed in Fig. 7-(c). Fig. 7-(d) also illustrates how the maxima spatially close in the first pyramidal decomposition fused into the second. At this point, all pixels with a value higher than 0 in $I_{R-max}^{PD_2}$ are considered to be pixel flowers:

$$I_{flowersBin}(x,y) = \begin{cases} 0 \; if \; I_{R-max}^{PD_2}(x,y) = 0 \\ 255 \; otherwise \end{cases}$$

Finally, the set of centroids, falling within ROI_{def}, of the connected components in $I_{flowersBin}$ represent the detected flowers (see Fig. 7, images (e) and (f)):

$$I_{flowers} = \left\{ centroid(CC_i) \middle| CC_i \subseteq I_{flowersBin} \land centroid(CC_i) \in ROI_{def} \right\}$$

It should be stressed that, since image $I_{flowersBin}$ derives from two steps of pyramidal decomposition, its size is four times less than the original one. Therefore, coordinates of centroids in $I_{flowers}$ must be scaled up in order to represent flower locations in terms of the original image size. Figure 8 shows some results of the whole process for flower segmentation.

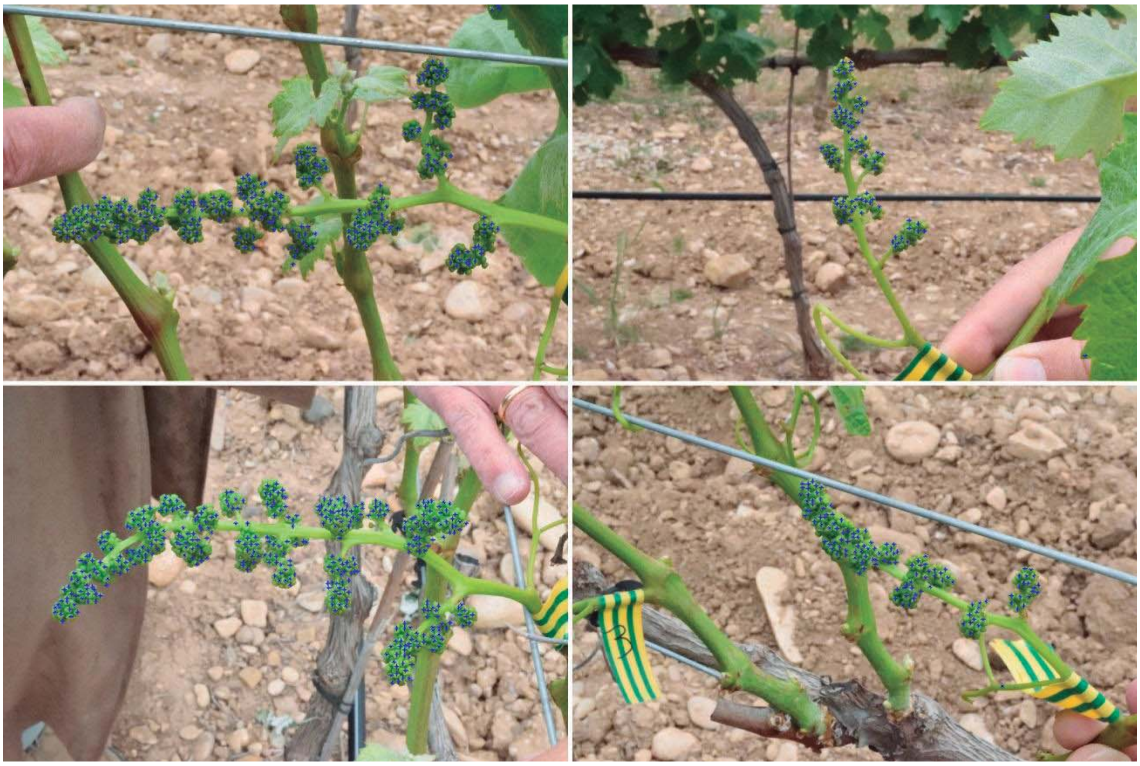

Fig. 8. *Illustration of flower segmentation results on four different images.*

3 Results and discussion

The presented methodology for segmenting flowers in grapevine inflorescence images is evaluated in the following section. Additionally, section 3.2 develops a study on different model approaches for estimating the total number of inflorescence flowers from information extracted from the image.

3.1 Performance evaluation of the presented segmentation algorithm for flower segmentation.

The algorithm was tested on the set of 40 images described in section 2.1. For evaluating its performance, the following metrics based on contingency tables for binary classification were employed:

$$RC = \frac{TP}{TP + FN}; PC = \frac{TP}{TP + FP}$$

Metric *RC* denotes *Recall*, and is the percentage of actual flowers detected by the algorithm. On the other hand, *PC* stands for *Precision*, which calculates the percentage of flowers correctly detected.

For making possible the application of the described metrics, a gold standard set was created. It was carried out by manually labelling flowers on each image in the set, making use of the software specifically developed to this effect. Thus, true positives (*TP*), false positives (*FP*) and false negatives (*FN*) were calculated as:

- *TP*: flowers automatically detected corresponding to actual flowers labelled in the gold standard.
- *FP*: flowers automatically detected, which do not correspond to actual flowers in the gold standard. Redundant *TPs* were also considered as *FP*.
- *FN*: actual flowers labelled in the gold standard which were not found by the segmentation algorithm.

Table 1 shows obtained results in terms of the *RC* and *PC* metrics. Figures in this table were calculated considering all the images together. In addition, results obtained in the previous work by Diago *et al.* (2014) are also included in this table for comparison purposes. It should be highlighted that results of Diago *et al.* (2014) were obtained on a different set of images, which were taken using the help of a dark cardboard as background. Furthermore, despite the authors collected 90 inflorescence images, the algorithm could be only evaluated on 15 of them. These facts make the rigorous comparison of both methods difficult, although some discussion can be brought up. The algorithm described in this paper shows evidence of being more balanced in terms of average *Precision* and *Recall*. The work by Diago *et al.* (2014) tends to produce less false positives, although this seems to significantly penalise the percentage of actual flowers that can be recognized. It could be justified by the application of a more conservative strategy. Furthermore, higher precision of the previous work could be logically expected, since the use of a dark cardboard as background avoids the huge ROI calculation

Table 1. Results of the proposed methodology compared to those obtained by Diago *et al.* (2014). Figures are given in terms of average *Precision* (*PC*) and *Recall* (*RC*); standard deviation obtained with both metrics is also presented. Diago *et al.* (2014) used a different dataset for calculating their results. This is why some features of that dataset and the one used in this study are detailed in this table.

Metric	Average	Standard deviation	Number of images	Grapevine varieties
This work				
Actual flowers in the gold standard set	225.65	89.8973	40	4
PC	0.8338	0.0971	-	-
RC	**0.8501**	0.1120	-	-
Diago et al. (2014)				
Actual flowers in photos	263.53	80.42	15	3
PC	**0.9290**	0.0300	-	-
RC	0.7430	0.0549	-	-

problem faced herein. In addition, the validation set used by Diago *et al.* (2014) was less diverse than the one used in this study, since it contained considerably fewer images and considered less grapevine varieties. In this respect, Table 2 details results of the presented methodology per variety. It can be noticed that accuracy, measured by *PC*, moderately varies among varieties. This can be justified by phenological development of the varieties, which was substantially more advanced for Tempranillo and Albariño. Flowers of these varieties were close to open. As was previously described, at this point, flower surface generates lobs around the opening point. When these lobs are sufficiently pronounced, they produce redundant maximum light reflections, thereby creating false positives. As a result, it can be inferred that even better results could have been obtained by taking the images in earlier phenological stages, even for Airen and Verdejo.

Table 2. Results of the segmentation methodology detailed per variety. Average and standard deviation values of *Precision* (*PC*) and *Recall* (*RC*) are given per variety. The average and standard deviation of flowers in the gold standard set (*GS*) are also given per grapevine variety.

Grapevine variety	$\overline{PC}$	σ_{PC}	$\overline{RC}$	σ_{RC}	$\overline{GS}$	σ_{GS}
Airen	0.8793	0.0903	0.8320	0.0696	284.14	93.79
Albariño	0.8016	0.0846	0.8320	0.0477	179.25	79.30
Tempranillo	0.7516	0.0918	0.8377	0.1525	160.75	66.47
Verdejo	0.8817	0.0463	0.8974	0.1272	240.2	59.88

3.2 Study on models for the estimation of the total number of flowers from flowers detected in images

Once flowers are counted on the image, the final step is, using the acquired information, the estimation of the actual number of flowers in the inflorescence. Studying the available bibliography in this sense, it can be concluded that one option has been explored. Poni *et al.* (2006) proposed the use of linear models to estimate the actual number of flowers in inflorescences of Sangiovese and Trebbiano grapevine varieties. Flowers were manually counted on images and linear regression was applied to correlate this information with actual inflorescence flowers. The two obtained regression equations, one for each variety, were proposed as estimation models. Diago *et al*. (2014) studied the use of estimation linear models more in depth, comparing the use of a unique variety-independent estimation model with the described previous approach. Both options were compared using the Pearson's correlation coefficient (R^2). The R^2 values obtained by the authors argued for the use of individual variety-dependent estimation models. Making an analysis of the described proposals, the following points can be concluded:

- The Pearson's correlation coefficient is not suitable for assessing the behaviour of an estimation model on its own. It gives an accurate idea about the trend similarity of the actual and estimated variables. However, R^2 does evaluate the performance of an estimation model.

- The fact that individual linear models showed good behaviour may argue for considering that inflorescences from different grapevine varieties have distinctive features. If this were true, it would imply that variables under modelling may have a non-linear relation, which would be interesting to assess.
- Models were not created and evaluated using a two-phase approach in which two disjoint sets should be used for obtaining and testing models.

As a result of these conclusions, what follows in this section is a comprehensive study on models for the estimation of inflorescence flowers using the number of flowers counted in an image.

A set of 48 images of the same varieties previously used (Airen, Albariño, Tempranillo and Verdejo) was acquired. For that, inflorescences were coded, photographed and cut after capture. Then, flowers were manually counted in a destructive manner. Counting results were registered individually attending to previously established coding. Finally, inflorescence flowers were also manually counted on images using the software specifically designed for this goal and figures were registered accordingly. At this stage, two disjoint datasets were created for model obtaining and evaluation. The first one, referred to as training set, was composed of 20 images, 5 per variety. The validation set was created using the remaining 28 images, 7 per variety.

Fig. 9-(a) represents individual linear models acquired using linear regression on the training set. Calculated model equations as well as R^2 values are given. In contrast, Fig. 9-(b) shows the described information for the case of variety-independent linear model calculation. Afterwards, calibrated models were employed to generate predictions using the validation set. Fig. 9-(c) analyses behaviour of the actual and predicted variables produced by individual models. Fig. 9-(d) illustrates the same feature for the case of the variety-independent linear model. As shown, R^2 values calculated for both approaches are considerably high, even higher than those obtained in other previous studies (see Table 3).

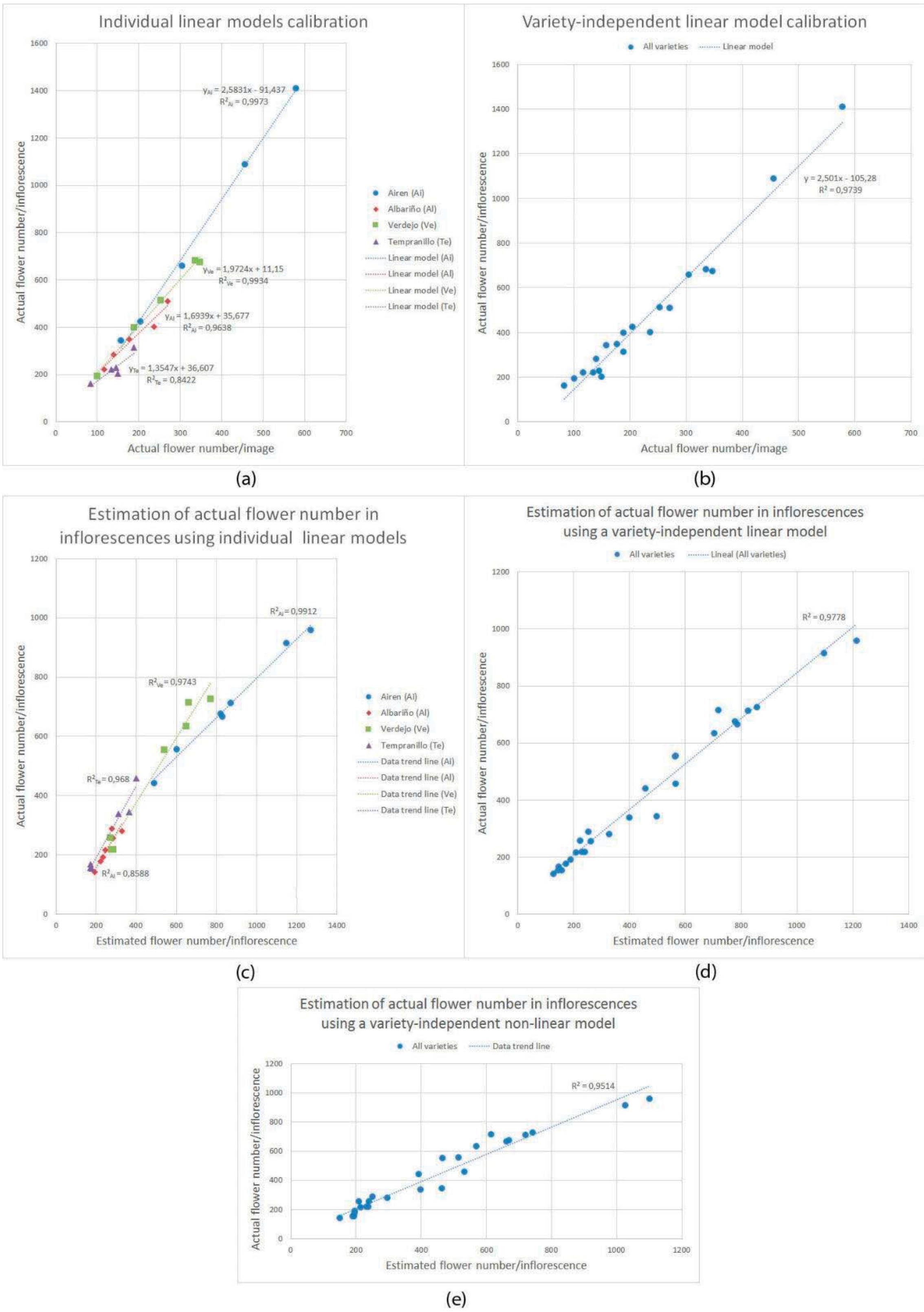

Fig. 9. *Comparison of different model approaches for actual flower estimation: (a) and (b) illustration of individual and variety-independent linear models calculation, respectively; (c) and (d) representation of performance of both approaches; and (e) performance representation of a non-linear variety-independent model. Root-mean-square error (RMSE) produced by (c), (d) and (e) are given in Table 3.*

This outcome has even more relevance taking into account that, in contrast to previous works, they were obtained on a validation set “unknown” for the model.

Table 3. R_2 values comparison of those obtained in this and other previous studies for different estimation model approaches.

Variety	Variety-independent linear model (R^2)	Variety-dependent linear model (R^2)	Variety-independent non-linear model (R^2)
		This work	
Airen	0.9912	0.9912	0.9945
Albariño	0.8588	0.8588	0.8761
Tempranillo	0.9680	0.9680	0.9556
Verdejo	0.9743	0.9743	0.9789
Total	0.9778	0.9528	0.9514
		Diago *et al.* (2014)	
Graciano	-	0.8100	-
Carignan	-	0.8900	-
Tempranillo	-	0.8700	-
Total	0.8100	-	-
		Poni *et al.* (2006)	
Sangiovese	-	0.8800	-
Trebbiano	-	0.8700	-
Total	-	-	-

The fact that results from previous experiments by other authors argued for individual models per variety opens the possibility to consider the evaluation of a non-linear approach. It would be justified if varieties would show inherent and distinctive features affecting flowers prediction. In an attempt of characterising them and evaluating a non-linear solution, a feature space composed of the following axes was defined:

- Number of flowers in the image:

$$f_1 = \#(I_{flowers})$$

- ROI area:

$$f_2 = \#(ROI_{def})$$

- Flower radius estimation:

$$f_3 = r_{flower}$$

- Flower density:

$$f_4 = \frac{f_1}{f_2}$$

- Flower area:

$$f_5 = \log(\pi * f_3^2)$$

The flower radius was estimated by calculating the average of the minimum distances among flowers (this is the flower diameter estimation) and dividing this result by 2. Then, a multilayer feed-forward backpropagation neural network was implemented for obtaining the non-linear estimation model. The neural network had 5 input neurons fed by the defined descriptors, a hidden layer with two neurons and an output; the transfer function was set to linear. The neural network was trained on the training set and tested using the validation set. As shown in Fig. 9-(e) and Table 3, correlation between the actual and estimated variables for the case of the non-linear model is lower than those obtained with the linear approaches, although it is high in absolute terms.

At this point, the three tested approaches have provided a high correlation between actual and predicted variables. To accurately assess and compare the predictive potential of all options, the root-mean-square error (*RMSE*) is proposed:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{Fl}_i - Fl_i)^2}{n}}$$

where $\widehat{Fl}_i$ and Fl_i are the predicted and actual flower number values of the i*th* image in the validation set, respectively. Table 4 includes results in terms of RMSE produced by the three studied approaches on the validation set. They are detailed per varieties and well as considering all of them together. Taking into account global results, there is not any observable justification for claiming the use of individual linear estimation models. This is a remarkable issue, since it is unmatched by other previous conclusions. Furthermore, the use of a unique linear estimation model for all varieties simplifies the prediction problem significantly. With regard to results of the non-linear model, despite global results are promising and even better than those produced by any other, this should be carefully discussed. In the authors' opinion, the suitability of the non-lineal model should be proven with a wider set of varieties so as to verify with more confidence the new feature space. In other words, in spite of being really promising, further research is considered necessary before accepting the increased complexity derived from the use of a non-linear model.

Table 4. Root-mean-square error (*RMSE*) produced by each model estimating the total number of flowers per inflorescence from the number of flowers in inflorescence image. Results are detailed per variety and also given considering all together.

Variety	Variety-independent linear model (*RMSE*)	Variety-dependent linear model (*RMSE*)	Variety-independent non-linear model (*RMSE*)
Airen	138	180	**72**
Albariño	24	40	**19**
Tempranillo	57	**43**	61
Verdejo	75	**29**	63
Total	84	95	**58**

4 Conclusion

This paper proposes a new methodology for flower segmentation in digital images of inflorescences of *Vitis vinifera* L. It is mainly based on mathematical morphology and pyramidal decomposition. The algorithm is capable of functioning under field conditions and without the need of placing a black cardboard behind the inflorescence. This supposes an advantage with respect to previous works since, besides making easier the process of taking the images, it also opens the door to its integration in vehicles and autonomous robotic platforms after further research. On the other hand, it has been found that several considerations prior to taking captions could even improve the obtained results. Taking photos in earlier phenological stages, using the row side at the sun or capturing the inflorescence with enough perspective are, among others, actions easily achievable that could benefit the obtained results.

Additionally to the above mentioned, a rigorous study and comparison of different models for actual number of flowers per inflorescence estimation, using the number of flowers in an image as input information, is developed. As a result, suitability of the use of variety-dependent linear models previously pointed out in the literature has been discarded in favour of employing a unique variety-independent linear model. This issue constitutes an important discovery in this field, since it greatly generalises and simplifies the solution for estimating the actual flower number per inflorescence. Besides the classical option based on models created by means of linear regression, a non-linear estimation model has also been presented along with a promising set of descriptors. Results obtained with this approach outperform linear options. In spite of this, in the authors' opinion, this line needs further research before arriving at definitive conclusions. In effect, suitability of the developed feature space needs to be verified on a wider range of varieties. Moreover, once this is confirmed, it has to be assessed whether the gained accuracy compensates the utilisation of a more complex solution.

Acknowledgments

Authors would like to thank the ADER agency of the La Rioja regional government and the Spanish Ministry for Economy and Competitiveness in the funding of the projects VINETICS and AGL2011-23673, respectively. Authors are also grateful to the grapevine nursery Vitis Navarra for allowing us to take data used in the research described in this paper.

Appendix A. Mathematical background

Mathematical morphology is a nonlinear image processing used to extract structures of interest from the image. Comprehensive manuals about this technique can be found in Serra (1982) and Soille (2004). Nevertheless, for completeness purposes, a brief review of morphological operators used in this paper is carried out in this appendix.

Let f be a greyscale image. Image f is a mapping of a subset D_f of $\mathbb{Z}^2$, which is the definition domain of the image, into a bounded set of nonnegative integers N_0:

$$f: D_f \subset \mathbb{Z}^2 \rightarrow \{0, \ldots, t_{max}\}$$

where t_{max} is the maximum value of the data type used (e.g., 255 for 8-bit images, 1 for binary images, …). The complementary image of *f*, denoted as f^c, is defined for each pixel *x* as the maximum value of the data type used minus the value of the image *f* at pixel *x*:

$$f^c(x) = t_{max} - f(x)$$

The intersection of two greyscale images *f* and *g* is defined as

$$f \wedge g = \min[f(x), g(x)]$$

where *min* stands for the minimum operation. Similarly, the union of two images *f* and *g* would be

$$f \vee g = \max[f(x), g(x)]$$

being *max* the maximum operation.

The structuring element is a basic and essential tool in mathematical morphology used to study the morphology of objects in images. Mathematically, a structuring element is defined as a subset *B(x)* of $\mathbb{Z}^2$ centered at point *x*, whose shape is designed to describe shapes like circles, lines, diamonds, etc.

The morphological erosion of image *f* with structuring element *B*, $\varepsilon_B(f)$, is given by the expression:

$$[\varepsilon_B(f)](x) = \min_{b \in B} f(x+b)$$

Hence, it is the minimum value of the image in the neighbourhood defined by the structuring element when its origin is at *x*. The effect of erosion is expanding dark regions.

The dual operator of erosion is dilation. The morphological dilation of image *f* with structuring element *B*, $\delta_B(f)$, is defined as follows:

$$[\delta_B(f)](x) = \max_{b \in B} f(x+b)$$

Therefore, it is the maximum value of the image in the neighbourhood defined by the structuring element when its origin is at *x*. Dilation expands bright regions in the image.

Combining erosion and dilation, two new operators called opening (γ) and closing (φ), are obtained:

$$\gamma_B(f) = \delta_B(\varepsilon_B(f))$$

$$\varphi_B(f) = \varepsilon_B(\delta_B(f))$$

Opening removes those bright objects in the image that can be completely covered by the structuring element. Conversely, closing performs the dual operation, removing dark objects in the image completely covered by the structuring element.

Another interesting operator is the top-hat transformation. It emphasizes bright details in the image that are smaller than the structuring element *B*. Its formulation is:

$$TH_B(f) = f - \gamma_B(f)$$

Operators described are complemented by geodesic transformations. The geodesic dilation is the iterative unitary dilation of an image *f*, called marker, with respect to the mask *g*. Marker *f* must be contained within mask *g*. Mathematically speaking, the operator is defined as:

$$\delta_g^{(n)}(f) = \delta_g^{(1)}\left[\delta_g^{(n-1)}(f)\right], being\ \delta_g^{(1)}(f) = \delta_B(f) \wedge g$$

The morphological reconstruction by dilation of a mask image *g* from a marker image *f*, is the geodesic dilation of *f* with respect to *g* until idempotence. It is denoted by:

$$R_g^{\infty}(f) = \delta_g^{(i)}(f)$$

where *i* is such that:

$$\delta_g^{(i)}(f) = \delta_g^{(i+1)}(f)$$

Similarly, a partial reconstruction of a mask *g* from a marker *f* is calculated by performing *n* times the geodesic dilation of *f* with respect to *g*:

$$R_g^{n}(f) = \delta_g^{(n)}(f)$$

Using the geodesic reconstruction, a fill-hole operator can be defined. A hole in a greyscale image is defined as a set of connected points surrounded by connected components of value strictly higher than those in the hole. The following operator fills all holes in an image:

$$\psi(f) = [R_f^{\infty}(f_\vartheta^c)]^c$$

being f_ϑ the boundary image of *f* .

References

Nuske, S., Achar, S., Bates, T., Narasimhan, S., Singh, S., 2011. Yield estimation in vineyards by visual grape detection. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), New York, E.E.U.U., pp. 2352–2358.

Nuske, S., Wilshusen, K., Achar, S., Yoder, L., Narasimhan, S., Singh, S, 2014. Automated visual yield estimation in vineyards. Journal of Field Robotics 31 (5), pp. 837-860.

Font, D., Pallejà, T., Tresanchez, M., Teixidó, M., Martinez, D., Moreno, J., Palacín, J., 2014. Counting red grapes in vineyards by detecting specular spherical reflection peaks in RGB images obtained at night with artificial illumination. Computers and Electronics in Agriculture 108, pp. 105-111.

Diago, M.P., Correa, C., Millán, B., Barreiro, P., Valero, C., Tardáguila, J., 2012. Grapevine's yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions. Sensors 12, pp. 16988-17006.

Roscher, R., Herzog, K., Kunkel, A., Kicherer, A., Töpfer, R., Förstner, W., 2014. Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. Computers and Electronics in Agriculture 100, pp. 148-158.

Dunn, G.M., Martin, S.R., 2004. Yield prediction from digital image analysis: A technique with potential for vineyard assessments prior to harvest. Aust. J. Grape Wine Res 10, pp. 196–198.

May, P., 2004. Flowering and Fruitset in Grapevines. Lythrum Press, Adelaide.

Dry, P.R., Longbottom, M.L., McLoughlin, S., Johnson, T.E., Collins, C., 2010. Classification of reproductive performance of ten winegrape varieties. Aust J Grape Wine Res 16, pp. 47–55.

Galet, P., 1983. Precis De Viticulture. Dehan, Montpellier.

Carbonneau, A., Deloire, A., Jaillard, B., 2007. La Vigne: Physiologie, Terroir, Culture. Éditions Dunod, Paris.

May, P., 2000. From bud to berry, with special reference to inflorescence and bunch morphology in Vitis vinifera L. Aust J Grape Wine Res 6, pp. 82–98.

Keller, M., Kummer, M., Vasconcelos, M.C., 2001. Reproductive growth of grapevines in response to nitrogen supply and rootstock. Aust J GrapeWine Res 7, pp. 12–18.

Poni, S., Casalini, L., Bernizzoni, F., Civardi, S., Intrieri, C., 2006. Effects of early defoliation on shoot photosynthesis, yield components, and grape composition. Am J Enol Vitic 57, pp. 397–407.

Diago, M.P., Sanz-Garcia, A., Millan, B., Blasco, J., Tardaguila, J., 2014. Assessment of flower number per inflorescence in grapevine by image analysis under field conditions. J Sci Food Agric 94, pp. 1981–1987.

Coombe, B.G, 1995. Adoption of a system for identifying grapevine growth stages. Aus. J. Grape Wine Res. 1, pp. 104-110.

Serra, J., 1982. Image Analysis and Mathematical Morphology – Vol I. Academic, London, U.K.

Soille, P., 2004. Morphological Image Analysis – Principles and Applications – Second edition. Springer – Verlag, Berlin, Germany.

Otsu, N., 1978. A threshold selection method from gray-scale histogram. IEEE Trans. Syst. Man Cybern. 8, pp. 62–66.

Burt, P.J., 1981. Fast Filter Transforms for Image Processing. Computer Graphics and Image Processing 16, pp. 20–51.