Assessment of grape cluster yield components based on 3D descriptors using stereo vision

Ivorra, E.[1*]; Sánchez, A. J[1]; Camarasa, J.G.[1]; Diago M.P.[2] and Tardaguila, J.[2]

1 Departamento de Ingeniería de Sistemas y Automática, Universitat Politècnica de València. Camino de Vera s/n, 46022 Valencia, Spain

2 Instituto de Ciencias de la Vid y del Vino (Universidad de La Rioja, CSIC, Gobierno de La Rioja). 26006 Logroño. Spain.

*Author for correspondence: Eugenio Ivorra

Address: Edificio 8G - Acceso D - Planta 1
Ciudad Politècnica de la Innovación
Universitat Politécnica de Valencia
Camino de Vera, s/n
46022 VALENCIA – SPAIN
E-mail: euivmar@upvnet.upv.es

Phone : +34 686506624

Fax: + 34 96 387 98 16

Abstract

Wine quality depends mostly on the features of the grapes it is made from. Cluster and berry morphology are key factors in determining grape and wine quality. However, current practices for grapevine quality estimation require time-consuming destructive analysis or largely subjective judgment by experts.

The purpose of this paper is to propose a three-dimensional computer vision approach to assessing grape yield components based on new 3D descriptors. To achieve this, firstly a partial three-dimensional model of the grapevine cluster is extracted using stereo vision. After that a number of grapevine quality components are predicted using SVM models based on new 3D descriptors. Experiments confirm that this approach is capable of predicting the main cluster yield components, which are related to quality, such as cluster compactness and berry size ($R^2 > 0.80$, $p < 0.05$). In addition, other yield components: cluster volume, total berry weight and number of berries, were also estimated using SVM models, obtaining prediction R^2 of 0.82, 0.83 and 0.71, respectively.

Keywords: grape quality; cluster yield components; *Vitis vinifera* L; non-invasive technologies; stereo-vision; 3D descriptors;

1. Introduction

Due to the economic importance of the wine industry worldwide, innovative methods and technologies are being developed and applied to increase the quality of wine. The quality of wine is partially subjective, as it depends on the consumer's taste and preferences. However, it also depends on objective parameters. Of all the factors that influence the quality of the wine, the most important is the quality and features of the grapes it is made from.

Cluster compactness and berry size are two key factors of grapevine fruit quality. Current practices to assess these quality parameters require time-consuming destructive analysis or largely subjective judgments by experts. Unfortunately, in addition to the lack of objectivity of these parameters, the short and limited time available for analysis during harvest time and the lack of measurement tools as well as their high cost, among other factors, make it difficult to assess grape quality. For this reason, developing non-destructive grape supervision analysis, which increases the objectivity (Tello & Ibáñez, 2014) or automates the estimation (Roscher et al., 2014), represents a huge technological advance compared to the current practices.

Nowadays, 2D vision systems are widely used in the agri-food industry/agribusiness with proven results (Benlloch, Agustí, Sanchez, & Rodas, 1995; Brosnan & Sun, 2004; Diago, Sanz-Garcia, Millan, Blasco, & Tardaguila, 2014…). Moreover, huge advances in 3D sensor technology are leading to new opportunities and challenges. Specifically, one of the sectors which can clearly benefit from 3D computer vision technology is viticulture; for example, to forecast the quality of wine grapes (Whalley & Shanmuganathan, 2013).

1.1 Three-dimensional computer vision methods

There are many methods for extracting 3D information from 2D images. These methods can be classified into active and passive methods. Active methods emit light patterns into the scene to analyze their behavior, while passive methods only analyze the behavior of ambient light.

Some active methods are based on structured light (Udomkun, Nagle, Mahayothee, & Müller, 2014; Verdú et al., 2013). The simplest form of this is the projection of light stripes onto the surface, enabling the 3D position of the surface to be calculated from its intersection with the stripes.

Passive methods, such as stereo photogrammetry (Cyganek & Siebert, 2011), consist of two camera views of the surface from slightly different locations. Corresponding features are matched between the two images and the 3D surface is then constructed by triangulation. An advantage of photogrammetry over structured light is that the natural appearance of the surface is captured as a normal part of the process. A disadvantage is that it relies on there being enough features in the surface texture (visual appearance) for matching to take place.

Although stereoscopy is very simple and has a low cost when compared to other 3D techniques, not many studies using this technique for food inspection can be found in the available literature. This is mainly because automatic matching of points from stereo pairs is a difficult task for computer-based image analysis. However, there are some interesting works using stereoscopy for detecting weeds (Sachez & Marchant, 2000), measuring the thickness of wheat grains (Sun, Berman, Coward, & Osborne, 2007), estimating firmness in salmon

(Quevedo & Aguilera, 2010) and measuring the volume of apple slices in a drying process (Sampson, Chang, Rupasinghe, & Zaman, 2014). In all these works, edge features are used to solve the correspondence problem.

1.2 Grapevine quality components

Cluster architecture and berry morphology and distribution are key factors in determining grape quality. On the one hand, compact clusters show favorable conditions for the development of different grape pests and diseases (especially *Botrytis cinerea*). On the other hand, the number of interior berries increases with cluster compactness. These interior berries may not receive the sunlight needed to achieve an adequate phenolic maturity, leading to a heterogeneous ripeness of the cluster. At present, cluster compactness is visually estimated by experts using OIV descriptor No. 204 (OIV, 2007). This descriptor categorizes a cluster into one of five groups, quantified by 1, 3, 5, 7 and 9, where number 1 indicates “berries in grouped formation with many visible pedicels” and number 9 indicates “misshapen berries”. This estimation is subjective, but objective compactness estimations are required to allow comparisons between different works. Recently (Tello & Ibáñez, 2014) have evaluated some destructive methods for an objective estimation of grape cluster compactness.

Berry size is related to the skin-to-pulp ratio of the berry and the concentration of skin-located compounds that play a key role in wine quality. Two pieces of research have been published recently regarding the estimation of grape berry size. (Cubero et al., 2014) presents a method to estimate the size and weight of isolated wine-grapes taking into account the peduncle. (Roscher et al., 2014) detects circular structures on in-field images which are potentially berries and classifies them into the ‘berry’ or ‘non-berry’ class by utilizing a conditional random field. These two works estimate berry sizes from one image using a calibration object with a known size, which is located at a certain distance from the berries in the scene. Instead of this, it will be interesting to use a stereoscopic system to estimate berry size (without the constraint of always having a known calibration object in the scene).

This work contributes to accomplishing an objective and non-destructive estimation of the main grape yield components. To achieve this, a new approach has been developed to generate a partial three-dimensional model of the grape cluster using 3D computer vision technology. The approach is intended to be used alongside an inspection system consisting of two cameras arranged in a stereoscopic fashion. The system outputs two images of the grapes, one from each camera. These images, along with the camera calibrations, are then used to provide a partial 3D reconstruction of a grape cluster. Once this 3D model is obtained, 3D descriptors are automatically extracted to estimate the main cluster quality components and other yield components.

The experiments, obtained plant material and images are presented in Sections 2.1 and 2.2. In Section 2.3 the proposed 3D reconstruction approach is presented. Section 2.4 explains the 3D descriptors and the models proposed to estimate quality and yield components. The experiments and the results obtained are shown and discussed in Section 3. Finally, the conclusions of the paper are presented in Section 4.

2. Material and methods

2.1. Description of grapevine cluster samples

The study was carried out with 100 grapevine clusters from ten different red grapevine (*Vitis vinifera* L.) varieties: Grenache, Pinot Noir, Graciano, Monastrell, Mencia, Bobal, Cabernet Sauvignon, Tempranillo, Merlot and Carignan. All clusters were collected at Vitis Navarra Nursery vineyards (Navarra, Spain), just prior harvest in 2011.

At a first stage, cluster compactness was rated according to OIV descriptor No. 204 (OIV, 2007) by a panel formed by 10 experienced judges. Secondly, yield components, including the number of berries per cluster, cluster volume, and cluster and berry weight, were destructively measured at the University of La Rioja (Spain). The weight of each cluster was determined using a scale (Blauscal, AC-5000), and the morphological volumes were measured through the volume of water displaced by immersion in a bucket filled with water. Once clusters were destemmed by hand, the number of berries was counted. In addition, 15 berries per cluster were randomly chosen to measure their length and width using digital calipers (Mitutoyo, CD-15DCX). For the latter two features, the average of the 15 measures was used.

2.2. Image acquisition system

Images were taken using a Bumblebee2 stereo camera (Point Grey Research Inc, Richmond, BC, Canada) model BB2-08S2C-25 configured in automatic mode (white balance, gain and shutter). Images were synchronized and had a resolution of 1024 x 768 RGB color pixels from 0 to 255 values per channel.

Image acquisition was performed in lab conditions but the image acquisition set-up was designed to be near field conditions. The sample grape cluster was fixed from its peduncle, as if hanging from the vine (Fig. 1). The illuminants employed were four pairs of fluorescent tubes (Osram L 18W/965 BIOLUX) with a color temperature of 6500 K. They were distributed and oriented at different angles, which produced highlights on the sample (Fig. 2). The distance between illuminant and samples was around 0.35 m. It was decided that the light intensity should not be changed because it would have a lower effect than the color changes between samples. The camera can be placed at a distance of 0.10 m to 0.25 m from the sample. In this set-up, the furthest distance was chosen to work with the lowest resolution of the sample projection (around 170x250 pixels). As a result, this set-up resulted in a difficult image acquisition scenario.

A controlled background was needed to calculate one of the proposed 3D descriptors for assessing compactness.

Image acquisition was performed using the program PGR FlyCapture v.1.22 (Point Grey Research Inc, Richmond, BC, Canada).

2.3. Stereoscopic 3D reconstruction of grape clusters

The proposed approach is intended to be used alongside an inspection system consisting of two cameras arranged in a parallel stereoscopic fashion. The system outputs two images of the grape cluster, one from each camera. These images, along with the known camera calibrations, are then used to automatically obtain a partial 3D reconstruction of a grape cluster, which can be refined manually by a user.

Unfortunately, this approach has some inherent difficulties. Firstly, some of the grapes in the internal layers of the cluster are totally or partially hidden behind other berries. This fact impedes the reconstruction of the berries that are completely hidden. However the 3D descriptors can still be extracted from the reconstructed grape berries. This also means that the grape berries that are partially hidden will be harder to generate due to missing information. In addition, for it to be feasible to solve the problem, grape berries were modeled mathematically as perfect spheres. In fact, though, they may actually be slightly deformed, more akin to an ovoid, which may be somewhat troublesome in certain extreme cases.

In order to overcome these difficulties, several computer vision techniques were used. These included: i) the 3D implementation of the Hough transform (Woodford, Pham, Maki, Perbet, & Stenger, 2014), ii) feature detection and iii) feature matching. Also, the approach presented here involved extensive use of direct and inverse perspective projection at various stages of the automatic process.

A new 3D reconstruction tool has been developed in C++ using Microsoft Visual Studio 2010 Professional and the Qt 4.8.0 UI framework. Image processing was performed using the openCV library v.2.3.1. This 3D reconstruction tool provides a clearly defined workflow, including an automatic approach and a manual refinement part. In the automatic phase, the tool generates as many correct grapes as possible without user supervision. The effectiveness of this method varies greatly on a case by case basis, as some grape varieties may be easier to reconstruct than others, depending primarily on how prominent the difficulties discussed above are.

2.3.1. Automated approach

This approach automatically generates a partial three-dimensional model of the grape berries without the user's help. In order to do so, the following problems were solved.

Firstly, the original images were rectified to accomplish a parallel camera configuration, where the baseline (b = 0.1197 m) was aligned with the Xs-axis and therefore the epipolar constraints could be easily applied to solve the correspondence problem. In this case, only vertical edge features were used, since the horizontal edges belong to the same epipolar line and this increases the risk of obtaining wrong correspondences.

Edges are features which fit perfectly in scenarios with a high degree of variability in color or illumination. In this case, vertical edges were extracted from the rectified images and for each vertical edge in the rectified images a vector of descriptors was estimated. This vector consists of the gradient module (m_i), the sines (s_i) and the cosines (c_i) of the gradient orientation. The range of distances allowed between the camera and the sample, described at image acquisition set-up as shown in Fig. 1, achieved accuracy with reconstruction errors smaller than 1 mm at occlusion edges. Therefore, the occlusion edges of the berries can be used to solve the reconstruction problem obtaining low errors.

The measure of similarity between features in the left and right images was calculated by the Euclidean distance between their normalized vector of descriptors [m_i, s_i, c_i]. Feature matching consists of an optimization problem where features in a row in the left image must be matched with features in the same row in the right image. This correspondence problem was solved using dynamic programming, where a tree of possible solutions is explored. A node of this tree contains two column indexes (one for each image), a list of matching pairs and the correspondence cost. The two indexes pointed to the features that the algorithm was trying to correspond in this node (Fig. 3).

Each node generated three new branches. The first branch took into account the case when the feature in the left image was occluded and therefore did not correspond to any feature in the right image. In this case, the son node incremented its left index and added a constant to its correspondence cost. The second branch considered the case when the feature in the left image matched with the feature in the right image. In this case the node incremented its two indexes, included the new matching pair in the list and added the similarity distance to its correspondence cost. And finally, the third branch assumed that the feature in the right image was occluded and therefore did not correspond to any feature in the left image. In this last case, the node incremented its right index and added a constant to its correspondence cost.

The algorithm expanded the tree and found the optimal solution, which was the list of matching pairs on the leaf node with the minimum cost.

For each matching pair of vertical edges, a 3D point was obtained by triangulation. In addition, others features (not only vertical edges) could be reconstructed. For example, points between berries, which have a dark appearance in the two images. The 3D points, reconstructed using vertical edges, were the input to a Hough transform which had been adapted to detect spheres in the four dimensional parameter space (x_c, y_c, z_c, r). After the accumulation phase, a minimum number of votes were set for detecting spheres. A refinement step was performed in order to detect and remove some wrong spheres, such as overlapping spheres or spheres which include reconstructed points between berries. It is important to mention that Hough transform is a robust technique against possible noise like highlights.

2.3.2. Manual refinement

In the manual phase, the user could reconstruct additional berries using the 3D reconstruction tool. This could be achieved by two methods. The first required the user to input the coordinates (position and radius) of the grape. While this method certainly worked, it may be somewhat unintuitive and slow. For this reason, an alternative, more visual method was designed and developed. In this case, in order to reconstruct a grape the user just needed to specify 5 points of its border in one of the original images and its center in the other one. With these data, the tool generated a 3D grape berry automatically, without the need for any further action. This enabled the user to reconstruct a grape simply by entering a total of 6 points in a totally visual, intuitive and quick way. Furthermore, the tool assisted this process by providing the user with a visual aid in the form of previsualizations of the final result in both of the original images and axis restrictions whenever possible. With this method the user would not need to be an expert to use the tool, as the process is simple, easy and straightforward.

One of the objectives of this work was to make the 3D reconstruction tool as intuitive as possible. Therefore, the tool has a graphical interface (GUI) that enables the user to visualize the model and interact with it. The GUI that can be seen in Fig. 4 and it is divided into two main parts: the 2D viewports and the 3D viewport.

On the one hand, there is a 2D viewport, where all the information related to the 2D space, such as the original images of the grape, is shown. As mentioned above, the reconstruction process takes two images of the grape as input; therefore, the tool has two 2D viewports, one for each camera. These viewports can also be used by the user to enter input to the tool, such as the 6 points required in the manual reconstruction phase.

A 2D viewport also has some features that are not directly related to the reconstruction process but that can greatly help the user. The most important of these features is the projection of the model onto the original images. The result is a new image which consists of the original image with the projection of the current model drawn over it. With this new image the user can easily check if the model correctly fits the grapes in the original images.

On the other hand, the 3D viewport provides a fully three-dimensional visualization of the current model. The user can select individual grapes and edit their parameters, as well as navigate through the 3D space. The 3D viewport is immediately updated with new information to reflect the current state of the model. This means that, whenever the user adds a new grape using the system explained above, the 3D viewport will show the changes on the fly. The same applies to the automatic phase: the moment it finishes its execution, the 3D viewport will be updated.

Finally, the reconstructed model can be saved in an open, easily-readable format, meaning that the models can be easily used by other applications that want to display or perform certain operations on them.

2.4. 3D descriptors and models

There are many morphological descriptors for the characterization of the grapevine cluster which can be measured manually. However, with the cluster 3D model, non-destructive descriptors can be calculated automatically (see Table 1). In this work, six 3D descriptors were extracted automatically from the 3D model: cluster volume, berry size, number of berries per cluster, concavity measure, intersection between berries, and number of berries per area. The last 3D descriptors are new descriptors proposed in this work to assess grapevine cluster compactness that cannot be measured by hand.

In order to calculate the concavity measurement (holes between berries) an image segmentation should first be performed. This segmentation was performed in two steps: firstly, the definition of a region of interest (ROI) and secondly, classification of pixels into three classes: background, holes and berries.

The first step, ROI definition, calculates the difference between the convex hull of the cluster and the berries from the 3D model. This difference estimates the potential concavities. These estimated potential concavities were extracted from the 2D images based on the projection of the 3D model. The convex hull of the cluster was calculated for the image based on a 2D Delaunay triangulation. This triangulation was calculated from the projected sphere centers of the 3D model. Then, using the triangles obtained, a convex hull was created. However, some triangles from this convex hull were removed to ensure a better fit to the projected shape of the cluster, specifically focusing on removing large concavities. This refinement consisted of removing the triangles that had at least one vertex belonging to the convex hull and in which the longest edges are three times longer than their heights. Pixels inside this refined convex hull were used as the first region. The second part of this step was to make a second region as the union of the ellipses from the projection of the 3D model. Finally, the ROI was defined as the difference between the first region and the second region.

Fig. 5 shows an example of this step on the Bobal sample 1. The projected berries can be seen in green, the Delaunay triangulation in blue, the convex hull in red and the removed triangles in purple. In Fig. 6, pixels inside the ROI were colored in red.

The second step was the color segmentation of the pixels in the ROI into 3 classes. This pixel classification was based on the *k*-nearest-neighbors pattern recognition approach (Sánchez, Albarracin, Grau, Ricolfe, & Barat, 2008). Five different grape colors were self-trained for each pair of images using a k-means clustering of the pixel colors inside the projected ellipses of the 3D model. Five colors were also trained for the background determination. Then, for each pixel inside the ROI a difference to these trained colors was calculated. If the lowest difference in absolute values was with a background color, then it was classed as background. If the lowest difference was with a self-trained grape color, then if the difference was positive (pixel color was brighter than self-trained grape color) it was classified as grape color, otherwise it was classified as a hole with the difference as an approximate measure of the depth.

The depth value for all the hole pixels was accumulated plus the accumulated value of all the background pixels inside the ROI, taking their depth as 255. The mean of both images was calculated for each sample. This value was called the concavity measure and it is the only 3D descriptor that requires a controlled background.

The intersection between spheres was approximated using the Eq. 1 where *c* is the center of the sphere, *r* its radius and *n* the total number of grapes in the 3D model.

$$d = \begin{cases} \left| \|c_1 - c_2\| - r_1 - r_2 \right|, & if\ \|c_1 - c_2\| - r_1 - r_2 < 0 \\ 0, & if\ \|c_1 - c_2\| - r_1 - r_2 \geq 0 \end{cases}$$

$$I = d * d * \min(r_1, r_2) \qquad (1)$$

$$Im = \frac{1}{n} * \sum_{i=1}^{n} I_i$$

The number of berries per area was approximated as the area of the ROI divided by the number of berries. This value was averaged using the stereo pair of images.

The cluster volume was calculated using the information from the 3D model following this equation (Eq. 2).

$$V = \sum_{i=1}^{n} \frac{4 * \pi * {r_i}^3}{3} \qquad (2)$$

Another value extracted from the 3D model was the berry size as the mean volume of the grape berries in the 3D model (Eq. 3).

$$V_m = \frac{1}{n} * V \qquad (3)$$

The number of berries was the number of reconstructed berries in the 3D model for each sample.

The 3D descriptors calculations were performed using our own code developed using the image processing toolbox of Matlab R2008a (The Mathworks, Natick, Massachusetts, USA) and loading the previously reconstructed 3D models.

2.5. Statistical analysis

Kendall’s Tau-b (Kendall, 1970) correlation coefficients were calculated between the 3D descriptors and the compactness measured by the visual evaluation panel.

Support Vector Machine regression models (using nu-support vector regression) (Schölkopf, Smola, Williamson, & Bartlett, 2000) were used to predict compactness and some morphological components (berry size, cluster volume, cluster weight and number of berries) using the calculated 3D descriptors as input data. The model is based on a number of support vectors (samples selected from the calibration set) restricted by the nu parameter and non-linear model coefficients which define the non-linear mapping of variables (calculated 3D descriptors).

The samples were pseudo randomly divided 2/3 for building the model and 1/3 for testing it. It was adjusted so that all the different grape species were represented for building and testing.

To estimate the fit of the calibration data with the models developed, the Root-Mean-Square Error of Calibration (RMSEC) was used. It is defined as:

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \qquad (4)$$

Where $\hat{y}_i$ are the values of the predicted variable when all samples are included in the model formation, y_i are the known values and n is the total number of samples.

A random cross-validation method was employed to evaluate the models developed for the samples used in the calibration model. In this method, subsets of 8 (n/8) random samples are used to test the model developed without them. This method was iterated three times with different samples and its results averaged to achieve more reliable validations.

The Root-Mean-Square Error of Cross-Validation (RMSECV) was used to evaluate and compare the accuracy of the different SVM models developed using the random cross-validation method described previously. RMSECV is based on Eq. 4, but in this case the parameter $\widehat{y_i}$ was the value of the variable estimated using a model that was built without the removed group of samples ($i \in RemovedSamples$).

The prediction accuracy of the model was estimated using the root-mean-square error of prediction (RMSEP). RMSEP is calculated exactly as in Eq.4, except that the estimates $\hat{y}_i$ refer to 33 new samples which were not involved in model building.

All statistical procedures were performed on PLS Toolbox 6.5 (Eigenvector Research Inc., Wenatchee, Washington, USA), a toolbox extension within the Matlab 7.6 computational environment (The Mathworks, Natick, Massachusetts, USA).

3. Results and discussion

The efficiency of the 3D reconstruction tool was tested by generating the 3D model of 100 clusters of grapes of different varieties. All models were successfully generated using a combination of the automatic and manual phase (Fig. 7). The process of obtaining an almost exact reconstruction of the visible part of the cluster never took more than 10 minutes, often less depending on the complexity of the cluster.

Applying only the automatic reconstruction phase, the mean success rate was 20% correctly reconstructed berries per view (approximately 10 berries) (Fig. 8). The lowest rate was 10% when the automatic phase was applied to “Tempranillo” clusters. The rest of the varieties provided better reconstruction ratios, the best case being “Bobal”, with 27.78%.

Using the aforementioned features of the 2D viewport the model developed was checked against the original images of the cluster, obtaining successful results with projected errors lower than 5 pixels (Fig. 9). Taking into account the acquisition set-up defined in Fig. 1, it was proven that if the error of projected edges is lower than 5 pixels, the reconstruction error will be lower than 1 mm. The only visible grapes that could not be generated were the ones that: a) were visible only in one image and hidden in the other, thus making it impossible to generate a 3D version of it; or b) could not be properly identified because of visual noise. All the other visible grapes could be reconstructed with minimal error.

At this point, it is important to highlight the advantages of having a digitalized, 3D model of the grapes. The most important one is the fact that the desired 3D descriptors can be extracted automatically, of course. However, it is important to note that it also allowed the model to be saved easily, virtually forever.

The benefits of this implication are significant because it is possible to create a record of models of different wine varieties over the years. Moreover, it is also possible to recover a previously analyzed model and extract new descriptors for it that were not needed before.

To correctly assess these benefits one can compare it to the traditional system where only measures or images can be stored. This means that in order to analyze the grape clusters with new criteria it is necessary to perform the measurement again. Apart from the time needed to do this, those grape clusters may not be available any more.

3.1. Cluster quality components

A Kendall tau study was performed to correlate with compactness for the new 3D descriptors concavity measure, intersection between berries and number of berries per area. It was also calculated for the index CI-12, which (Tello & Ibáñez, 2014) claims can be used for compactness measures. As it can be seen in Table 2, the CI-12 result confirmed a good correlation, as reported by the authors. The concavity measure descriptor presented the highest Pearson correlation coefficient, whereas the number of berries per area was poorly correlated.

Fig. 10 and Fig. 11 depict the image segmentation results of two samples of Cabernet Franc with two different classes of compactness 3, a loose cluster, and 7, corresponding to a tight cluster. Hole depths inside the cluster were coded as blue being deeper the higher the value. As it can be seen, there were significantly more background and hole pixels in Fig. 10 than in Fig. 11. This information was collected by the previously defined concavity measurement. Therefore, it seems logical that there was a high correlation value with the compactness obtained with the Kendall tau study.

The results of the calibrated SVM model for predicting compactness using these three 3D descriptors: concavity measure, intersection between berries and number of berries per area, can be seen in Table 3. The inclusion of the 3D descriptors intersection between berries and number of berries per area in the SVM model for grape cluster compactness enhanced the results by around 10% in comparison with using the concavity measure alone. Although the Kendal tau results in a poor direct correlation for the 3D descriptor of number of berries per area (Table 2), it helped to model some extreme cases, which explains the better results when it is used.

It should be noted that only one model was generated for ten different grapevine cultivars, which increases the complexity of the problem. Prediction results were moderate (R^2 Pred = 0.80), but it can be observed in Fig. 12 that the maximum error for test samples was around ± 2 so it was misclassified in the worst case as one class above or below.

Another SVM model was built regarding the berry size quality component. The results (Table 3) were good, with an error in calibration and prediction of less than 0.23 cm^3 and an R^2 above 0.82. The model built was very robust because the samples used for the prediction fit even better than those used in the calibration (RMSEP< RMSEC). It should be mentioned, however, that for the compactness descriptors, this 3D descriptor was calculated based only on the automatic reconstruction of the 3D model and it did not need a controlled background.

3.2. Other cluster yield components

Three different SVM models were also built for each of the morphological descriptors estimated (cluster volume, total berry weight and number of berries). These models were built using the same distribution of samples for calibration and testing as those used for the quality components. Just one 3D descriptor was used for each yield component (the 3D descriptor with the same name as the yield component, except for weight, which was calculated with the cluster volume). As can be seen in Table 3, the results for cluster volume and total berry weight showed a determination coefficient (R^2) in prediction higher than 0.82. For the manual measurement of cluster volume, the complete cluster (including pedicels and branches) was measured, while for the image method employed, only the information about the reconstructed berries was used, so it seems coherent that a better result was obtained using berry weight alone. The outcomes were lower for number of berries (approx. 24 berries of error in prediction) probably because many berries were occluded inside the cluster (especially for the samples with greater compactness).

4. Conclusions

The research focused on estimating grape yield components using stereo vision. Due to the difficulty of making the correct correspondence between the pair of images, a 3D reconstruction tool was developed to obtain an accurate 3D model of the samples with a reconstruction error of less than 1 mm. It has been proven that the tool actually achieves the goal of reconstructing the model successfully while complying with the standards of usability, speed and user-friendliness. In fact, using the 3D models obtained only from the automatic approach, berry size was estimated by an SVM model with an R^2 in prediction higher than 0.82. Once three-dimensional models were available, new 3D descriptors were extracted from them to assess the compactness quality component: concavity measure, intersection between berries and number of berries per area. These descriptors were evaluated and compared with the state-of-the-art index CI-12 on 100 different samples from 10 different varieties. The concavity measure descriptor gave a Kendal tau correlation of -0.71 ($p<0.01$) compared with the 0.52 ($p<0.01$) obtained by CI-12. An SVM was developed using these new 3D descriptors with an R^2 in prediction higher than 0.80, with a maximum classification error of one class. In addition, other yield components: cluster volume, total berry weight and number of berries, were estimated using SVM models, obtaining the following R^2 in prediction, respectively: 0.82, 0.83 and 0.71.

The results achieved show the capability of this technique for solving the problem of having an accurate and objective tool for measuring cluster compactness. In addition, the fact that the berry size quality component was estimated automatically without a controlled background made this technique very feasible for use under field conditions. It would be very useful, for example, to assess grapevine quality components in inter varietal studies (such as genetic association studies).

Grape clusters were measured in a difficult image acquisition scenario with the aim of simulating field conditions. Future work will focus on extrapolating these results into the field scenario.

Acknowledgements

This work has been partially funded by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria de España (INIA – Spanish National Institute for Agriculture and Food Research and Technology) through research project RTA2012-00062-C04-02, support of European FEDER funds, UPV-SP20120276 and AGL2011-23673 project.

References

Benlloch, J., Agustí, M., Sanchez, A., & Rodas, A. (1995). Colour segmentation techniques for detecting weed patches in cereal crops. *Proc. of the 4th Workshop on Robotics in Agriculture and the Food-Industry*, 71–81.

Brosnan, T., & Sun, D.-W. (2004). Improving quality inspection of food products by computer vision––a review. *Journal of Food Engineering*, *61*(1), 3 – 16. doi:http://dx.doi.org/10.1016/S0260-8774(03)00183-3

Cubero, S., Diago, M. P., Blasco, J., Tardáguila, J., Millán, B., & Aleixos, N. (2014). A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis. *Biosystems Engineering*, *117*(0), 62–72. doi:10.1016/j.biosystemseng.2013.06.007

Cyganek, B., & Siebert, J. P. (2011). *An introduction to 3D computer vision techniques and algorithms*. John Wiley & Sons.

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. *Journal of the Science of Food and Agriculture*, *94*(10), 1981–1987. doi:10.1002/jsfa.6512

Kendall, M. (1970). Rank Correlation Methods (4th edn.) Griffin. *London, UK*.

Quevedo, R., & Aguilera, J. M. (2010). Computer Vision and Stereoscopy for Estimating Firmness in the Salmon (Salmon salar) Fillets. *Food and*

Bioprocess Technology, *3*(4), 561–567. doi:10.1007/s11947-008-0097-3

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*, 148–158.

Sachez, A. J., & Marchant, J. A. (2000). Fusing 3D information for crop/weeds classification. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on* (Vol. 4, pp. 295–298 vol.4). doi:10.1109/ICPR.2000.902917

Sampson, D. J., Chang, Y. K., Rupasinghe, H. P. V., & Zaman, Q. U. (2014). A dual-view computer-vision system for volume and image texture analysis in multiple apple slices drying. *Journal of Food Engineering*, *127*(0), 49–57. doi:10.1016/j.jfoodeng.2013.11.016

Sánchez, A. J., Albarracin, W., Grau, R., Ricolfe, C., & Barat, J. M. (2008). Control of ham salting by using image segmentation. *Food Control*, *19*(2), 135–142. doi:10.1016/j.foodcont.2007.02.012

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New Support Vector Algorithms. *Neural Computation*, *12*(5), 1207–1245. doi:10.1162/089976600300015565

Sun, C., Berman, M., Coward, D., & Osborne, B. (2007). Thickness measurement and crease detection of wheat grains using stereo vision. *Pattern Recognition Letters*, *28*(12), 1501–1508. doi:10.1016/j.patrec.2007.03.008

Tello, J., & Ibáñez, J. (2014). Evaluation of indexes for the quantitative and objective estimation of grapevine bunch compactness. Vitis - , 53 (1), pp. 9-16. *Journal of Grapevine Research*, *53*, 9–16.

Udomkun, P., Nagle, M., Mahayothee, B., & Müller, J. (2014). Laser-based imaging system for non-invasive monitoring of quality changes of papaya during drying. *Food Control*, *42*(0), 225–233. doi:10.1016/j.foodcont.2014.02.010

Verdú, S., Ivorra, E., Sánchez, A. J., Girón, J., Barat, J. M., & Grau, R. (2013). Comparison of TOF and SL techniques for in-line measurement of food item volume using animal and vegetable tissues. *Food Control*, *33*(1), 221–226. doi:10.1016/j.foodcont.2013.02.031

Whalley, J., & Shanmuganathan, S. (2013). Applications of image processing in viticulture: A review.

Woodford, O., Pham, M.-T., Maki, A., Perbet, F., & Stenger, B. (2014). Demisting the Hough Transform for 3D Shape Recognition and Registration. *International Journal of Computer Vision*, *106*(3), 332–341. doi:10.1007/s11263-013-0623-2

Table01

Descriptors	Morphological (measured by hand)	3D (automatic measures)	Units
Cluster length	Non-destructive	Non-destructive	$m*10^{-2}$
Cluster width	Non-destructive	Non-destructive	$m*10^{-2}$
Cluster volume	Non-destructive	Non-destructive	$m^3 *10^{-3}$
Cluster weight	Non-destructive	Estimated	$kg*10^{-3}$
Berry size	Non-destructive	Non-destructive	$m^3 *10^{-6}$
Number of berries per cluster	Destructive	Estimated	
Seeds per berry	Destructive	Occluded	
Pedicel length	Destructive	Occluded	$m*10^{-2}$
Rachis weight	Destructive	Occluded	$kg*10^{-3}$
First to seventh rachis node length	Destructive	Occluded	$m*10^{-2}$
Ramifications per cluster	Destructive	Occluded	
Concavity measure	Not possible	Non-destructive	
Intersection between berries	Not possible	Non-destructive	m^3*10^{-6}
Number of berries per area	Not possible	Non-destructive	

Table 1 Morphological and 3D descriptors

Table 02

Compactness descriptors	Rho
CI-12: $\mathrm{B.weight}/\mathrm{B.length}^2$	0.530**
Concavity measure	-0.710**
Intersection between berries	0.569**
Number of berries per area	-0.205*

Table 2 Kendall tau results for compactness descriptors **p<0.01 *p<0.05

Table 03

	Number of SVs	RMSEC	RMSECV	RMSEP	R^2 Cal	R^2CV	R^2 Pred
Compactness	20	0.920	1.139	0.817	0.886	0.826	0.808
Berry size	16	0.214	0.227	0.180	0.842	0.820	0.830
Cluster volume	36	28.204	46.945	56.798	0.944	0.845	0.822
Berry weight	42	19.549	36.672	44.656	0.966	0.880	0.830
Number of berries	23	13.307	23.821	23.791	0.947	0.826	0.714

Table 3 SVM results of the cluster components

Figure 01

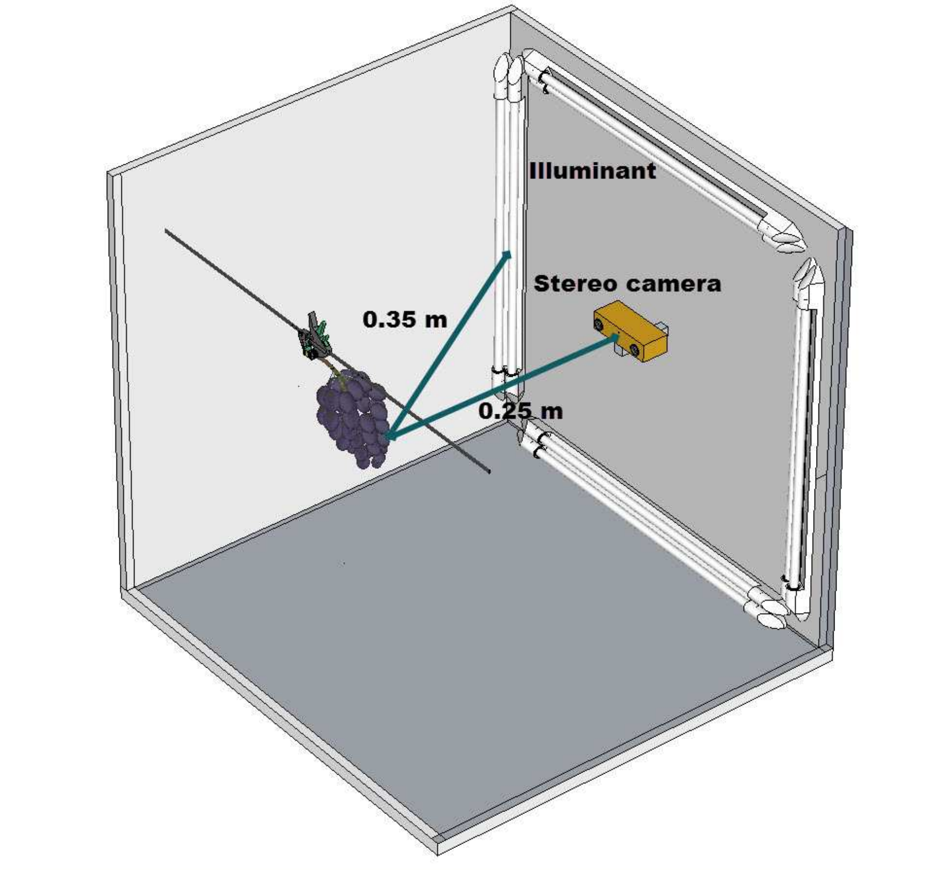


Figure 02

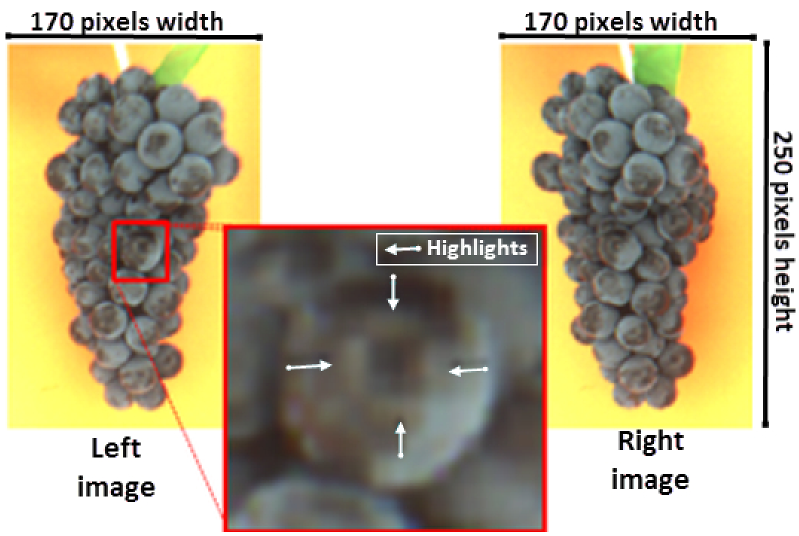

170 pixels width
170 pixels width
250 pixels height
Highlights
Left image
Right image

Figure 03

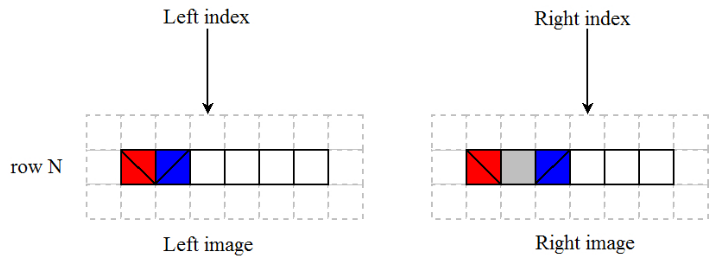
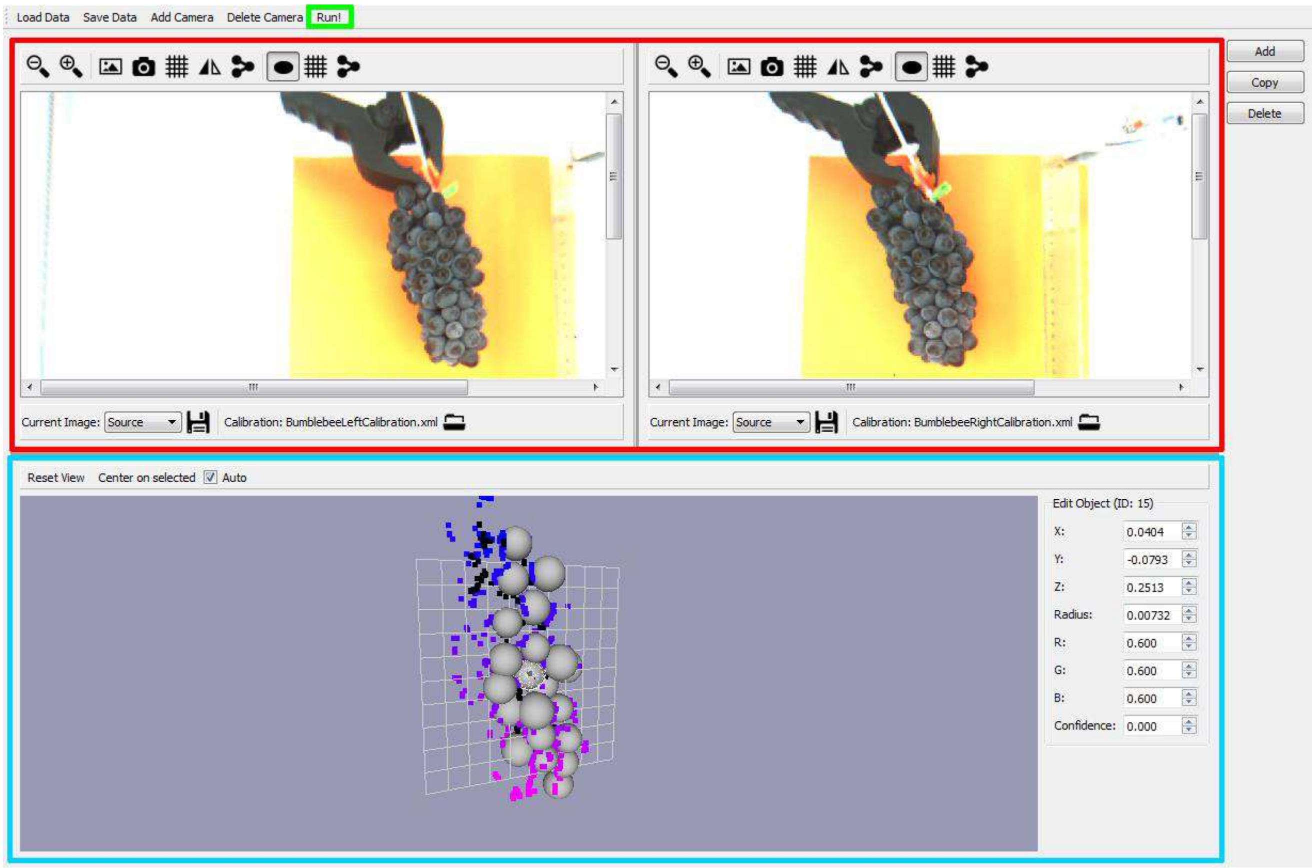


Figure 04



Load Data
Save Data
Add Camera
Delete Camera
Run!
Add
Copy
Delete
Current Image: Source
Calibration: BumblebeeLeftCalibration.xml
Current Image: Source
Calibration: BumblebeeRightCalibration.xml
Reset View
Center on selected
Auto
Edit Object (ID: 15)
X: 0.0404
Y: -0.0793
Z: 0.2513
Radius: 0.00732
R: 0.600
G: 0.600
B: 0.600
Confidence: 0.000

Figure 05

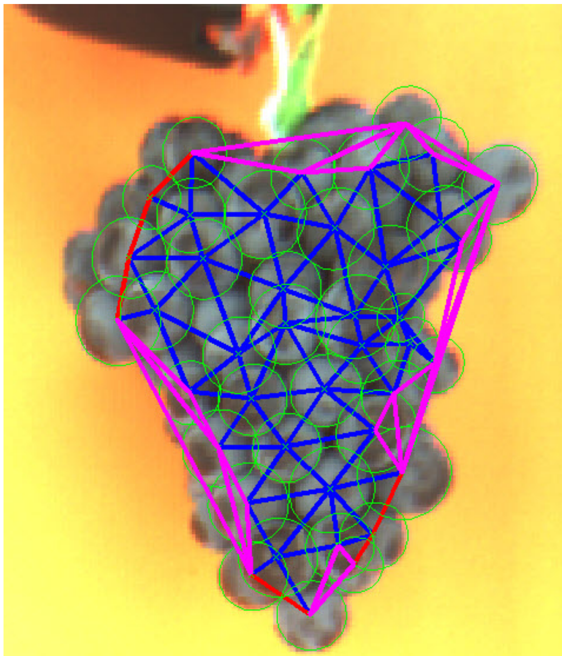


Figure 06

Figure 07

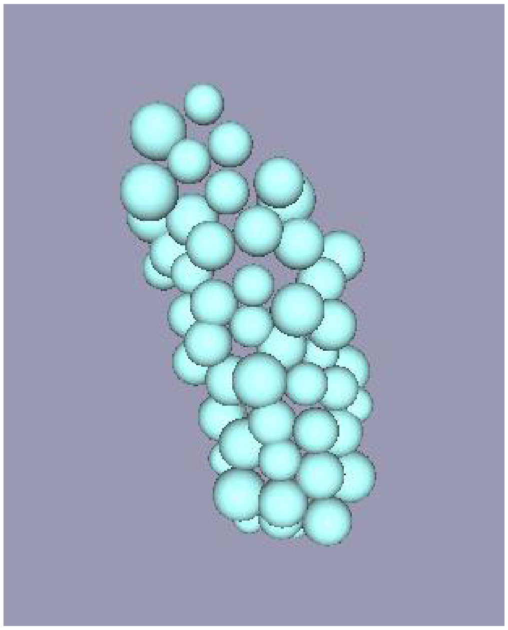


Figure 08

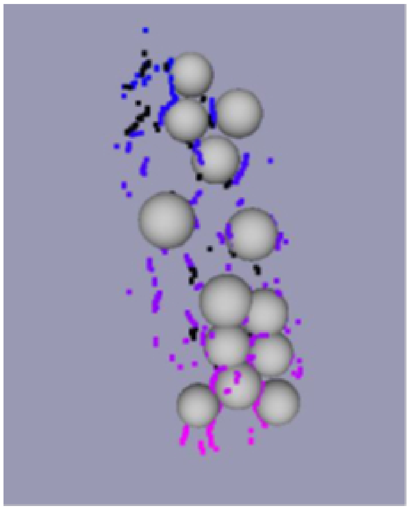


Figure 09

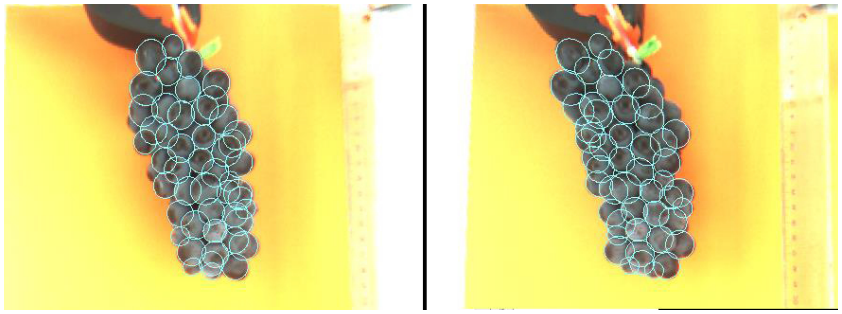


Figure 10

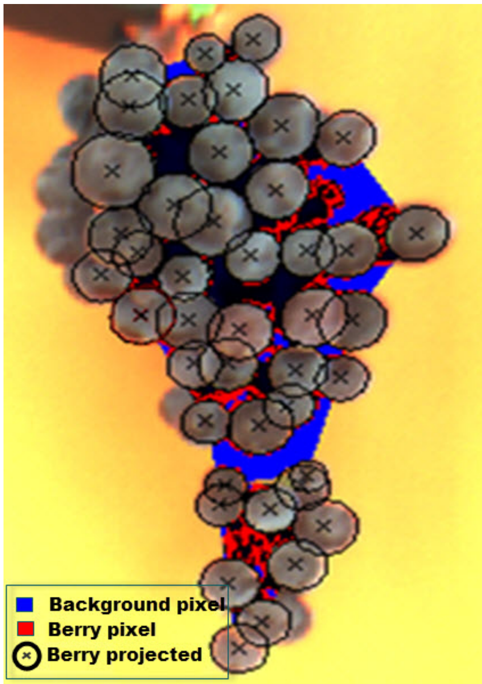


Figure 11

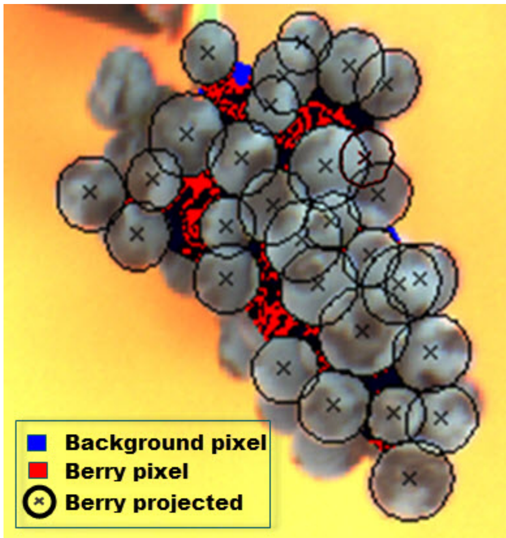
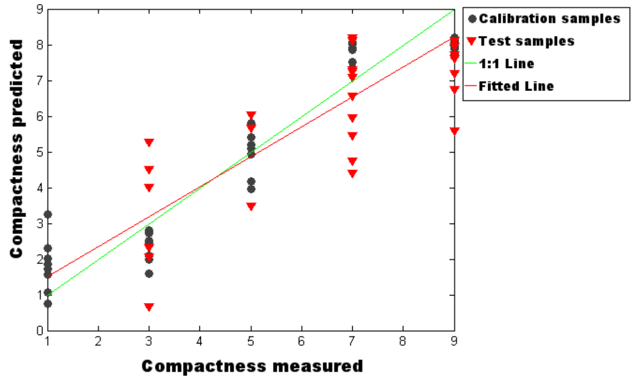


Figure 12



Calibration samples
Test samples
1:1 Line
Fitted Line
Compactness predicted
Compactness measured

Fig. 01 Image acquisition set-up.
Fig. 02 Images captured with low resolution and highlights (Bobal sample 6)
Fig. 03 A node example. A pair of features with the same color represents a matching pair. The gray color represents an occluded feature.
Fig. 04 3D reconstruction tool interface. The 2D viewports are marked in red. The 3D viewport is marked in blue.
Fig. 05 Convex hull created by Delaunay triangulation.
Fig. 06 Region of interest selected for pixel classification.
Fig. 07 3D model refined by hand.
Fig. 08 Berries detected using Hough transform. Color points are the 3D points extracted from the vertical edges. Black points represent 3D points between berries.
Fig. 09 Checking the 3D model on the rectified images.
Fig. 10 Image segmentation of sample 2 Cabernet Franc (Compactness class 3).
Fig. 11 Image segmentation of sample 2 Cabernet Franc (Compactness class 7).
Fig. 12 SVM compactness results.

Table 1 Morphological and 3D descriptors
Table 2 Kendall tau results for compactness descriptors $**p<0.01$ $*p<0.05$
Table 3 SVM results of the cluster components