Computers and Electronics in Agriculture 110 (2015) 17–26

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Vineyard yield estimation by automatic 3D bunch modelling in field conditions

Mónica Herrero-Huerta [a,*], Diego González-Aguilera [a], Pablo Rodriguez-Gonzalvez [a], David Hernández-López [b]

[a] Department of Cartographic and Land Engineering, University of Salamanca, Higher Polytechnic School of Avila, Hornos Caleros 50, 05003 Avila, Spain
[b] Institute for Regional Development (IDR), University of Castilla La Mancha, 02071 Albacete, Spain

ARTICLE INFO

Article history:
Received 26 May 2014
Received in revised form 7 August 2014
Accepted 1 October 2014

Keywords:
Yield estimation
Photogrammetry
Computer vision
Vitis vinifera L.
Automation
Non-invasive technologies

ABSTRACT

This manuscript focuses on developing a workflow for determining the productivity of vineyards in a novel and innovative way, ensuring flexibility and simplicity in data acquisition, automation in the process and high-quality results, using low cost sensors. The non-invasive system proposed allows the determination of yield at cluster level by combining close-range photogrammetry and computer vision. Bunches are reconstructed in 3D from images processed with Photogrammetry Workbench software (PW) developed by the authors. Algorithms and techniques were combined to estimate the most relevant parameters in the productivity of a vineyard: volume, mass and number of berries per bunch. To validate the workflow proposed, a sample of laboratory tests based on dimensional analysis of the clusters together with the single count of berries, were analyzed to establish the groundtruth. The results achieved from the scaled models and different estimation parameters were contrasted. The results confirm the feasibility of the proposed methodology, providing scalability to a comprehensive analysis of the productivity of the vineyard and affording a constant operational improvement and proactive management.

1. Introduction

The development of innovative technologies in viticulture to monitor vineyards is encouraged because of their huge environmental and economic impact on society. Computer vision systems are highly suitable for this purpose because they are a non-contact and non-destructive technique (Chherawala et al., 2006). The application scenarios of digital image analysis cover yield estimation, quality evaluation, disease detection and grape phenology (Whalley and Shanmuganathan, 2013). One of the main concerns in the wine industry along history has been the accurate and objective estimation of the yield and the oenological potential of vineyards: *yield forecasting* or harvest forecasting. Currently, this is still an unresolved issue and is of great technical and economic importance (Wolpert and Vilas, 1992; Clingeleffer et al., 2001; Dunn and Martin, 2004). A precise vineyard yield estimation allows more efficient grapevines to be obtained, their qualitative potential to be established, and the production of higher-quality wines (Dunn and Martin, 2003). Roby et al. (2004) checked that berry size is a determining factor for winegrape quality where the berry small size is related to lack of water and the variation in berry size indicates an inhomogeneous maturation. Also, this estimation influences decision-making techniques, such as the execution of vineyard score sheets.

Currently, the methods applied at industrial scale to estimate the productivity of vineyards are destructive, labour- and time-demanding (Martin et al., 2003) and, therefore, economically non-viable. Furthermore, several methods are based mostly on a visual inspection of the vineyard so the final data cannot guarantee reliability and accuracy. In light of this, a specific software (Grape Forecaster) was developed (Martin et al., 2003) to calculate the final harvest, allowing only historical information on yield components and variations from year to year to be collected. Blom and Tarara (2009) proposed a method based on the tension of the wires of the conduction system, which is expensive. Grape yield assessment was studied from the point of view of water availability and its effects on yield and berry quality attributes (Serrano et al., 2012) by expensive measuring techniques. For these reasons, in recent years image analysis has begun to be applied in viticulture in attempts to assess the vegetative state or performance of vineyards in some countries such as Australia

* Corresponding author. Tel.: +34 920353500x3820.
E-mail address: monicaherrero@usal.es (M. Herrero-Huerta).

http://dx.doi.org/10.1016/j.compag.2014.10.003
0168-1699/

(Dunn and Martin, 2004) or the U.S. (Nuske et al., 2011). More concretely, Dunn and Martin (2004) captured in field images using a white screen behind the canopy and extracted the colour features to classify the berry clusters, using manual thresholds and tolerances for the segmentation. The work of Nuske et al. (2011) involves a visible light camera boarded in a small vehicle to survey the entire vineyard. The berries are detected based on a 2D radial symmetry transform to extract their center, and then are classified based on their colour and texture. Finally a clustering is applied to remove false positives. In a recent work (Nuske et al., 2012) the use of calibration data from prior harvest data enhanced the previous results. Moreover, 2D computer vision techniques have also been applied to individual strains for the identification of plant elements (Herrero Langreo et al., 2010) or counting individual berries using a flatbed scanner (Battany, 2008). A 2D grapevine yield and leaf area estimation was done by Diago et al. (2012), who used a visible light camera to capture images in-field using a white screen behind the canopy. Their approach involves the computation of Mahalanobis colour distance for a supervised classification application. Approaches to obtaining 3D models from photographs for plant analyses were carried out at a laboratory with individual tomato plants (Aguilar et al., 2008), but without effective automation and failing to translate this technology to field conditions. More recently, Djuricic et al. (2014) employed an active sensor, multi-echo laser scanner, for grape berries detection, through the laser intensity.

This paper aims to establish a method for estimating vine production in field conditions using low-cost, non-invasive methods based on computer vision (Sethian, 1999). Thus, 3D models of bunches were obtained and dimensional analyses were carried out to calculate the different yield components. The development of the proposed method had to face additional difficulties, such as the partial 3D models generated by *in-situ* bunch documentation (only the visible side of the bunch), or having to deal with the occlusions and geometrical complexities of the strain itself. Techniques and algorithms to estimate the components of vineyard yields were developed. The recovery of the non-visible side in the model was achieved using convex hull techniques (Barber et al., 1996). The main output of this workflow was an accurate and precise predictive method aimed at eliminating the subjectivity deriving from the spatial and temporal variability of grape production. This would allow vine growers to take decisions in advance.

2. Image-based modelling

For precision agriculture purposes, a combination of Photogrammetry quality and computer vision algorithms (flexibility, automation and efficiency) is required. An image-based modelling technique based on this combination allows 3D models to be obtained from 2D images via two main steps: first, the automatic determination of the view of each image taken at vineyards; second, the automatic computation of the 3D coordinates for the generation of a dense and scaled 3D model of the scene.

This is a non-destructive and non-invasive technology with low associated costs. The methodology allows data to be acquired remotely with great efficiency, affording the radiometric and geometric characteristics of objects with a high degree of accuracy and detail in complex scenarios. Moreover, photogrammetry has advantages over others sensors, such as expensive laser systems (Lumme et al., 2008) or gaming sensor technologies, which are subject to difficulties in external daylight scenarios (Lange et al., 2011) as well as having a reduced working range.

To develop the proposed method, several robust descriptors for feature extraction and matching were implemented and tested, the SIFT variation (ASIFT) being the one that provided the best results in this study, where variations in geometry and lighting were very common. Last, but not least, several camera calibration models, such as the Brown model or the Fraser model, were integrated to allow working with any type of camera, including low-cost smartphones and tablets (Akca and Gruen, 2007).

Fig. 1 schematizes the workflow of the PW software (González-Aguilera et al., 2013) based on image-based modelling.

2.1. Image acquisition protocol

The image acquisition protocol is the key to success of the developed process since these images represent the input data of the workflow, and thus their position (spatial and attitude) will affect the final accuracy (i.e. in terms of perspective ray intersection) and completeness (i.e. in terms of overlap between images) of the 3d model.

Prior to image acquisition, the scene must be analyzed, including the lighting conditions because these will influence the exposure values, and the aperture and shutter speeds of the sensor. To this end, images should be acquired without strong variations in illumination, avoiding overexposed areas and ensuring sharpness, together with an occlusion analysis due to the presence of obstacles that will affect the image acquisition protocol and the overlaps between adjacent images. The shortest available focal length of the camera should be chosen and must be held constant throughout the image acquisition process to keep the internal camera parameters stable.

Regarding the geometric conditions of the camera shots, the objective is to establish an image acquisition protocol to reconstruct the grapevine of interest, guaranteeing the completeness and best accuracy of the resulting 3D model to perform the dimensional analysis of each cluster. It should be remarked that finding an optimal image network can be complex, in particular in scenes with strong depth variations and occlusions. Therefore, the key at this point is to establish a guideline, based on simple geometric constraints, to perform image acquisition at vineyards (Fig. 2). For an extensive vineyard study, at least five images (one master and four slaves) must be taken of each grapevine. The master image should cover the study grapevine and represents the origin of the coordinate system. This image has to be taken in a frontal way and framing the principal part of the grapevine or, if possible, including all of it. The overlaps between the slaves and master image must be high (80–90%) and must always maintain a slight convergence (i.e. optical axis) (maximum 15°) so that image matching will be ensured during the orientation phase. Regarding depth, this should be chosen according to the image scale or the desired resolution. This image acquisition protocol ensures the completeness and quality of the final model.

The automatic scaling of the 3D point clouds is resolved by incorporating self-scaled algorithms based on the automatic recognition of targets of known dimension which should appear in the photograph.

2.2. Features extraction and matching

One of the most critical steps in this process is the extraction and matching of features (lines and points) with high accuracy and reliability. This constitutes the framework that supports the whole process by providing the necessary information to resolve the spatial and attitude positions of images (orientation), camera self-calibration and, finally, grapevine 3D reconstruction. In addition, agricultural scenes usually show variations in scale, perspective and illumination, so classical descriptors based on grey levels such as area-based matching (ABM) (Joglekar and Gedam, 2012) and least square matching (LSQ) (Gruen, 1985) are useless. To this end, more sophisticated and robust descriptors were tested: smallest univalue segment assimilating nucleus (SUSAN)

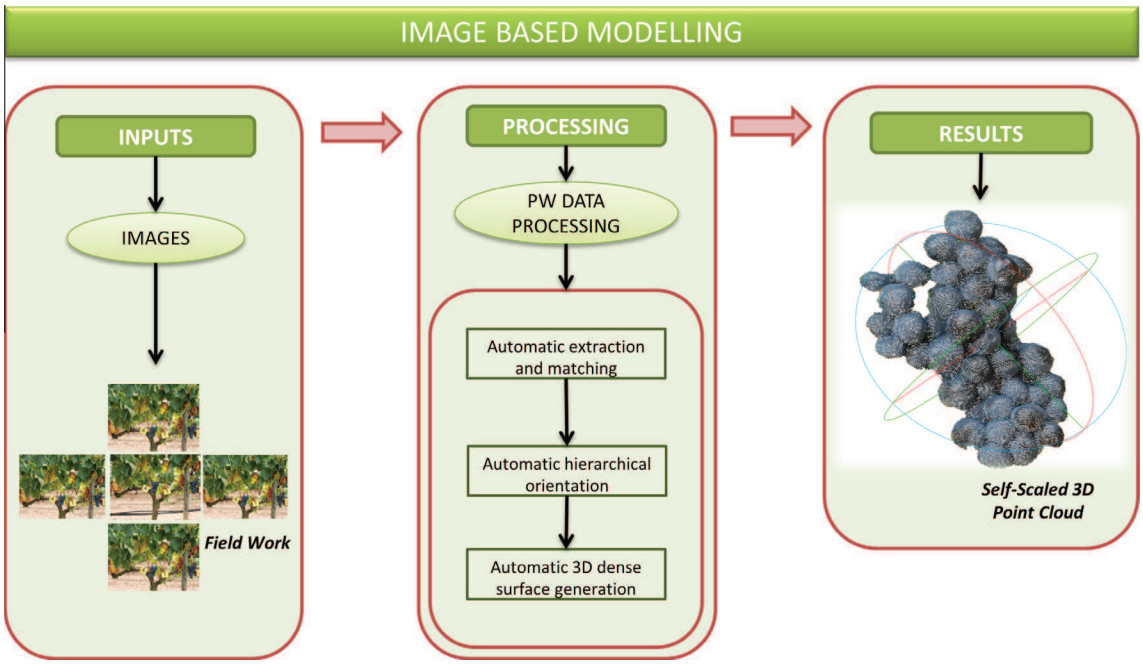

Fig. 1. Workflow of image-based modelling technique.

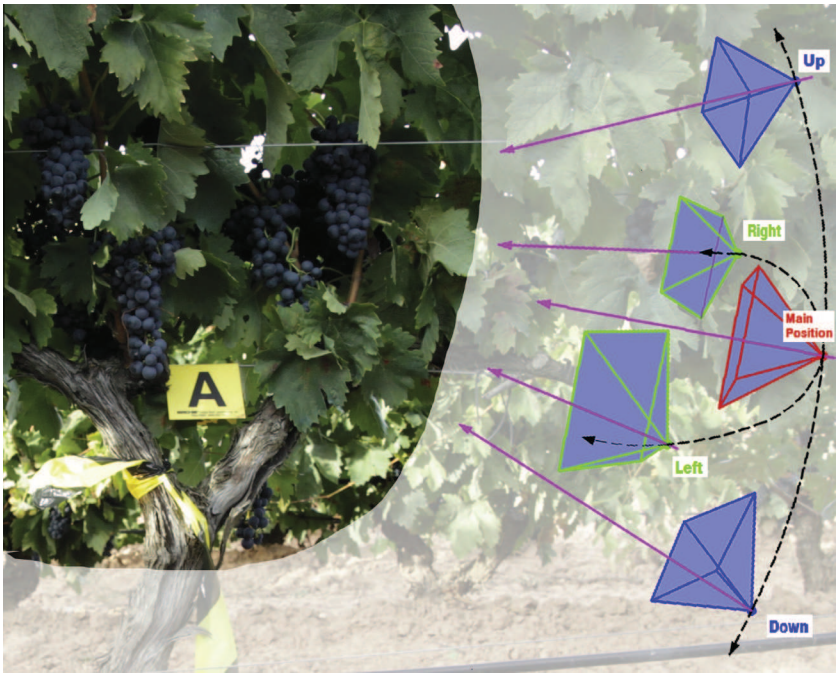

Fig. 2. Protocol of image acquisition of the study grapevines.

(Smith and Brady, 1997); scale invariant feature transform (SIFT) (Lowe, 2004); maximally stable extremal regions (MSER) (Matas et al., 2004) and speeded up robust features (SURF) (Bay et al., 2006). Unfortunately, all these algorithms become ineffective when there are considerable variations in perspective between images.

In this sense, a variation of the SIFT algorithm, called affine scale invariant feature transform (ASIFT) (Morel and Yu, 2009), has been incorporated to the PW software. It permits the consideration of two additional affinity parameters that control the perspective of the images. These are the angles corresponding to two perspective angles of the camera's optical axis, the ϖ angle (tilt) and the φ angle (axis) (Eq. (1)). In this way, the ASIFT algorithm allows one to cope with images that have a high degree of perspective, which is very common in vineyards. The result is an invariant descriptor that considers scale, rotation, movement and significant deformations due to the perspective between images. This result provides the next expression:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = H_{\lambda} R_1(\kappa) T_1 R_2(\varpi) = \lambda \begin{bmatrix} \cos\kappa & -\sin\kappa \\ \sin\kappa & \cos\kappa \end{bmatrix} \cdot \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\varpi & -\sin\varpi \\ \sin\varpi & \cos\varpi \end{bmatrix} \quad (1)$$

where A is the affinity transformation that contains the scale (λ), rotation (κ) around the optical axis (swing) and perspective parameters corresponding to the inclination of the camera optical axis (φ (tilt)), the vertical angle between the optical axis and the line perpendicular to the image plane and the horizontal angle between the optical axis and the fixed vertical plane (ϖ(axis)).

The matching process is carried out by SIFT descriptors. These descriptors are first matched according to the Euclidean distance (Lowe, 1999), after which they are filtered by the optimized random sampling algorithm (ORSA) (Moisan and Stival, 2004; Moisan et al., 2012). This algorithm is a variant of random sample consensus (RANSAC) (Fischler and Bolles, 1981) with an adaptative criterion to filter outliers by epipolar geometry constraints.

2.3. Hierarchical image orientation

Computing the spatial and attitude position of images is not an easy mathematical task. Furthermore, image orientation is based on features related to extraction and matching, so the presence of outliers could be quite common. In this sense, the method proposes a hierarchical approach for obtaining an approximate and relative image orientation in an arbitrary coordinate system based on computer vision. It can be refined and improved in an absolute and scaled coordinate system for the set of images thanks to photogrammetry.

Firstly, a relative image orientation is achieved using independent models obtained from the fundamental matrix calculated with the Longuet–Higgins algorithm (Longuet-Higgins, 1987). One of the greatest advantages of the fundamental matrix is that it is independent of any scene, so no knowledge of the internal parameters or initial approaches of the camera is required.

Secondly, once the relative attitude and spatial position of images has been obtained, a global adjustment of all images (bundle-adjustment) is performed by an iterative and least-squares process based on the collinearity condition (Kraus, 1993). The coordinates of the ground control points are incorporated for absolute georeferencing (Eq. (2)). These ground coordinates are added to the orientation process as artificial targets located around the vineyard. In cases in which the internal calibration parameters of the camera are unknown (principal distance, principal point and lens distortion), this second step allows these parameters (self-calibration) to be incorporated into the equation as unknowns (Quan, 2010).

$$\begin{aligned}(x-x_0)+\Delta x &= -f\frac{r_{11}(X-S_X)+r_{21}(Y-S_Y)+r_{31}(Z-S_Z)}{r_{13}(X-S_X)+r_{23}(Y-S_Y)+r_{33}(Z-S_Z)}\\ (y-y_0)+\Delta y &= -f\frac{r_{12}(X-S_X)+r_{22}(Y-S_Y)+r_{32}(Z-S_Z)}{r_{13}(X-S_X)+r_{23}(Y-S_Y)+r_{33}(Z-S_Z)}\end{aligned} \quad (2)$$

where x, y are the image coordinates of a point of the scene; X, Y, Z are the coordinates of the scene point expressed in a local reference system; r_{ij} are the rotation matrix coordinates; S_X, S_Y, S_Z are the spatial coordinates of the camera point of view; f is the camera focal length; x_0, y_0 are the principal point coordinates of the image, and X, Y are the radial and tangential distortion parameters of the lens.

2.4. Automatic dense surface generation

Starting from the robust image orientation, a dense matching process was developed. It is based on the semi-global matching technique (SGM) (Hirschmuller, 2005; Deseilligny and Clery, 2011). Applying the projective equation (Hartley and Zisserman, 2003) (3), it permits the generation of a dense model resulting from the determination of a 3D coordinate per pixel.

$$x_k = \boldsymbol{C}(\boldsymbol{D}(\boldsymbol{R}_i(X_k - S_i))) \quad (3)$$

where X is the 3D point; x is the point corresponding to the image; $\boldsymbol{R}$ is the camera rotation matrix; S is the camera projection center; $\boldsymbol{C}$ is the internal calibration function; $\boldsymbol{D}$ is the lens distortion function, and the subscripts k and i are related to point and image, respectively.

The SGM process consists of minimizing an energy function (4) through the eight basic directions that a pixel can take. This function is composed of a cost function, $\boldsymbol{M}$, which reflects the degree of similarity of the pixels between two images, x and x', together with the incorporation of two restrictions, and P_1 and P_2, which show the possible presence of outliers in the SGM process.

$$E(D) = \sum_x (M(x, D_x) + \sum_{x' \in N_x} P_1 T[|D_x - D_{x'}| = 1] + \sum_{x' \in N_x} P_2 T[|D_x - D_{x'}| > 1]) \quad (4)$$

where $E(D)$ is the energy function to be minimized on the basis of the disparity between the homologous characteristics; the function C evaluates the level of similarity between the pixel p and its counterpart q through the disparity Dp, while the P_1 and P_2 terms correspond to two restrictions to avoid outliers in the dense matching process due to the disparity of one pixel or a higher number of them, respectively.

In addition, a third constraint was added to the SGM process. This consisted of epipolar geometry derived from photogrammetry (Hartley and Zisserman, 2003) and can enclose the search space per pixel in order to reduce the huge computational cost. In particular, based on the computed fundamental matrix, $\boldsymbol{F}$, the following search band is defined: $\boldsymbol{l}' = x\boldsymbol{F}$, which establishes that the corresponding point of x should be along the epipolar line $\boldsymbol{l}'$. In other words if two image points (x and x') correspond, then the epipolar line $\boldsymbol{l}'$ is defined $\boldsymbol{l}' = \boldsymbol{F}\boldsymbol{x}$. As a result, it will generate a dense model with multiple images, obtaining more optimal processing times.

3. Estimation of vineyard yields

The measurement of vineyard yields is accomplished with a set of production components, such as volume, mass and the number of berries per bunch (Greven, 2007). This study attempts to establish these components, demonstrating the feasibility of scaling the proposed methodology to the total yield estimation of the vineyard ($\mathrm{kg} * \mathrm{m}^{-2}$). The current method improves classic and alternative methods (Kurtural and ÓDaniel, 2007) and the empirical relations proposed by Greven (2007). Different algorithms were implemented to analyze clusters from the metric point cloud dimensionally, obtaining the volume and weight, as well as determining the number of berries, following two computational strategies, as shown in Fig. 3. The first one is automatic and relies on the point cloud generated. The second one is a semi-automatic process and requires computer-aided design (CAD) models from the point cloud.

Both strategies are fed by external variables, the average volume of the berries (Vm) and the average density of the bunches (Dm) calculated by averaging the data acquired in laboratory tests. These variables depend on the area and year of the crop.

$$Vm = \frac{\sum_{i=1}^{i=n}(V/B)i}{n} \quad (5)$$

$$Dm = \frac{\sum_{i=1}^{i=n}(W/V)i}{n} \quad (6)$$

where Vm is the average berry volume calculated by means of the real bunch volume (V), the number of bunches (n) and the real number of berries per bunch (B); Dm is the average density of the bunch, and W is the real weight of the bunch.

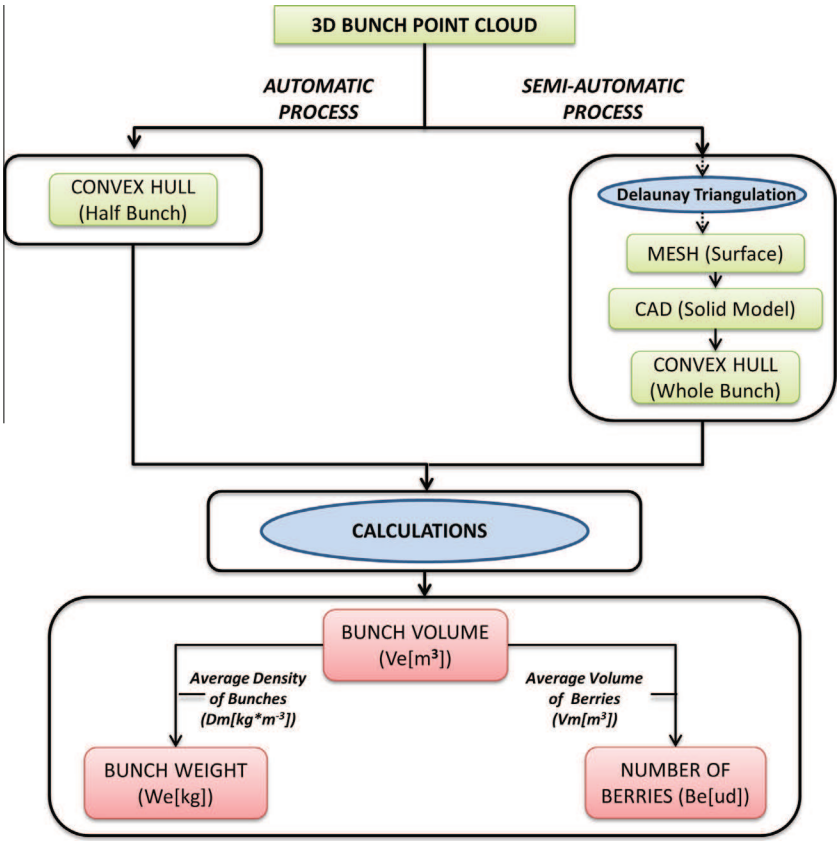

Fig. 3. Different computational strategies of vineyard production components from 3D bunch point cloud.

The two methods described are based on different modelling techniques of the 3D point cloud of the cluster. The first one is an automatic process supported by generating the convex hull (O'Rourke, 1994) of the 3D point cloud of the visible side of the cluster. This volume limited by the hull (*Vc*) collects half of the whole cluster due to the field of view of the camera and the limitations of the branches and leaves, closing the hidden side of the cluster by a flat surface. Because the convex hull includes empty spaces where there are no berries, as well as the exclusivity of the visible side of the bunch it is necessary to include an empirical correction factor (*K*) (Nuske et al., 2011; Nuske et al., 2012) that will refine the estimated volume (*Ve*).

The second strategy is based on a semi-automatic process supported by reverse engineering procedures that allows CAD models to be generated from 3D point clouds (DeLuca et al., 2006). It is noteworthy that there are still serious barriers in our knowledge of full automation in the conversion to CAD solid models, especially in the case of complex objects (Gonzalez-Aguilera et al., 2012). This approach requires the triangulation of the point cloud from the cluster to achieve the spatial topology required to model it. The meshing algorithm chosen was 3D Delaunay triangulation (Golias and Dutton, 1997) to obtain a TIN (Triangle Irregular Network). With this approach, the volume enclosed by the convex hull (*Vc*) of the whole cluster can be achieved, from modelling the individual berries of the visual side of the bunch and the subsequent symmetry. As in the automatic approach, an empirical correction factor (*K*) that refines the estimated volume (*Ve*) is required.

Fig. 4 shows both computational processes.

This second approach is more tedious. To obtain a CAD model close to reality, the mesh has to be repaired previously. This step uses the approximation of Attene (2010), which incorporates several automatic and sequential tasks:

- Filling of holes through algorithms of planar triangulations (Barequet and Sharir, 1995) or more complex approaches based on interpolators of radial basis function (Branch et al., 2006), the minimum distance (Dolenc and Makela, 1993) and the measurement of angles (Bøhn and Wozny, 1992; Varnuška et al., 2005).
- Repair of meshing gaps, based on minimum threshold distance algorithms (Rock and Wozny, 1992; Barequet and Kumar, 1997).
- Removal of topological noise, allowing the mesh to be re-triangulated locally (Guskov and Wood, 2001).
- Removal of geometric noise by algorithms that apply filters as anti-aliased Laplacians in general or specific zones (Fan et al., 2008).

After this reconstruction, the mesh has been modelled to obtain a solid CAD comprising two stages: segmentation and adjustment of basic primitives. As proposed by Wang et al. (2012), the segmentation process involves 3 steps:

- Plane segmentation using the region growing technique.
- Segmentation of quadric surfaces from knowledge of the vertex curvature obtained by fitting the local vicinity (Besl and Jain, 1988).
- Segmentation of free surfaces relying on the region growing technique to maximize the number of vertices connected topologically that can be generalized as a free surface of the B-spline type.

Once the geometry has been segmented, modelling continues with the setting of basic primitives. Thus, berries were approximated to spheres, extracted by an iterative adjustment of the minimization of the Euclidean distance from the mesh.

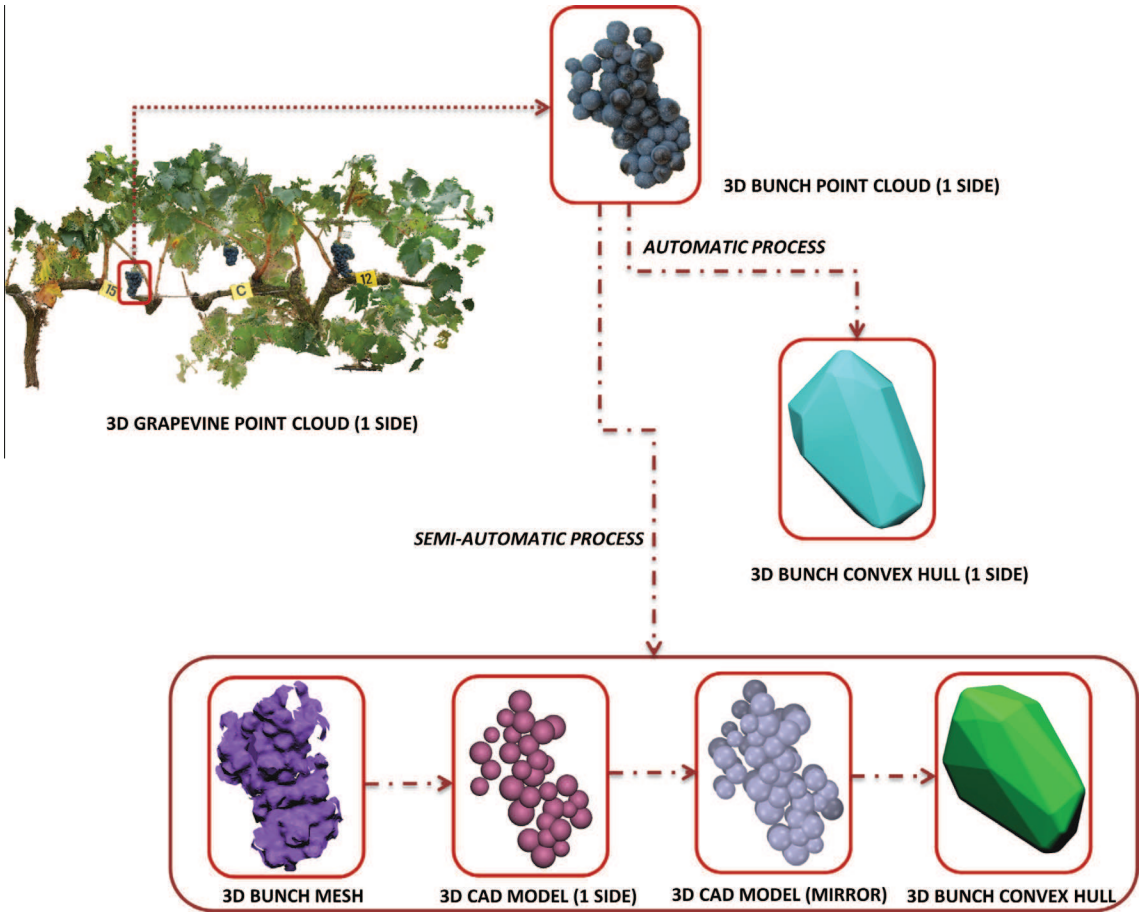

Fig. 4. Workflow of different computational processes to obtain the convex hull of the cluster.

The mathematical models implemented in both approaches are detailed below, where Eq. (7) shows the bunch volume estimated (*Ve*); bunch weight is estimated (*We*) in Eqs. (8) and (9) calculates the number of berries estimated per bunch (*Be*). *Vc* represents the volume limited by the convex hull and *K* is the empirical correction factor:

$$Ve = Vc * K \tag{7}$$

$$We = Vc * K * Dm \tag{8}$$

$$Be = Vc * K / Vm \tag{9}$$

Finally, a correlation between the calculated components and the groundtruth by a sampling of laboratory tests was established to analyze and validate the results acquired with both strategies.

4. Experimental results

Data collection was performed on the 20th of October 2013, on a day with bright and homogeneous lighting, at an experimental dry-farmed cv. Tempranillo (*Vitis vinifera* L.) vineyard of 1.05 ha, located 6 km from Logroño (lat. 42°26' N; long. 2°30' W; 455 m asl, La Rioja, Spain). Many studies have been conducted previously at this location (Romero et al., 2010; Vicente Renedo et al., 2007).

Tempranillo vines (clon RJ-26) were grafted onto Ritcher-110 rootstock and planted in 1995, following a between-row and within-row spacing of 2.90 m × 1.15 m respectively, with an East–West orientation. This corresponds to a density of 2998 vines/ha. The grapevines were spur-pruned on a bilateral cordon and trained to a VSP trellis system. The trellis featured a supporting wire at 0.70 m, two wires at 1.00 m above the ground for protection against wind damage, and a pair of movable shoot-positioned wires at 1.45 m.

Fig. 5. Image of trellis vineyards selected for this study.

Table 1
Technical specifications of the photographic sensor.

Canon EOS 500D	
Type	CMOS
Sensor size	22.3·14.9 mm
Total pixels	15.1 Mpíxels
Image size	4752·3168 pixels
Focal length	17 mm

Fig. 6. Example of matching points between images.

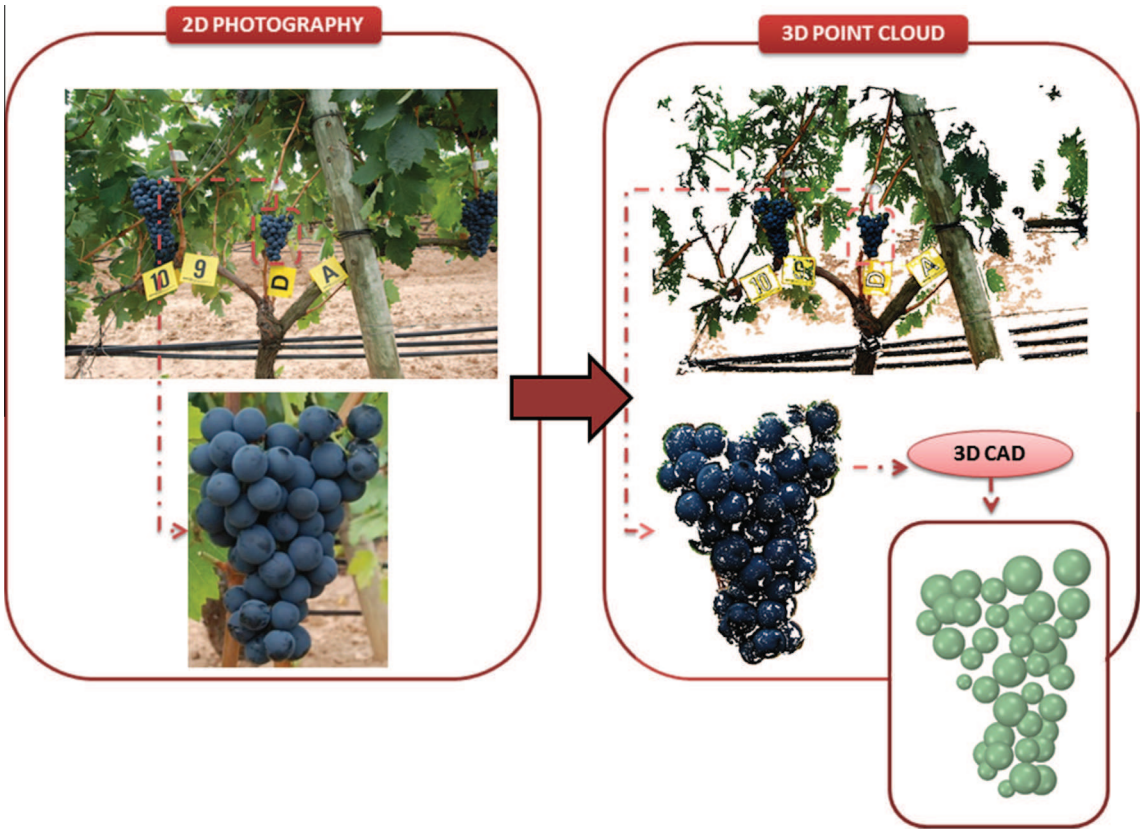

Fig. 7. 3D point cloud and CAD model of a cluster example from 2D images.

To validate the proposed method, 20 clusters from 14 trellis vines were chosen, taking photographs according to the protocol defined in Section 2.1. The selected strains were lightly defoliated to allow one side of the study clusters to be seen. Before taking the photographs, several artificial yellow targets were placed on the ground to allow the automatic scaling of point clouds (Fig. 5).

The main features of the camera used are shown in Table 1:

Images were taken using a fixed short focal length, with an of f/5.6 aperture, an exposure time faster than 1/160 s, ISO quality of 200 to decrease image noise, and without flash, owing to the shadow effects. The distance to the grapevine was approximately 2 m in order to maintain it inside the camera depth of field (DOF). The aperture was chosen to achieve an optimal compromise between exposure time, resolution and the DOF.

Photograph acquisition at such close distances requires the consideration of lens parameters, such as the Modulation Transfer Function (MTF), so the strains were centered in the image to achieve better resolution. Additionally, the MTF was taken into account for lens aperture selection (the lens "sweet" spot).

3D point cloud generation involves the identification and matching of homologous points between images, exemplified in Fig. 6 by straight white lines. The average number of homologous points between pairs of images was 580, obtained automatically through Micmac algorithm (Deseilligny and Clery, 2011). One of the keys to success in the matching process passes through acquiring pair of images with small baseline and thus guarantying high overlap between images. Moreover, the ORSA algorithm used in the matching process consists in introducing an *a contrario* (Desolneux et al., 2000) criterion to avoid the hard thresholds for inlier/outlier discrimination. Thus, the ORSA algorithm finds the right balance between the critical parameter σ and number of inliers by controlling the number of false alarms (Moisan et al., 2012).

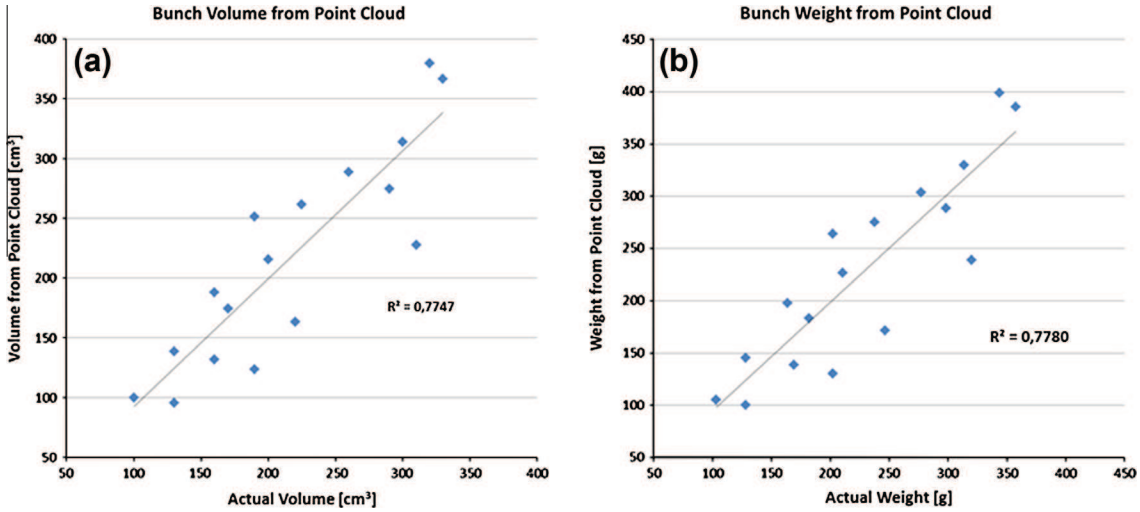

Fig. 8. Bunch volume (a) and weight (b) obtained automatically from point cloud.

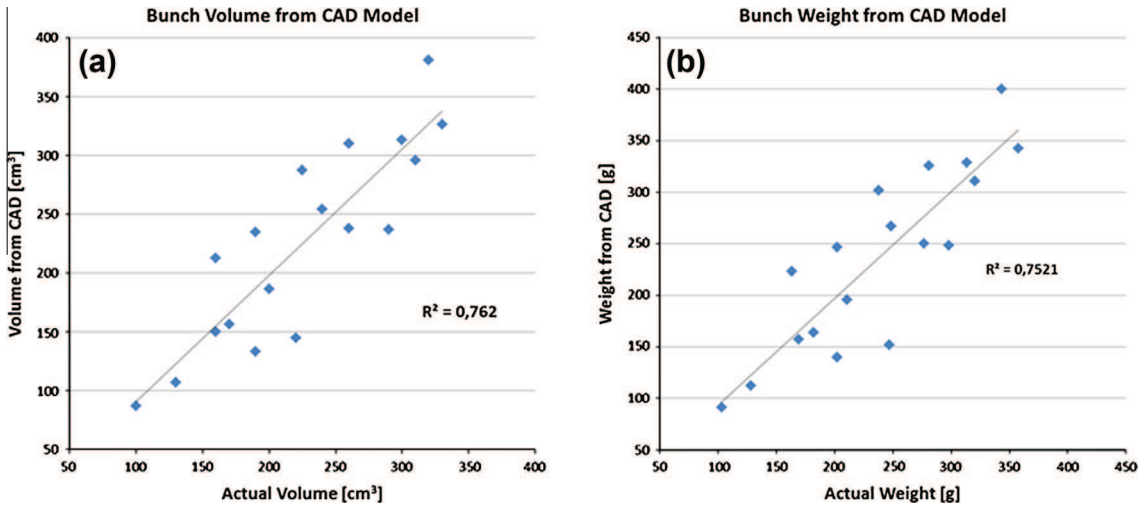

Fig. 9. Bunch volume (a) and weight (b) obtained semi-automatically from CAD model.

The dense point clouds generated have 4,276,000 points per vineyard, while the clusters alone have 96,460 points (as mean value). These points are utilizated to create the bunch mesh, having 90,779 points with a ground sample distance of 0.3 mm and a size of 8.3 kB (as mean value).

Fig. 7 shows scaled 3D models of a cluster example and CAD solids through 2D images, which were the inputs of the method proposed to estimate the vineyard yield.

The actual yield components were established by a laboratory test: volume calculation (immersion), weighing, and individual counting of berries per cluster. These laboratory tests which perform as a groundtruth were contrasted with estimated parameters, providing the error of each estimation. Owing to the nature of point clouds, the robust statistical parameter for defining the adjustment is MAD (median absolute deviation) (Mosteller and Tukey, 1977; Sachs, 1984) with respect to the median of the estimated values. Additionally, the Pope test (Pope, 1976) was applied for the detection and removal of the outliers. The final results are shown in Figs. 8–10.

- A coefficient of determination of 0.77 was achieved by comparing the actual volume of the cluster versus the volume estimated from the point cloud (Fig. 8a). The MAD had a value of 34 cm^3 when the median of the estimated volumes was 216 cm^3. In the case of bunch weight, an average coefficient of determination of 0.78 g was obtained (Fig. 8b), with a MAD of 35 g when the median was 227 g.
- On comparing the actual volume of the bunch with the estimated volume from CAD modelling, a coefficient of determination of 0.76 was obtained (Fig. 9a). The MAD was 33 cm^3, the median taking a value of 236 cm^3. When bunch weight was analyzed, a coefficient of determination of 0.75 was obtained (Fig. 9b), the MAD 35 g being when the median was 248 g.
- On testing the actual number of berries and the number of berries estimated by the point cloud, a determination value of 0.80 was obtained (Fig. 10a). The MAD of this fitting was 15 units, the median being 98 units. When this estimation was performed with CAD modelling, a coefficient of determination of 0.78 was obtained (Fig. 10b). In this case the MAD was 15 units when the median was 114 units.

From this analysis, it is possible to conclude that the semi-automatic process does not provide significant improvements, in contrast to the automatic one, which also has the advantage of a

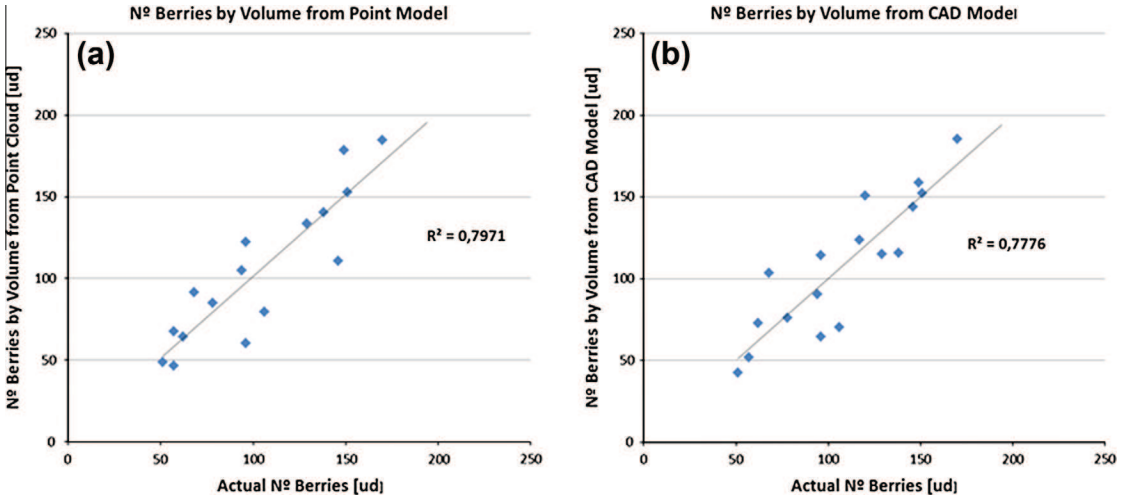

Fig. 10. Number of berries obtained automatically from point cloud (a) and number of berries obtained semi-automatically from CAD model (b).

reduced processing time and does not require human interaction. It significantly decreases subjective variability as well as featuring reduced associated costs and time saving. Furthermore, comparing the coefficients obtained with other yield estimation studies, this method can be cataloged as a great accurate one. Related to this assertion, Diago et al. (2012) achieved a R^2 of 0.73 between the observed and predicted yield values. Dunn and Martin (2004) succeeded a R^2 of 0.85 between grape weight and the ratio of grape pixels to total image pixel. Both cases used approaches based on supervised classification methodologies and an limited 2D analysis of bunch morphology. Additionally, both approaches required the employment of white screen to avoid the background influence which limits their effectiveness and automation in field.

5. Conclusions

This research presents a non-invasive low-cost method with application to precision viticulture that consists of vineyard yield estimation. The method is fast, reliable, robust and objective, and could serve to take decisions in advance regarding actions to be taken in vineyards, predicting the results and planning the vintage optimally. The method developed relies on the adaptation of the PW software tool, which allows the metric reconstruction and dimensional analysis of bunches from grapevine images taken in field conditions. The groundtruth established in laboratory tests ensures that the method is precise in complex agronomic scenarios and in trellis vine configurations, which are more complex to photograph than classic pruned configurations. The advantages of PW include process automation, sufficient quality to generate dense resolution models equivalent to the pixel size of the image, and low cost and ease of use.

Regarding the two processes developed to estimate vineyard production parameters (point cloud-automatic and CAD-semiautomatic), the results show that there is no significant improvement in CAD berry modelling. This is why the automatic methodology using integrated algorithms in PW software was selected for the present work. In addition, working only on one side of the cluster, the visible side, is feasible for approximating the whole bunch with an accurate and precise fit.

This methodology can be extrapolated to the whole of the vineyard, measuring its yield. Because this method is based on non-invasive technologies through a passive sensor, the constraints mainly depend on weather conditions, such as homogeneous lighting and the absence of rain and strong winds. By using artificial light and simple screens to diffuse light, several restrictions can be overcome. Since future prospects will address cluster compactness and its relationship to volumes, another action to be developed in the future will be to design and build a platform that supports the configuration of the image acquisition protocol described in Section 2.1 and that can be loaded on a mobile platform such as a quad. This would also allow the optimal distances between sensors to be determined and calibrated and hence the possibility of scaling models without targets, significantly optimizing the time devoted to field work and data processing.

Acknowledgments

Authors would like to thank the Instituto de Ciencias de la Vid y el Vino (ICVV) from the CSIC, the University of La Rioja and the Department of Agriculture of the Government of La Rioja, for allowing us to use their facilities and their collaboration during the experimental phase of this research. Special thanks to Javier Tardáguila, Mª Paz Diago, Borja Millán and Rubén Urraca.

References

Aguilar, M.A., Pozo, J.L., Aguilar, F.J., Sanchez-Hermosilla, J., Páez, F.C., Negreiros, J., 2008. 3D surface modelling of tomato plants using close-range photogrammetry. Int. Archives Photogr., Remote Sens. Spatial Informat. Sci., 330.

Akca, D., Gruen, A., 2007. Generalized Least Squares Multiple 3D Surface Matching. Institute of Geodesy and Photogrammetry, ETH Zurich.

Attene, M., 2010. A lightweight approach to repairing digitized polygon meshes. Visual Comput. 26 (11), 1393–1406.

Barber, C.B., Dobkin, D.P., Huhdanpaa, H., 1996. The quickhull algorithm for convex hulls. ACM Transact. Math. Software (TOMS) 22 (4), 469–483.

Barequet, G., Kumar, S., 1997. Repairing CAD models. In: Visualization'97, Proceedings, IEEE, pp. 363–370.

Barequet, G., Sharir, M., 1995. Filling gaps in the boundary of a polyhedron. Comput. Aided Geom. Des. 12 (2), 207–229.

Battany, M., 2008. A practical method for counting berries based on image analysis. In: 2nd Annual National Viticulture Research Conference, University of California, Davis, USA, 4–5, July 9–11.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: speeded up robust features. In: Computer Vision–ECCV 200. Springer, Berlin Heidelberg, pp. 404–417.

Besl, P.J., Jain, R.C., 1988. Segmentation through variable-order surface fitting. Pattern Anal. Machine Intell., IEEE Trans. on 10 (2), 167–192.

Blom, P.E., Tarara, J.M., 2009. Trellis tension monitoring improves yield estimation in vineyards. HortScience 44 (3), 678–685.

Bøhn, J.H., Wozny, M.J., 1992. Automatic CAD-model repair: shell-closure. In: Proc. Symp. on Solid Freeform Fabrication, pp. 86–94.

Branch, D., Dang, L.C., Hall, N., Ketchum, W., Melakayil, M., Parrent, J., Baron, E., 2006. Comparative direct analysis of type Ia supernova spectra. II. Maximum light. Publ. Astron. Soc. Pac. 118 (842), 560–571.

Chherawala, Y., Lepage, R., Doyon, G., 2006. In food grading/sorting based on color appearance trough machine vision: the case of fresh cranberries. Informat. Commun. Technol., 1540–1545, ICTTA '06. 2nd, 0–0 0.
Clingeleffer, P.R., Martin, S., Krstic, M., Dunn, G.M., 2001. Crop development, crop estimation and crop control to secure quality and production of major wine grape varieties: a national approach: final report to grape and wine research development corporation. Grape and Wine Research & Development Corporation.
DeLuca, J.G., Gall, W.E., Ciferri, C., Cimini, D., Musacchio, A., Salmon, E., 2006. Kinetochore microtubule dynamics and attachment stability are regulated by Hec1. Cell 127 (5), 969–982.
Deseilligny, M.P., Clery, I., 2011. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images. In: Proceedings of the ISPRS Symposium, 3DARCH11, pp. 269–277.
Desolneux, A., Moisan, L., Morel, J.M., 2000. Meaningful alignments. Int. J. Comput. Vision 40 (1), 7–23.
Diago, M.-P., Correa, C., Millán, B., Barreiro, P., Valero, C., Tardaguila, J., 2012. Grapevine yield and leaf area estimation using supervised classification methodology on RGB images taken under field conditions. Sensors 12, 16988–17006.
Djuricic, A., Weinmann, M., Jutzi, B., 2014. Potentials of small, lightweight and low cost multi-echo laser scanners for detecting grape berries. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XL-5, 2014. In: ISPRS Technical Commission V Symposium, 23–25 June 2014, Riva del Garda, Italy.
Dolenc, A., Makela, I., 1993. Some efficient procedures for correcting triangulated models. In: Proceedings of Solid Freeform Fabrication Symposium, pp. 126–134.
Dunn, G.M., Martin, S.R., 2003. The current status of crop forecasting in the Australian wine industry. ASVO Seminar Series: Grapegrowing at the Edge, 24.
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 (3), 196–198.
Fan, Y.Z., Tam, B.S., Zhou, J., 2008. Maximizing spectral radius of unoriented Laplacian matrix over bicyclic graphs of a given order. Linear and Multilinear Algebra 56 (4), 381–397.
Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.
Golias, N.A., Dutton, R.W., 1997. Delaunay triangulation and 3D adaptive mesh generation. Finite elements Anal. Des. 25 (3), 331–341.
Gonzalez-Aguilera, D., Del Pozo, S., Lopez, G., Rodriguez-Gonzalvez, P., 2012. From point cloud to CAD models: Laser and optics geotechnology for the design of electrical substations. Opt. Laser Technol. 44 (5), 1384–1392.
González-Aguilera, D., Guerrero, D., Hernández-López, D., Rodríguez-Gonzálvez, P., Pierrot, M., Fernández-Hernández, J., 2013. PW, Photogrammetry Workbench. <http://www.isprs.org/catcon/catcon6.aspx> (accessed 30.04.14).
Greven, M., 2007. Manual for yield forecasting in Sauvignon Blanc grapes. Marlborough Wine Research Centre.
Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. South African J. Photogr., Remote Sens. Cartogr. 14 (3), 175–187.
Guskov, I., Wood, Z.J., 2001. Topological noise removal. In: 2001 Graphics Interface Proceedings. Ottawa, Canada, 19.
Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.
Herrero Langreo, A., Barreiro Elorza, P., Santamaria, D., Paz, M., Baluja, J., Ochagavia, H., 2010. Pixel classification through mahalanobis distance for identification of grapevine canopy elements on rgb images.
Hirschmuller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition. CVPR 2005. In: IEEE Computer Society Conference on, vol. 2, pp. 807–814.
Joglekar, J., Gedam, S.S., 2012. Area based image matching methods—a survey. Int. J. Emerg. Technol. Adv. Eng. 2, 130–136.
Kraus, K., 1993. Photogr. Fundamentals Standard Processes, vol. 1. DummlersVerlag, Bonn, Germany, ISBN 3-427-78684-6.
Kurtural, K., ÓDaniel, B., 2007. Crop estimation in vineyards. UK Cooperat. Extens. Service Univ. Kentucky-College Agric., 45.
Lange, B., Chang, C., Suma, E., Newman, B., Rizzo, A.S., Bolas, M., 2011. Development and evaluation of low cost game-based balance rehabilitation tool using the microsoftkinect sensor. Engineering in Medicine and Biology Society, EMBC, In: 2011 Annual International Conference of the IEEE, pp. 1831–1834.
Longuet-Higgins, H.C., 1987. A computer algorithm for reconstructing a scene from two projections. In: Fischler, M.A., Firschein, O. (Eds.), Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, pp. 61–62.
Lowe, D., 1999. Object recognition from local scale-invariant features. In: International Conference on Computer Vision, Corfu, Greece.
Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60 (2), 91–110.
Lumme, J., Karjalainen, M., Kaartinen, H., Kukko, A., Hyyppä, J., Hyyppä, H., et al., 2008. Terrestrial laser scanning of agricultural crops. Int. Archives Photogr., Remote Sens. Spatial Informat. Sci. 37, 563–566.
Martin, S., Dunstone, R., Dunn, G., 2003. How to Forecast Wine Grape Deliveries Using Grape Forecaster Excel Workbook Version 7. GWRDC, Adelaide, Australia, p. 100.
Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. Image Vis. Comput. 22 (10), 761–767.
Moisan, L., Stival, B., 2004. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. Int. J. Comput. Vision 57 (3), 201–218.
Moisan, L., Moulon, P., Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. Image Processing On Line, 10.
Morel, J.M., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. SIAM J. Imag. Sci. 2 (2), 438–469.
Mosteller, F., Tukey, J., 1977. Data Analysis and Regression. Addison-Wesley, Upper Saddle River, NJ.
Nuske, S., Achar, S., Bates, T., Narasimhan, S., Singh, S., 2011. Yield estimation in vineyards by visual grape detection. Intelligent Robots and Systems (IROS). In: 2011 IEEE/RSJ International Conference on, pp. 2352–2358.
Nuske, S., Gupta, K., Narasimhan, S., Singh, S., 2012. Modeling and Calibrating Visual Yield Estimates in Vineyards. In: Yoshida, K., Tadokoro, S. (Eds.), Field and Service Robotics, vol. 92. Springer, Berlin Heidelberg, pp. 343–356.
O'Rourke, J., 1994. Computational Geometry. C. Press Syndicate of the University of Cambridge, p. 118.
Pope, A.J., 1976. The statistics of residuals and the detection of outliers.
Quan, L., 2010. Image-Based Modeling. Springer, New York.
Roby, G., Harbertson, J.F., Adams, D.A., Matthews, M.A., 2004. Berry size and vine water deficits as factors in winegrape composition: anthocyanins and tannins. Aust. J. Grape Wine Res. 10 (2), 100–107.
Rock, S.J., Wozny, M.J., 1992. Generating topological information from a bucket of facets. In: Solid Freeform Fabrication Symposium Proceedings. The University of Texas at Austin, Austin, pp. 251–259.
Romero, I., García-Escudero, E., Martín, I., 2010. Effects of leaf position on blade and petiole mineral nutrient concentration of Tempranillo grapevine (*Vitis vinifera* L.). Am. J. Enol. Viticulture 61 (4), 544–550.
Sachs, L., 1984. Applied Statistics: A Handbook of Techniques. Springer-Verlag, New York, p. 253.
Serrano, L., González-Flor, C., Gorchs, G., 2012. Assessment of grape yield and composition using the reflectance based water index in Mediterranean rained vineyards. Remote Sens. Environ. 118, 249–258.
Sethian, J.A., 1999. Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science, vol. 3. Cambridge University Press.
Smith, S.M., Brady, J.M., 1997. SUSAN—a new approach to low level image processing. Int. J. Comput. Vision 23 (1), 45–78.
Varnuška, M., Parus, J., Kolingerová, I., 2005. Simple holes triangulation in surface reconstruction. In: Proceedings of ALGORITMY, pp. 280–289.
Vicente Renedo, T., Martínez García, J., Chavarri, J.B., Marín, M., González Marcano, S., Domínguez Miguel, N., Villar Capellán, R., 2007. Preservation of genetic heritage of authorized vinifera grapes and rootstocks in Rioja DOC [Designation of Origin]. In: Foro Mundial del Vino, 5, Logroño (España), 28–30 Mar 2006. GLR, CAGDR.
Wang, J., Gu, D., Yu, Z., Tan, C., Zhou, L., 2012. A framework for 3D model reconstruction in reverse engineering. Comput. Ind. Eng. 63 (4), 1189–1200.
Whalley, J., Shanmuganathan, S., 2013. In Applications of image processing in viticulture: A review. In: 20th International Congress on Modelling and Simulation, Adelaide, Australia, pp. 531–537.
Wolpert, J., Vilas, E., 1992. Estimating vineyard yields: introduction to a simple, two-step method. Am. J. Enol. Viticul. 43 (4), 384–388.