Computers and Electronics in Agriculture 144 (2018) 26–36

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Automated early yield prediction in vineyards from on-the-go image acquisition

Arturo Aquino, Borja Millan, Maria-Paz Diago, Javier Tardaguila*

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

A R T I C L E I N F O

Keywords:
Grapevine early yield prediction
Image analysis
Non-invasive sensing
Precision viticulture

A B S T R A C T

Early grapevine yield assessment provides information to viticulturists to help taking management decisions to achieve the desired grape quality and yield amount. In previous works, image analysis has been explored to this effect, but with systems performing either manually, on a single variety or close to harvest-time, when there are few rectifiable agronomic aspects. This study presents a solution based on image analysis for the non-invasive and in-field yield prediction in vines of several varieties, at phenological stages previous to veraison, around 100 days from harvest. To this end, an all-terrain vehicle (ATV) was modified with equipment to autonomously capture images of 30 vine segments of five different varieties at night-time. The images were analysed with a new image analysis algorithm based on mathematical morphology and pixel classification, which yielded overall average Recall and Precision values of 0.8764 and 0.9582, respectively. Finally, a model was calibrated to produce yield predictions from the number of detected berries in images with a Root-Mean-Square-Error per vine of 0.16 kg. This accuracy makes the proposed methodology ideal for early yield prediction as a very helpful tool for the grape and wine industry.

1. Introduction

Among all collectable data from a vineyard, grapevine yield estimation outstands for its economical relevance (Wolpert and Vilas, 1992; Martin et al., 2002; Dunn, 2010), and also for being key to help optimizing plant growth and to improve fruit quality (Dunn and Martin, 2003). Yield variability within a vineyard has been proved to be high (Bramley and Hamilton, 2004). Classical yield estimation methods, which consist on manual collection and weighting of the crop yield in a given and limited number of plants previous to harvest is tedious and insufficient to obtain representative yield data. Consequently, non-invasive imaging-based methods are being investigated to make possible the efficient and continuous capture of detailed information from vines throughout their life cycle (Spalding and Miller, 2013; Li et al., 2014).

Grapevine yield is determined by the yield components, defined as the number of clusters, the number of berries per cluster and the berry weight (Tardaguila et al., 2012). Alternatively, yield can also be estimated from the total number of berries and the berry weight (Nuske et al., 2014). Whatever the case, the number of berries is the most labile variable determining yield (Anderson et al., 2008). It is highly influenced by the weather conditions during inflorescence development and berry-set, when it gets fully established and remains mostly invariable until harvest (May, 2004).

Imaging-based developments aimed at yield estimation can be found in the literature under two differentiated approaches. A first set of proposals is framed within the manual acquisition of images, thereby focusing on the discrete analysis of samples from the vineyard. Contrary, the second set studies yield estimation by means of on-the-go image acquisition using modified human-driven, or autonomous vehicles, with the ambition of making possible the continuous and massive analysis. Within the first approach, Chamelat et al. (2006) presented a method based on pixel classification to detect grape-pixel aggregations in vine RGB images manually taken in the field under daylight. Reis et al. (2011) also proposed the detection of grape-pixel aggregations in outdoor RGB images that were taken using the camera flash. They employed colour analysis and non-linear image processing to detect aggregations of white-grape pixels. The method was later improved and extended to work with red grapes also (Reis et al., 2012). On the other hand, Rahman and Hellicar (2014) studied the detection and classification of undeveloped and mature clusters of white grapes. Their development consisted on the extraction of the clusters from the background as a first step, to finally perform pixel classification by means of texture analysis. Other approaches involved colour histogram classification, RGB thresholding and fuzzy clustering to estimate cluster weight in vine images (Liu et al., 2013). A more complex capturing device was developed by Fernández et al. (2013), which consisted on a

* Corresponding author.
E-mail address: javier.tardaguila@unirioja.es (J. Tardaguila).

https://doi.org/10.1016/j.compag.2017.11.026
Received 24 March 2017; Received in revised form 14 November 2017; Accepted 17 November 2017

sensor rig composed of a CCD camera and a servo-controlled filter wheel. This device was used to collect multispectral imagery, which was analysed employing a sequential masking algorithm based on k-means clustering. Aquino et al. (2017) proposed a methodology based on mathematical morphology and false positives filtering by means of supervised learning to count the number of berries in RGB cluster images taken in the vineyard with a smartphone under natural sunlight. Very recently, Mack et al. (2017) described a method based on 3D reconstruction of high-resolution laser range data to model sampled grape clusters.

Regarding the methods for on-the-go vineyard monitoring, the literature is rather limited. Some works involved image acquisition after veraison (Font et al., 2015; Rose et al., 2016), very close to harvest, when the time window for cluster thinning, to either regulate the yield or to promote increased allocation of carbohydrates to specific clusters (to improve fruit ripening) is very limited. Font et al. (2015) equipped a ground vehicle with a reflex camera and artificial lighting to capture vine images of a single table grape variety (Flame seedless) at night-time right before harvest and reported yield prediction errors of 16%. More recently, a multi-view-stereo system coupled to a LED external illumination system was mounted in an autonomous moving platform (0.3 km/h), to acquire night-time images, and the estimation of the number of berries in the cluster images was attempted (Rose et al., 2016). The choice of earlier yield estimation, before veraison, was taken by several authors. Likewise, Nuske et al. (2014) used a tractor, which moved around 5 km/h, to mount RGB cameras and artificial illumination for image capture in the vineyard at night time. They collected data from six grapevine varieties in several growing seasons and vineyards, at various timings before harvest, and developed a supervised classifier for berry detection fed with a wide set of descriptors (above 30) of texture, colour and shape. Average yield estimation accuracy of 11% was only provided for Flame seedless, using images acquired at harvest time (Nuske et al., 2014). A different approach was followed by Liu et al. (2017), who developed a computer vision system for early grape yield estimation based on shoot detection using videos acquired with a low-cost camera at daytime using a white background which moved alongside the row.

In view of the missing factors and lessons learned from previous works, the goal of the present study was the development of a comprehensive technological solution for automated early grapevine yield prediction based on images acquired on-the-go at a speed similar or faster than most agricultural vehicles. In contrast to other available tools focused mostly on yield estimation of a single variety at pre-harvest, this proposal analyses vine images of tens of clusters of five different grapevine varieties at earlier stages; concretely at phenological stages between berry-set and cluster closure (around 100 and 120 days before harvest) This earlier stage at which yield information is acquired could significantly improve its impact, since grape-growers could perform viticultural practices to rectify certain key parameters more effectively. Additionally, a secondary goal of the present work was to simplify the classifier for berry detection and counting and to evaluate the potential development of a unique solution involving multiple grapevine varieties.

2. Materials and methods

2.1. Experimental design

The trials were carried out during season 2015 in a commercial vineyard located in Falces (lat. 42°27′46.0″N; long. 1°48′12.9″W; Navarra, Spain). Five grapevine (*Vitis vinifera* L.) varieties (red and white) were considered for this study: Albariño, Cabernet Sauvignon, Syrah, Tempranillo and Viognier. The vines were planted in year 2009 on rootstock Richter 110 in N-S orientation, trained to vertical shoot positioning (VSP) trellis system with 2 and 1 m inter-row and intra-row distances, respectively. Vines were defoliated from node one to six only on their east side after berry set. Depending on the variety, at the time of image acquisition in the field (23rd June 2015), the clusters were at phenological stages between K and L, according to the scale proposed by Baggiolini (1952). Following this scale, phenological stage K refers to that at which berries have the 50% of their final size; it is also denoted as pea-size stage. With regard to phenological stage L, also called cluster closure, it is reached when berries have about 70% of their final size, and they start to touch each other within the cluster.

2.2. Acquisition of reference data

For every cultivar, ten consecutive sampling segments composed of three adjacent vines each, were labelled and delimited ($5\ varieties \times 10\ segments \times 3\ vines$); thirteen additional segments of non-defoliated vines were also selected. Each three-vine segment constituted a unique sampling point in which the produced yield was individually weighted and registered at the end of the season using a hanging scale (Kern CH15K20, Kern & Sohn GmbH, Balingen, Germany).

2.3. Image acquisition

The vine segments were photographed ‘on-the-go’ at night time without user intervention. To this effect, a sensor-equipped all-terrain vehicle (ATV) was driven through the vineyard at 7 km/h. Image capture automation was achieved by adapting the vehicle to incorporate the following elements:

- A mirror-less RGB camera (Sony α7II, Sony, Tokyo, Japan) equipped with a Zeiss 24/70 mm lens with optical stabilisation. This camera mounts a 24 Mpx CCD sensor, incorporates a 5-axis image stabilisation system and provides high shutter speed, quick image storage and low noise generation (Fig. 1(a)). These features allowed to capture and store on-the-go three images per wheel-spin at 7 km/h, producing high quality images despite the vibrations caused by the ATV’s engine and the irregular ground’s surface. For the experiments, the camera was set in manual mode, configuring the aperture in f/4, shutter speed in 1/2500 s, ISO sensitivity in 5000 and focus in manual mode.
- A white-light LED panel (Fig. 1(a)) to provide controlled artificial illumination for vineyard monitoring at night time. By means of illumination and camera parametrization, it was possible to isolate in the image the vines under evaluation from those in the adjacent row. Should image acquisition be conducted at day time, the sunlight would equally illuminate the whole scene, producing images in which the vines under study were hardly distinguishable from those at their back (this fact is illustrated in Fig. 2).
- A modular and flexible structure built with commercial aluminium profiles for sensor attachment. The structure consisted on an upfront and a rear tray, plus an adjustable arm to be installed in the front tray (Fig. 1(a)). The arm was designed to mount the camera and illumination system, making the combination adjustable to different vineyard heights and widths. The arm was adjusted for the camera to be at around 1.5 m from the canopy.
- An inductive sensor installed in the rear axle for camera triggering; the sensor produced three activation pulses per wheel-spin (Fig. 1(b)).
- A GPS receiver (Leica Zeno 10 Global Positioning System; Heerbrug, St. Gallen, Switzerland) for image georeferencing (Fig. 1(c)).
- A custom-built electronic control system for managing the signals generated by the installed devices (GPS and inductive sensor) and triggering the camera using an isolated signal (Fig. 1(d)). The system also allowed for data storage in an SD-card and for showing capture-status information in a 4.9″ TFT screen.