A Lightweight Method for Grape Berry Counting based on Automated 3D Bunch Reconstruction from a Single Image

Scarlett Liu[1], Mark Whitty[2] and Steve Cossell[3]

Abstract— **Berry counting is an integral step towards grape vine yield estimation. As a traditional yield estimation step, counting berry by human hand is tedious and time consuming. Recent methods have approached this using specialized stereo cameras and lighting rigs which are impractical for a large scale field application. This paper presents a lightweight method for generating a representative 3D reconstruction of an individual grape bunch from a single image from one side of the bunch. The results were poor prior to the application of a sparsity factor to compensate for bunches of varying sparsity, with the final result being an absolute average accuracy of 87.6% and average error of 4.6%, with an R^2 value of 0.85. These results show promise for *in vivo* counting of berry numbers in a non-computationally expensive manner.**

Keywords: Grape, Berry, Viticulture, Image Processing, 3D Bunch Reconstruction

I. INTRODUCTION

Yield estimation in viticulture is notorious for producing poor estimates due to range of sampling factors and dependency on subjective interpretation of the state of vine maturity. This poor estimation costs hundreds of millions of dollars each year in contract adjustments, harvest logistic management, oak barrel purchases and tank space allocation amongst others. The structure of vineyards means aerial imagery is only able to contribute a small amount to the yield estimation, and other on ground estimation methods are time consuming. Recent work by Nuske [1] in the US has shown the potential for image processing to speed up this analysis as well as generate unbiased estimates which are orders of magnitude smaller than manual estimates, leading to substantial cost savings.

As to traditional yield estimation in vineyards, berry number is a critical parameter for early forecasting production since the number of berries remains stable after fruit setting [2]. Also the ratio between of berry number per bunch and bunch size is one of many factors governing the quality of the fruit at harvest. At current vineyards, counting berry is accomplished by hand, which is work intensive and time consuming. [3], [4], [5] demonstrated the advantages of image processing on yield components analysis for the sake of saving time and energy for grape production forecast. [6], [4] applied image processing techniques for berry counting one side of a bunch, achieving average R^2 value of 0.92 and 0.82 between actual berries and detected berries per bunch. However, the image processing algorithm proposed in paper [6] can not be utilised after véraison since the reflection on berry skin is affected by pruine (which causes matte surface on berries on both green and purple grapes). As the work presented by Diago [4], a dataset with 70 bunches from 7 varieties was tested, with a R^2 value varying from 0.62 to 0.95 based on 10 bunches for each variety (0.817 for 7 caltivars in average). Leaving the image techniques described by the author alone, 10 bunches is not representative for validating image processing procedure in one cultivar. Especially for Cabernet Sauvignon as well as Shiraz which are famous for non-uniform bunch shape, [4] obtained the lowest R^2 value with 0.62 based a single image of Cabernet Sauvignon from 7 cultivars.

Except detecting berries from one side by processing one image, other work [5], [7] showed the advantages of performing 3D reconstruction of grape bunches for the purpose of estimating the number of grapes in a bunch by stereo images. Their accuracy improved achieved an R^2 value of 0.78 opposed to more traditional 2D estimation techniques [3] which have been a staple for the image processing community [8], [9]. Their 3D reconstruction relies on substantial manual input (semi-automatic) for each bunch, which is tedious even given an impressive user interface and thus cannot be applied on a large scale for reliable yield estimation. As to the scope of experiment, data sets in paper [5], [7] are small, 10 bunches from one cultivar (10 cultivars) and 20 bunches from 14 vines in one block, respectively. Also R^2 achieved in both paper are 0.71 and 0.78, which is not satisfied for practical implementation in current vineyards. In addition, a specialized stereo camera arrangement was required, along with controlled lighting conditions, limiting the applicability to *ex vivo* analysis. Stereo cameras also have a minimum range which restricts the level of detail which may be achieved by moving closer, meaning in field application within the confines of a sprawling canopy is impractical.

In order to increase of these image processing methods, low cost and simpler solutions are needed that can be applied by farmers on the ground. Thus objective of this paper is to form a representative 3D reconstruction of grape bunches from a single image for the purpose of accurate berry counting. The use of a single image only is a key feature, which simplifies the data capture process and keeps the cost manageable, to the point where cameras such as those

[1]Scarlett Liu is with School of Mechanical and Manufacturing, University of New South Wales, 2052 Sydney, Australia `sisi.liu@unsw.edu.au`

[2]Mark Whitty is with School of Mechanical and Manufacturing, University of New South Wales, 2052 Sydney, Australia `m.whitty@unsw.edu.au`

[3]Steve Cossell is with School of Mechanical and Manufacturing, University of New South Wales, 2052 Sydney, Australia `scos506@gmail.com`

contained in current smart phones can be used. The lack of equipment overhead is particularly attractive to farmers in difficult economic environments.

In the remainder of this paper, **Section II** details the proposed method before **Section III** describes the experimental data and methods used to validate the method. Results follow in **Section IV** before conclusions are drawn and recommendations are made for future work in **Section V**.

II. Fundamentals

Two major image processing components form the basis of the berry counting method. First, a 3D reconstruction of the bunch is formed to give an initial estimate of the number of berries. A sparsity factor is then calculated from the color of the berries and used to generate a final estimate of the number of berries. This paper is based on three assumptions:

1) The actual number of berries in a bunch is equal to the number of berries that fit in a volumetric shell derived from a single image of a bunch.
2) All sizes of invisible berries follow the normal distribution of sizes of visible berries same bunch / sub-bunch.
3) Sparse factor has an effect on estimating the number of berries per bunch.

A. 3D reconstruction and initial estimate of berry number

Given an image of a single bunch of grapes, the outline of the bunch is extracted from R channel by Otsu's method [10]. The image is rotated until its major axis is approximately vertical. Each point on the outline is considered a candidate berry location to which a circle is fitted using a Hough transform, as demonstrated in **Fig.1**. These circle locations and diameters are used to seed the 3D model by placing spheres of corresponding diameter in a single plane normal to the direction of view. For addressing overlapping of berries at the edge of a bunch, neighbours searching within specific distance is applied for finding two berries that have extreme metrics within this distance. Then the berry with largest metric is moved forward in z-direction (normal to the paper plane) pixel by pixel while the berry with smallest metric is moved backward until there is no overlapping.

Beginning at the top detected berry, the 3D model is populated using the following process until the bottom of the bunch is reached:

1) Find the first and last pixel of a horizontal section through the image and subtract the diameter of one berry.
2) Revolve this section about a vertical axis through its centre, forming a virtual circle.
3) Randomly pick a sphere diameter within the observed range of berry diameters.
4) Moving around the circumference of the virtual circle, attempt to place a new sphere at regular (1 degree) intervals.
5) At each interval, place a sphere at that location on the circumference only if no intersection with any existing sphere is detected.

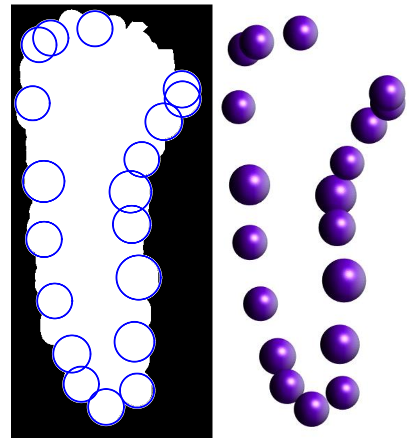

Fig. 1: Berries as seeds on the edge of a bunch

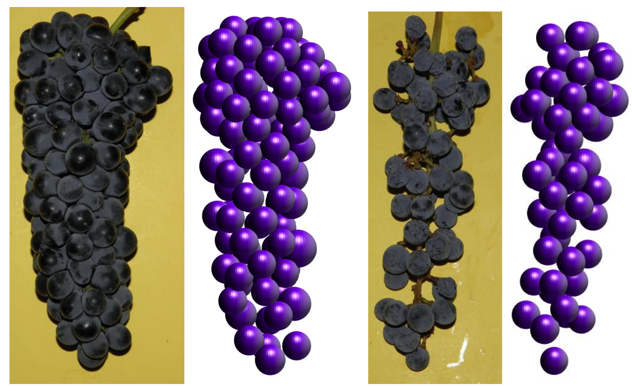

Fig. 2: 3D reconstruction by a single image

6) Move down a defined step size (in this paper, step size is two pixels) and repeat. Once the model is fully populated, the number of berries is tallied and denoted as Initial Berry Number (IBN).

As to **step 3)** above, Hough Transform is applied on a image of a bunch to detect all visible berries and a normal distribution model is built based on all radius of detected berries. Then in aforementioned **step 3)**, a radius is randomly generated by the built normal distribution. **Assumption 1)** and **2)** are embedded here. **Fig.2** illustrated examples of single images and the corresponding shaded 3D reconstructions.

B. Sparsity factor

Assumption 1) refers a convex hull for a healthy and compact grape bunch. However, there are not always compact bunches so that IBN is not accurate enough for a bunch with loosen pattern by applying aforementioned image processing techniques. Hence in this paper, a sparse factor (SF) is proposed for defining the compactness of a bunch. This indicator will be used for final estimation of berry number. **Assumption 3)** is embedded here. In order to get SF for each bunch, each image was processed according to the following sequence of operations:

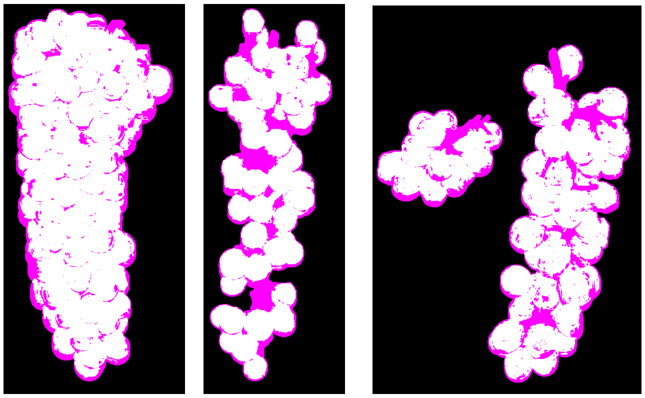

Fig. 3: Sparsity Factor Calculation

1) Automatically crop image to the outline of the bunch as detect above.
2) Automatically threshold the Red channel to obtain a binary image using Otsu's method.
3) Automatically threshold the Saturation (from HSV) channel to obtain a binary image using Otsu's method.
4) Calculate the area of each of these two thresholded images, giving AR and AS, as shown in **Fig.3**
5) Calculate the sparsity factor according to: $SF = (AR - AS)/AR$

C. Final estimate of berry number

The sparsity factor is then used to improve the estimation of the number of berries through the following formula:

$$BN = SF \times IBN \tag{1}$$

where BN is the final estimate of the number of berries.

III. Experiments

Data was collected by viticulturists at Treasury Wine Estates, Camatta Hills, California in September and October 2013. Photographs were taken of a total of 112 individual bunches randomly comprised of two red grape varieties Shiraz and Cabernet Sauvignon.

Images were captured at a resolution of 3968 x 2976 pixels using a consumer grade compact camera (Olympus SP600UZ) on automatic mode with the flash turned on. These images were then processed using Matlab according to the method in **Section II**. Firstly, a 3D reconstruction of each bunch was generated from a single image of that bunch, providing an initial estimate of the number of berries. Secondly, the sparsity factor for each bunch was calculated and applied to the initial estimate to obtain a final estimate of the number of berries.

Each bunch was then de-constructed, with manual counts of the number of berries on each bunch being recorded. In addition, the diameters of a small number of berries on each bunch were measured. The number of berries was compared with the final estimate from the proposed method, and the following metrics calculated:

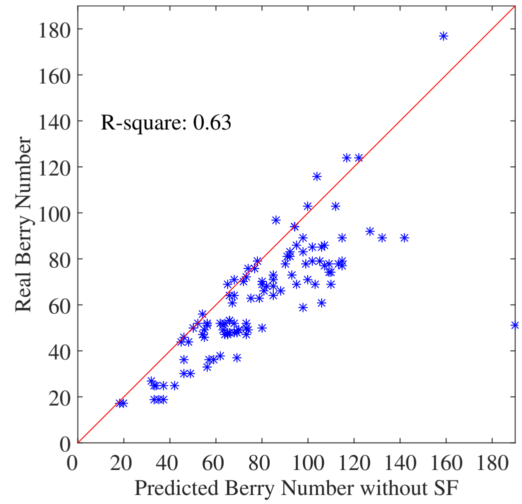

Fig. 4: Initial berry number estimation, with average absolute error 23.1%

Average absolute error: Taking the absolute values of the differences between the actual and estimated number of berries divided by the actual number of berries and then averaging these differences over all bunches.

Accuracy: 1 minus average absolute error.

Average error: Taking the values of the differences between the actual and estimated number of berries divided by the actual number of berries and then averaging these differences over all bunches.

R^2 **value**: Based on a linear correlation between the actual and estimated number of berries.

IV. Results

Fig.4 shows the relationship between the actual and initially estimated berry number. **Fig.5** shows the relationship between the actual and finally estimated berry number. The sparsity factor ranged from 0.32 to 0.89.

On the dataset of 112 images described above, an average absolute value error of 12.4% (accuracy is 87.6%) was achieved and average error is 4.6%. The proposed method generated an R^2 value of 0.63 using the initial estimate of berry number. Following application of the sparsity factor, a final R^2 value of 0.85 was achieved.

The processing time was approximately 0.5 seconds per image, prior to any optimization.

This method fits berries to the outer profile of the bunch, which matches in field observations as to the structure of real grape bunches and produces aesthetically pleasing models. It is notable that larger bunches induced larger errors in the method, most likely due to a larger number of interior berries.

A comparison about proposed method and other three berry counting methods is demonstrated in **table I**. In terms of processing type, the proposed method in [4] requires calibrating the relationship between visible and invisible berries in a testing data set, while approaches presented in paper [5], [7] need human interaction with software. So these three methods are not totally automatic while the proposed

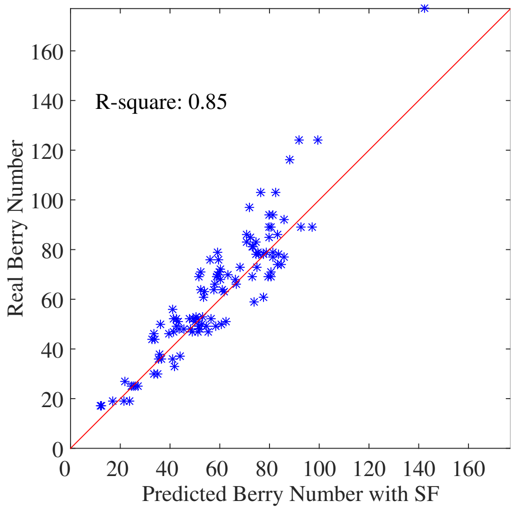

Fig. 5: Final berry number estimation by multiplying sparse factor, with average absolute error 12.4%

approach can estimate berry number based on one image of a bunch straight away.

TABLE I: Comparison of performance with other's works

Method	Cultivar	Data Set Size	Type	R^2
Diago [4]	Cabernet Sauvignon	10 bunches	Semi-auto	0.62
Ivorra [5]	10 cultivars	100 bunches	Semi-auto	0.71
Herrero [7]	Tempranillo	20 bunches	Semi-auto	0.78
Proposed Method	Cabernet Sauvignon and Shiraz	112 bunches	Automatic	0.85

V. CONCLUSIONS

This paper has presented a lightweight method for estimating the 3D structure of grape bunches from a single image. Experiments on two varieties of red grapes showed an average absolute accuracy of 87.3% relative to the actual number of berries on a bunch. The method achieved an R^2 value of 0.85 using a linear relationship between the estimated and actual number of berries. These results were obtained with nothing more than a standard compact camera.

The proposed 3d model based on one image also works on a bunch with distinguishing shoulder, as shown in the second row of **Fig 6**. But it cannot achieved a good estimation of berry numbers on a bunch with overlapping shoulders. Also this work is limited to purple bunch since the sparse factor is achieved by color operation. Future work will focus on extending this work to green grapes and more bunch shapes, and fitting visible berries into the exact position in its 3D reconstruction model. In addition, comparison of the results with analysis of the same bunches as photographed *in vivo* is expected to demonstrate the viability of the method for reliably counting the number of berries and in turn estimating block yield.

The processing time may also be improved by using a larger distance between horizontal sections as per step (f) in

Fig. 6: 3D model of a bunch with distinguishing shoulder

Section II-A. Some varieties of grapes elongate noticeably following véraison, and this method could be extended to fitting ellipses and reconstruction using corresponding ellipsoids. Furthermore, the 3D structure may be used for large scale analysis of the bunch structure, as it allows rapid estimation of many bunch parameters which are tedious to calculate via existing manual methods.

ACKNOWLEDGMENT

The authors would like to thank Will Drayton, Franci Dewyer, Joseph Geller and Angus Davidson from Treasury Wine Estates in collecting the raw images used in this paper. One author is partly supported by the Chinese Scholarship Council.

REFERENCES

[1] S. Nuske, K. Gupta, S. Narasimhan, and S. Singh, "Modeling and Calibrating Visual Yield Estimates in Vineyards," *Field and Service Robotics*, pp. 343–356, 2012.

[2] S. Martin, R. Dunstone, and G. Dunn, "How to forecast wine grape deliveries using grape forecaster excel workbook version 7," Tech. Rep., 2003.

[3] S. Liu, S. Marden, and M. Whitty, "Towards Automated Yield Estimation in Viticulture," in *Proceedings of Australasion Conference on Robotics and Automation*. Sydney, Australia: araa.asn.au, 2013, pp. 2–4.

[4] M. P. Diago, J. Tardaguila, N. Aleixos, B. Millan, J. M. Prats-Montalban, S. Cubero, and J. Blasco, "Assessment Of Cluster Yield Components By Image Analysis," *Journal of the Science of Food and Agriculture*, July 2014.

[5] E. Ivorra, A. Sánchez, J. Camarasa, M. Diago, and J. Tardaguila, "Assessment of grape cluster yield components based on 3d descriptors using stereo vision," *Food Control*, vol. 50, pp. 273–282, 2015.

[6] M. Grossetete, Y. Berthoumieu, J.-P. Da Costa, C. Germain, O. Lavialle, G. Grenier, *et al.*, "Early estimation of vineyard yield: site specific counting of berries by using a smartphone." in *Infomation Technology, Automation and Precision Farming. International Conference of Agricultural Engineering-CIGR-AgEng 2012: Agriculture and Engineering for a Healthier Life, Valencia, Spain, 8-12 July 2012*. CIGR-EurAgEng, 2012, pp. C–1915.

[7] M. Herrero-Huerta, D. González-Aguilera, P. Rodriguez-Gonzalvez, and D. Hernández-López, "Vineyard yield estimation by automatic 3d bunch modelling in field conditions," *Computers and Electronics in Agriculture*, vol. 110, pp. 17–26, 2015.

[8] R. Chamelat, E. Rosso, A. Choksuriwong, C. Rosenberger, H. Laurent, and P. Bro, "Grape Detection By Image Processing," in *IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics*. IEEE, Nov. 2006, pp. 3697–3702.

[9] M. J. C. S. Reis, R. Morais, E. Peres, C. Pereira, O. Contente, S. Soares, A. Valente, J. Baptista, P. J. S. G. Ferreira, and J. Bulas Cruz, "Automatic detection of bunches of grapes in natural environment from color images," *Journal of Applied Logic*, vol. 10, no. 4, pp. 285–290, 2012.

[10] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.