Journal of Applied Logic 13 (2015) 643–653

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

Automatic grape bunch detection in vineyards with an SVM classifier

Scarlett Liu *, Mark Whitty

School of Mechanical and Manufacturing, University of New South Wales, Sydney 2052, Australia

ARTICLE INFO

Article history:
Received 11 December 2014
Received in revised form 11 May 2015
Accepted 29 May 2015
Available online 9 June 2015

Keywords:
Bunch detection
Image processing
Low cost computation
Precision viticulture
Yield estimation
Field application

ABSTRACT

Precise yield estimation in vineyards using image processing techniques has only been demonstrated conceptually on a small scale. Expanding this scale requires significant computational power where, by necessity, only small parts of the images of vines contain useful features. This paper introduces an image processing algorithm combining colour and texture information and the use of a support vector machine, to accelerate fruit detection by isolating and counting bunches in images. Experiments carried out on two varieties of red grapes (Shiraz and Cabernet Sauvignon) demonstrate an accuracy of 88.0% and recall of 91.6%. This method is also shown to remove the restriction on the field of view and background which plagued existing methods and is a first step towards precise and reliable yield estimation on a large scale.

1. Introduction

Precise yield prediction and forecasting in vineyards are predicted to help the Wine Industry save 100 million dollars per year since it provides the base for wine marketing and management. The supply chain from grape growing to wine marketing relies heavily on yield estimation at times well before harvest. Poor yield estimation or forecasting could damage relations between grape growers and wineries, considering both sides need to endure the cost of inaccurate harvest planning and intake purchasing. As a result, accurate yield estimation and forecasting are in high demand in the viticulture industry both for practical and academic reasons. Nowadays, yield estimation is still undertaken by human hands worldwide, and its precision depends on the scale of hand sampling. Large scale sampling is not a guarantee of successful yield forecasting since usually some vineyards are not uniform [4] and many variables may affect the accuracy as harvest approaches, such as temperature and water availability. At the same time, this hand sampling work is tedious, expensive, inaccurate and has human bias (humans tend to choose healthier and bigger bunches when they are doing sampling in the field) [13]. So an automatic yield estimation system based on

* Corresponding author.
E-mail addresses: sisi.liu@unsw.edu.au (S. Liu), m.whitty@unsw.edu.au (M. Whitty).

http://dx.doi.org/10.1016/j.jal.2015.06.001
1570-8683/

(a) An image from Block 11 (b) An image from Block 19 (c) An image without bunches from Block 19

Fig. 1. Three original images with different uninformative regions of rocks, grass, posts and trunks, captured by consumer-grade camera in vineyard during daytime. The three images were taken at different times from different blocks at the same distance from the cordon.

image processing in vineyards has become a mainstream research topic in recent years. For tackling this problem, Nuske et al. [14] developed a prediction method by detecting the berry number from images taken in a vineyard. This paper applied a Radial Symmetry Transform (RST) [12] to find keypoints, which are potential berry locations, from every single image as the initial step for later feature analysis. This novel work was the first one to process a large collection of ground truth vine images to generate a yield estimate for viticulture. However, before the keypoints are extracted, there is no image size reduction. The problem is that the computation of high level image processing methods becomes slower with increased image size. Take RST for instance, the theoretical order of computation is O(KN), where K is the total pixel number of the image and N is the size of the neighbourhood [12]. That is, the computational cost doubles if 50% of an image was not useful for yield estimation.

In reality, for accurate yield estimation in vineyards, images are usually captured by a camera which is mounted on a viticultural vehicle. With the aim of not missing bunches, the field of view is set large enough to photograph the grape bunches in every single shot. Other interrelated objects, such as rocks, posts, grass, and other grapevines behind the current row are visible in the images. Every single image taken along a row in each location in a vineyard needs to cover all bunches in vertical direction, which means that majority of the vine canopy needs to be captured in each image, as shown in Fig. 1. Since the berry size is small compared with the canopy, and extracting the details of berries, such as berry size, is vital to late yield calibration [11], large field of view images need to be processed. High resolution images are slow to process when some high level image processing methods, such as RST and Zernike Moments Extraction [9], are applied. For the required field of view, images captured in vineyards contain around 70% meaningless information for yield estimation, as Fig. 1 illustrates. This means 70% of the processing time is redundant. Therefore, this paper aims to reduce the image size or extract relevant windows before extracting details of berries in the images for accelerating the image processing tasks. Extracting relevant windows for further processing also allows an estimate of the number of bunches to be made.

Feature extraction and grapevine structure classification has been researched in recent years. In 2011, Correa et al. performed a comparison [2] of different Fuzzy C-Means (FCM) clustering algorithms to extract features from vineyard images. The 20 segmented images achieved an accuracy of 85%, 87% and 88% by Robust Fuzzy Possibilistic C-Means, FCM and FCM-GK (FCM with Gustafson–Kessel) methods respectively. In the next year, another two extended papers [2,7] by the same research team combined Support Vector Machines (SVM), K-Means and the Scale-Invariant Feature Transform (SIFT) to cluster different objects in vineyard images. In the same year, a classification of grapevine structures from uncalibrated image sequences was presented by Dey et al. [5]. In this paper, Structure-From-Motion (SFM) was implemented to build a 3D reconstruction of grapevines as a first step, then a saliency feature was applied for traversability analysis of point cloud data. An SVM and a conditional random field (CRF) were adopted for spatially smoothing the 3D model in the final step, improving the accuracy of classification. Mahalanobis measures were applied in papers [6,18] for grape, branch, leaf and background classification. The accuracy of classification in the