Computers and Electronics in Agriculture 114 (2015) 154–162

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Apple detection in nighttime tree images using the geometry of light patches around highlights

Raphael Linker *, Eliyahu Kelman

Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 18 November 2014
Received in revised form 10 March 2015
Accepted 5 April 2015

Keywords:
Artificial vision
Fruit localization
Specular highlights
Yield estimation

ABSTRACT

Detection of fruit in tree images has been the focus of numerous studies. Although most studies considered approaches based primarily on color analysis, the major drawback of such approaches is that the fruit apparent color depends not only on variety or physiological stage but also on illumination, which is inherently non-uniform within the canopy, even if artificial lighting is used. In the present work we developed a novel approach to detect apples in nighttime images by analyzing the spatial distribution of the light around highlights ("bright spots"). The approach is based on the observation that, under the artificial illumination used, apples exhibit strong specular reflection so that a small, but very bright, spot is visible on almost all apples. Each of these highlights serves as the center of a region of interest and is the seed of the investigated light patch. This patch is initially very small but its size is increased iteratively by annexing pixels with predefined decreasing gray level intensities. The evolution of the patch geometry is used to determine whether it corresponds to an apple. The approach was tested with two datasets containing over 360 images (close to 13,000 apples) acquired in the same 'Golden Delicious' orchard in July 2012 and August 2013. Twenty images from the 2012 dataset were randomly selected to develop and calibrate the procedure. The results of these 20 images were used to establish a linear relationship between the number of detected objects and the actual number of apples visible in the images ($R^2 \sim 0.75$). Applying the calibrated procedure to the remaining images of this dataset led to an estimate of 6739 apples compared to a visual count of 6195 apples (~9% overestimate). Analysis of the 2013 dataset, in which the apparent size of the apples was smaller, required only adjustment of the two parameters related to apple size. Following this adjustment, 12 images were randomly selected to determine the relationship between the number of detected objects and the actual number of apples ($R^2 \sim 0.74$). Using this relationship, the estimated number of apples was 6687, compared to the visual count of 6713 fruits.

1. Introduction

Localization of fruit on trees remains a challenging issue, which has potential applications ranging from early estimation of fruit load which could be used to manage the orchard (thinning operations, irrigation, etc.) to localization of mature fruit for robotic harvesting or yield estimation. A first review devoted to fruit detection was published in 2000 by Jimenez et al. (2000a). More recently, Kapach et al. (2012) surveyed the advances and challenges still faced by computer vision for robotic harvesting. Despite more than 30 years of research, the unstructured, uncontrolled, cluttered outdoor environment which is typical of agricultural operations still presents many challenges for computer vision systems. Due to the complexity of the task various approaches have been investigated, ranging from a simple RGB camera to systems which provide three dimensional information (such as laser scanners, multi-views systems, structured light, time-of-flight, e.g. Jimenez et al., 1999, 2000b; Rakun et al., 2011; Barnea and Ben-Shahar, 2014), hyperspectral imaging (e.g. Annamalai and Lee, 2003; Safren et al., 2007) or thermal imaging (e.g. Stajnko et al., 2004; Bulanon et al., 2008; Wachs et al., 2010). The advantages and limitations of the various approaches are discussed in Kapach et al. (2012). The use of a standard RGB camera has obvious advantages in terms of cost and ease of operation, and such a simple configuration is still the focus of numerous studies (e.g. Kurtulmus et al., 2011; Linker et al., 2012; Payne et al., 2013; Zhou et al., 2012; Bansal et al., 2013; Kelman and Linker, 2014). Most studies which attempted to detect green fruit within green foliage with a single RGB camera under natural illumination concluded that the

* Corresponding author. Tel.: +972 4 8295902; fax: +972 4 8228898.
E-mail address: linkerr@tx.technion.ac.il (R. Linker).

http://dx.doi.org/10.1016/j.compag.2015.04.005
0168-1699/

uncontrolled illumination made it very difficult to achieve robust and reliable results. In order to overcome this, several studies have investigated the use of nighttime imaging under artificial illumination (e.g. Sites and Delwiche, 1988; Payne et al., 2014; Cohen et al., 2014). These studies considered the more homogenous illumination of the scene to be the main advantage of this technique. In the present study we investigated the usefulness of another feature of nighttime imaging, namely the fact that in such images convex surfaces exhibit a "bright spot" due to specular reflection (so-called specular highlights). Specular highlights may also exist in daytime images, and Mairon and Ben-Shahar (2014) recently reported the use of this feature for detecting sweet peppers in greenhouses. When a strong source of artificial light roughly aligned with the camera is used, as in nighttime imaging, these highlights are more pronounced. Although such highlights are usually considered a nuisance and numerous studies have been devoted to removing them (e.g. Tan et al., 2006; Blanc-Talon et al., 2009), Font et al. (2014) recently showed that these highlights could be used to estimate the number of grapes in nighttime images. A similar approach is followed in the present study. However, whereas Font et al. (2014) were able to use simple filtering techniques to mask the background (and hence most of the highlights not due to grapes), such background removal is not possible for typical tree images. In order to distinguish between highlights on apples and highlights on leaves, branches or other reflecting surface, we developed a procedure that relies on the spatial variation of the light intensity around the highlight.

The present work is part of a project whose ultimate goal is the development of a vision system for estimating orchard yield. As such, and contrary to vision systems for harvesting robots, the emphasis is not on exact localization of the fruits but rather on accurate estimation of the number of fruits in the images. Also, since each image covers a large portion of the tree, the apparent size of the fruit is much smaller than in images acquired for robotic harvesting.

2. Materials and methods

2.1. Datasets

The study involved two datasets of images captured in 'Golden Delicious' orchards at the Matityahu Research Station located in Northern Israel. The acquisition of both datasets was started one hour after sunset on cloudless nights. In order to create datasets that included trees with both high and low fruit loads, images were acquired in two areas known to produce very different yields. In each area images were acquired in two rows, labeled H_1 and H_2, and L_1 and L_2, respectively. The planting pattern in both areas was 2.0 m by 4.5 m. The first dataset included 210 images of 35 trees captured on July 9, 2012 with a Canon PowerShot SD1200IS camera (Resolution $3648 * 2736$ pixels) placed approximately 3.0 m from the tree trunks. The images were stored in 24-bit RGB format using the "Fine JPEG" setting of the camera. A professional studio flash (VILEN RG1700, equipped with a reflector umbrella) synchronized with the camera was used for artificial illumination. The flash was placed manually in front of each tree and three images were acquired holding the camera at heights of roughly 100 cm, 170 cm and 220 cm in order to obtain overlapping images which covered the whole side of the tree. The second dataset, which included 156 images of 26 trees, was collected on August 6, 2013. Three identical cameras (Canon PowerShot G7. Resolution $3648 * 2736$ pixels) were mounted on a vertical pole at 110 cm, 180 cm and 240 cm from the ground, and the pole was mounted on a trailer which was pulled by an electric cart. Two additional poles were mounted on the trailer at 60 cm from the camera pole, and LED lights were mounted on these poles at heights of 130 cm and 250 cm. The cart was driven in the middle of the row and the distance between the cameras and the tree trunks was approximately 260 cm. The cameras were operated in fully automatic mode but without flash. The images were stored in 24-bit RGB format using the "Fine JPEG" setting of the camera. The images were not taken "on-the-fly" but the cart stopped in front of each tree to prevent image blurring.

Prior to the analysis all the images were converted to gray-level images according to

$$I = 0.2989 * R + 0.5870 * G + 0.1140 * B$$

where I denotes the intensity in the gray-level image and R, G and B are the red, green and blue values in the original image.

2.2. Description of the fruit localization procedure

2.2.1. Outline and terminology

As mentioned in Section 1, when the light source is roughly aligned with the camera, each convex surface exhibits one (and only one) specular highlight. Our preliminary visual inspection of nighttime images of apple trees acquired with artificial illumination showed that such a specular highlight was clearly visible on most apples. Although such highlights occurred also on leaves or other surfaces, it was observed that when the highlight was located on an apple, neighboring pixels with intensity above some threshold formed a circle, or part of a circle. This can be explained by the fact that apple diffuse reflectance is primarily isotropic. This observation led to development of the method detailed below: After identifying a local maximum of the light intensity, a sub-image centered at this point was extracted and normalized, and regions of connected pixels corresponding to decreasing light intensities were formed and analyzed. In the sequel, these regions are called light *patches*. Detection of apples and rejection of other objects was achieved through analysis of the geometrical properties of these light *patches*, and in particular with the help of their inscribed polygons.

Throughout the description of the algorithm, the sub-image shown in Fig. 1 is used as illustrative example. This image is

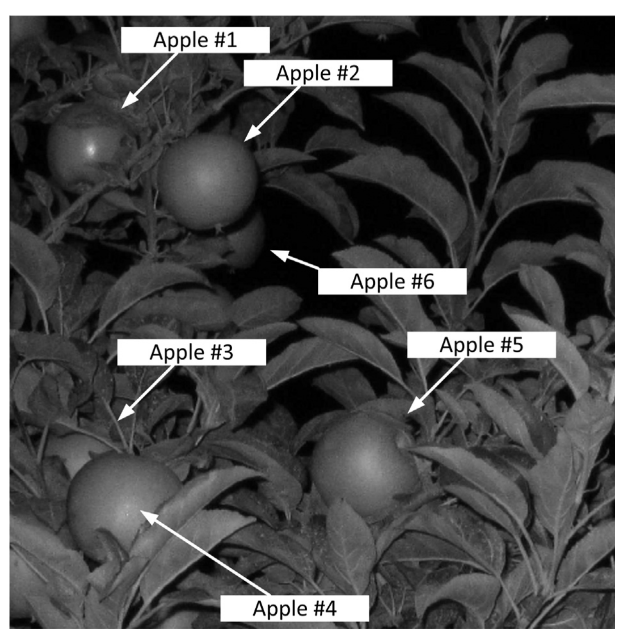

Fig. 1. A $1000 * 1000$ pixels region of a typical nighttime image.