Computers and Electronics in Agriculture 106 (2014) 75–90

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

# Recognition of clustered tomatoes based on binocular stereo vision

Xiang Rong [a,b], Jiang Huanyu [a,c], Ying Yibin [a,c,*]

[a] College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China
[b] College of Quality and Safety Engineering, China Jiliang University, 258 Xueyuan Road, Hangzhou 310018, China
[c] Key Laboratory of Equipment and Informatization in Environment Controlled Agriculture, Ministry of Agriculture, China

ARTICLE INFO

*Article history:*
Received 13 September 2013
Received in revised form 12 May 2014
Accepted 18 May 2014

*Keywords:*
Overlapping
Binocular stereo vision
Recognition
Depth map
Circle regression

ABSTRACT

To improve the applicability of the recognition method for clustered tomatoes, an algorithm based on binocular stereo vision was presented. First, a depth map of clustered tomatoes was acquired using a combination stereo matching method. Second, the noises in the depth map were removed using an eight-neighbor mode denoising method. Third, the clustered regions were classified into two types (i.e., overlapping and adhering regions) based on the depth difference between the front and back regions in a clustered region using an iterative Otsu method. Finally, different recognition methods were used for different types of clustered tomatoes. For adhering tomatoes, a recognition method based on edge curvature analysis was used for the edges in color image. For overlapping tomatoes, the same method was applied for the edges in color image, which were segmented into several parts by the edges in depth map after segmentation. A total of 189 pairs of stereo images were tested, and the recognition accuracy rate of clustered tomatoes was 87.9% when the leaf or branch occlusion rate was less than 25%. The acquisition distance and average execution time of this method were 300–500 mm and approximately 0.5 s, respectively. In conclusion, this method can realize the recognition of the clustered types and different types of clustered tomatoes, despite the serious occlusion of other tomatoes. Moreover, the headmost tomato in clustered tomatoes can be recognized based on depth information. This method can also realize the recognition of clustered tomatoes based on the images taken at different distances. However, the success rate of clustered tomatoes was not satisfactory when the occlusion was serious. Further research should focus on the improvement of the accuracy of stereo matching and the recognition of tomatoes occluded by leaves.



## 1. Introduction

To date, problems of labor shortage and high labor cost are increasingly becoming more serious along with the aging population in China. Harvesting is a time-consuming and laborsome procedure during fruit and vegetable production; thus, the automation of this procedure is urgently needed. As a very ideal solution, studies on harvesting robots have been of particular interest to researchers (Plebe and Grasso, 2001; Van Henten et al., 2002; Tanigaki et al., 2008; Rath and Kawollek, 2009; Shigehiko et al., 2010; Zhao et al., 2011). However, most of these studies remain at the experimental stage because of their insufficient ability to adapt to unstructured environment. The vision system, which is the main component of a harvesting robot, is used to acquire information on the unstructured environment. The functions of the vision system of a harvesting robot include fruit and vegetable maturity judgment (Wang et al., 2011), visual navigation (Yu and Zhao, 2009; Zheng et al., 2009; Wu et al., 2010), and obstacle recognition and localization (Cai et al., 2009, 2012; Bac et al., 2013), fruit and vegetable recognition and localization. Obstacle recognition includes image segmentation of obstacles, recognition and 3D reconstruction of plants of fruits and vegetables, and recognition of other obstacles, such as poles. Recognition of plants of fruits and vegetables includes recognition of stems, lateral branches, peduncles, and fruits with the same color with plants and leaves. Recognition of fruits and vegetables includes image segmentation of fruits and vegetables, recognition of fruits and vegetables occluded by leaves and branches, and recognition of clustered fruits and vegetables. Because of complexities associated with working objects and environments, determining these functions of the vision system of a harvesting robot is a challenging task. Research studies have been separately undertaken to determine every function of the vision system of a harvesting robot. Fruit and vegetable recognition is widely accepted to be a difficult task. Problems

* Corresponding author at: College of Biosystems Engineering and Food Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou 310058, China. Tel.: +86 0571 88982885.

*E-mail address:* yingyb@zju.edu.cn (Y. Ying).

http://dx.doi.org/10.1016/j.compag.2014.05.006
0168-1699/

associated with fruit and vegetable recognition that could affect the reliability of the vision system include drastic illumination changes, color similarities between withered leaves/branches and fruits, fruits growing in clusters, and fruits occluded by leaves or branches (Plebe and Grasso, 2001). Fruits are known to overlap with each other under natural conditions (e.g., the ratio of clustered to all tomatoes growing in the greenhouse of Zhejiang Agricultural Science Research Institute is 87.5%). Thus, recognition of clustered fruits is a main problem that must be addressed to improve the reliability of harvesting robots.

In the current studies, three recognition methods exist for clustered fruits. The first method is the watershed algorithm (Zhou et al., 2007), which is often used for adhering fruits. This method has problems of over-segmentation and under-segmentation. The second method is the Hough Transform method (Yao et al., 2008), which is based on the idea of shape modeling. This method can recognize clustered fruits based on the created fruit shape models. It has a good anti-jamming capability but has high time cost. The third method, which is based on the geometric shapes of fruits, is often used for quasi-circular fruits (Plebe and Grasso, 2001; Xun et al., 2007). This method is realized according to the following steps: edge points of clustered fruits with abnormal curvature values are removed using a curvature analysis method for all edge points of clustered fruits. Then, the clustered fruits can be recognized after the corresponding circle formulas for the remaining edge points have been created. Moreover, the accuracy rate of this method relies on the lengths of the edges of the clustered fruits, which are not occluded by other objects (such as leaves, branches, or other fruits in the same clustered region).

The aforementioned methods are all realized based on 2D image information. Clustered fruits can be correctly recognized when enough pixels are available for each fruit, which are not occluded by other objects. However, recognizing these fruits remains difficult when the overlapping ratio (i.e., the ratio of the area of the invisible part of a tomato occluded by other tomatoes to the total area of this tomato in a clustered region) is high. Fusing depth information with 2D image information is a better approach to solve this problem. This method is based on 3D image information, which is close to the real world scenes (Jiménez et al., 2000; Paula et al., 2006). The main problem with this method is how to quickly and accurately acquire depth information on fruits. Methods used to acquire depth information are mainly classified into two categories: passive and active range-finding methods (Kondo and Ting, 1998). Active range-finding methods include the laser range-finding method (Lee and Ehsani, 2008), the ultraphonic range-finding method (Regunathan and Lee, 2005), and the position sensing detector-based method (Tanigaki et al., 2008). Passive range-finding methods include those based on the movement of a visual sensor (Van Henten et al., 2002) and on the binocular stereo vision. Active range-finding methods are time consuming because they are realized by light scanning objects horizontally and vertically. The passive range-finding method based on binocular stereo vision is the most well known and widely used in agricultural robotics research because of its better real-time performance (Jiang et al., 2008).

As previously mentioned, these methods for the recognition of clustered fruits have disadvantages. However, in almost all studies, only one of these methods is used for all clustered regions in images, resulting in unsatisfactory performance of the recognition algorithm. Furthermore, high overlapping ratios lead to the occurrence of false negatives (i.e., several fruits are falsely considered as one fruit). In a cluster with most parts of the back fruit hidden by the front fruit, the manipulator of a harvesting robot is likely to be obstructed by the front fruit when the fruit to be harvested is the back fruit, regardless of the correct recognition of all fruits in this cluster. This phenomenon is attributed to the fact that the fruit to be harvested is randomly selected from all fruits in this cluster and that the fruit not hidden by other fruits in this cluster is not always selected first for harvesting. As a result, harvesting cannot be correctly completed.

In this study, after image segmentation and depth map acquisition, clustered tomatoes were divided into two types: adhering and overlapping. This classification was based on the depth difference between the front (i.e., the region with smaller depth values in the clustered region) and back regions (i.e., the region with larger depth values in the clustered region) in each cluster using an iterative Otsu method for the depth map. Different recognition methods were utilized for different types of clustered tomatoes. For adhering regions, tomatoes were recognized using the method based on edge curvature analysis. For overlapping regions, the front region was segmented from the back region after depth map segmentation using the threshold produced from the iterative Otsu method. Then, only the tomatoes in the front region were recognized using the method based on edge curvature analysis. After harvesting the tomatoes in the front region of the overlapping region, the tomatoes in the back region of the overlapping region were subsequently recognized.

The objectives of this study are the following: (1) to realize the recognition of overlapping tomatoes when tomatoes were seriously occluded to each other; (2) to realize the cluster type recognition of clustered tomatoes based on the depth difference between the front and back regions in every clustered region and the recognition of different types of clustered tomatoes based on different recognition methods to improve the recognition accuracy rate of clustered tomatoes; and (3) to realize that the headmost tomato in an overlapping region was the first to be harvested to avoid the situation where the manipulator would be obstructed by the headmost tomato if the robot first harvested the back tomatoes.

This study mainly focused on the recognition of clustered tomatoes, which are slightly occluded by leaves or branches. Despite the limitation of this study, it does suggest a systematic recognition method for clustered tomatoes, which is very popular under natural conditions. The method can also be used for the recognition of other clustered fruits. The recognition of tomatoes seriously occluded by branches or leaves and other problems previously mentioned will be studied in the future.

## 2. Materials and methods

### 2.1. Equipment and image acquisition

The test equipment included the following: a binocular stereo camera (Bumblebee2; Point Grey Research Company, Canada) consisting of two color Sony CCDs (maximum resolution, $1024 \times 768$ and focus length of optical lens, 6 mm); a 1394 image grabber (MOGE) with a power adapter; a 1394 connecting line; a tripod; a laptop (Lenovo R400) with 1 GB RAMs and an Intel Core 2 Duo T6570 CPU; a Windows 7 operating system; and a VC++ 6.0 programming environment.

Plants were grown in flowerpots in a shed in Zhejiang University in China. Image acquisition was carried out when the tomatoes were ripe. The image acquisition conditions were as follows: distance of the image acquisition system, 300–1000 mm; dates, July 1, 2010 to July 25, 2010; time, 5:30 AM to 7:00 PM; weather, cloudy, rainy, overcast, and sunny; and lighting condition, front. Some images were also acquired in a laboratory where the lighting source was a fluorescent light.

### 2.2. Recognition method for clustered tomatoes

The flowchart of the recognition method for clustered tomatoes is shown in Fig. 1.