Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Vision-based extraction of spatial information in grape clusters for harvesting robots

Lufeng Luo [a,b], Yunchao Tang [c,*], Xiangjun Zou [a,**], Min Ye [a], Wenxian Feng [c], Guoqing Li [b]

[a] Key Laboratory of Key Technology on Agricultural Machine and Equipment, Ministry of Education, South China Agricultural University, 483 Wushan Road, Guangzhou 510642, China

[b] Tianjin University of Technology and Education, College of Mechanical Engineering, 1310 Dagun South Road, Tianjin 300222, China

[c] School of Civil and Transportation Engineering, Guangdong University of Technology, 100 Waihuan West Road, Guangzhou 510006, China

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form
23 August 2016
Accepted 25 August 2016
Published online 15 September 2016

Keywords:
Binocular stereo vision
Grape cluster
Bounding volume
Cutting point
Harvesting robots

Grapes are likely to have collisions and be damaged by manipulations when harvesting grape clusters. To conduct an undamaged robotic harvesting, this paper focuses mainly on locating the spatial coordinates of the cutting points on a peduncle of grape clusters for the end-effector and determining the bounding volume of the grape clusters for the motion planner of the manipulator. A method for acquiring spatial information from grape clusters is presented based on binocular stereo vision. This method includes four steps: (1) calibrating the binocular cameras and rectifying the images, (2) detecting the cutting points on the peduncle and the centres of the grape berries, (3) extracting three-dimensional spatial coordinates of the points detected in step 2, and (4) calculating the bounding volume of the grape clusters. A total of 300 images were captured in the vineyard and were tested to validate the method for the cutting point detection, and the success rate was approximately 87%. The accuracy of the localisation of the cutting points was determined under outdoor conditions, and the accuracy in the Z and X directions was 12 mm and 9 mm, respectively. The acquired bounding volume of the grape cluster was compared with manual measurements, and errors in the height and maximum diameter were less than 17 mm and 19 mm, respectively. The elapsed time of the whole algorithm was less than 0.7 s. The demonstrated performance of this developed method indicated that it could be used on harvesting robots.

1. Introduction

Grape harvesting is a time-consuming and labour-intensive procedure, and labour shortages and high labour costs are drastically becoming more serious along with the ageing population in China (Xiang, Jiang, & Ying, 2014). Thus, it is urgently needed to develop a robot for harvesting grapes in the vineyard. In the process of robotic harvesting, to complete an automatic harvest with undamaged grapes, the end-effector of the harvesting robot should avoid collisions with

* Corresponding author.
** Corresponding author.
E-mail addresses: ryan.twain@gmail.com (Y. Tang), xjzou1@163.com (X. Zou).
http://dx.doi.org/10.1016/j.biosystemseng.2016.08.026
1537-5110/

Nomenclature

Symbols

i	serial number of the detected lines
n	serial number of stereo image pairs
N	number of grape cluster samples
S_{GC}	pixel set of the grape cluster
x,y	pixel coordinates in the image
x_c, y_c	barycentre coordinates
x_t, y_t	pixel coordinates of the top point of the grape contour
x_0, y_0	centre coordinates of the berry
(x_{i1}, y_{i1}), (x_{i2}, y_{i2})	pixel coordinates of the detected line end
x_l, y_l	pixel coordinate of S_{GC} in the left image
X_w, Y_w, Z_w	three-dimensional coordinates under world coordinates
x_j, y_j, z_j	spatial coordinates of the front surface centre of the berries
$f(x,y)$	pixel values of (x, y) in the binary image
$L_i(x,y)$	equation of the detected line i
r	radius of the regression circle
θ	angle whose value range was from 0 to 360°
r_t	real radius of the berry
A	two-dimensional accumulator
B	baseline distance between the left and right camera
f	focal length
d	disparity between the left and right camera
L_{max}	maximum Euclidean distance between the left and right points of the grape contour
l_j	distance between the jth berry and the centre axis of the grape cluster
d_{jk}	distance between any two berries' centres, j and k
D_n	actual distance between the cutting point and the reference point in the x direction
$x(CP)_n$	x coordinate value of the cutting point acquired by the stereo-vision system
$x(RP)_n$	x coordinate value of the reference point on the calibration board
$z(CP)_n$	z coordinate value of the cutting point acquired by the stereo-vision system
$z(GT)_n$	depth value of the cutting point acquired by the ground-truth measurement device

Abbreviations

NCC	normalised cross correlation
ROI	region of interest
Roi_L	length of peduncle ROI
Roi_H	height of peduncle ROI

the grape clusters and other obstacles when it moves toward a target (Bac, Hemming, & van Henten, 2014). To plan a collision-free path, the location and bounding volume of the grape clusters must be sent to the motion planner. The grape character such as softness, irregular shape and contour make grasping the fruit directly not feasible, and thus, grasping and cutting the peduncle of the grape is an effective method for harvesting. Therefore, localisation of the cutting points on the peduncle and estimation of the bounding volume of the grape clusters have become essential functions of the harvesting robots.

In recent decades, many researchers have taken great interest in fruit-harvesting robots and have performed some studies on it, such as cucumber harvesting (Van Henten et al., 2003), strawberry harvesting (Hayashi et al., 2010; Kondo, Monta, & Hisaeda, 2001), tomato harvesting (Monta et al., 1998), apple harvesting (Si, Liu, & Feng, 2015; Zhao, Lv, Ji, Zhang, & Chen, 2011), aubergine harvesting (Hayashi, Ganno, Kurosaki, Arima, & Monta, 2003), litchi (Wang, Zou, Tang, Luo, & Feng, 2016) and de-leafing cucumber harvesting (Van Henten et al., 2006). However, most of these studies did not achieve commercialisation successfully because of certain factors, such as low success rates, low work efficiency, high costs, fruit damage and difficulty of detection in unstable illumination (Hayashi et al., 2010). To overcome these problems, the recognition and localisation of fruits via sensor equipment and image algorithms were widely studied by researchers from all over the world.

In the study of sensors, Jimenez, Ceres, and Pons (2000) adopted a laser-based computer vision system for fruit detection. This system was based on an infrared laser range-finder sensor that provided range and reflectance images and was designed to detect spherical objects in unstructured environments. Additionally, Kondo, Shibano, and Mohri (1994) proposed a grape identification method early by using its spectral properties. Mehta and Burks (2014) utilised an inexpensive perspective transformation-based range estimation method that positioned 3D fruits in citrus harvesting based on a monocular camera-in-hand. To recognise clustered tomatoes, Xiang et al. (2014) used a binocular stereo vision to realise the recognition of the clustered tomatoes.

In addition, Berenstein, Ben, Shapiro, and Edan (2010) utilised the difference in the edge distribution between the grape clusters and the foliage to detect grape clusters for an autonomous selective vineyard sprayer. The bunches of grapes were detected in a natural environment based on their colour images (Reis, Morais, & Peres, 2012). The number of grape berries was counted by detecting specular spherical reflection peaks in RGB images obtained at night by artificial illumination (Font, Pallejà, Tresanchez, Teixidó, & Martinez, 2014). The grapes were recognised and localised by extracting the external rectangle of the grape contours for a grape bagging robot (Yang et al., 2013). Kai, Lining, and Zhe (2013) proposed an improved "super green" colour feature model (2G-R-B) to achieve target recognition and positioning of the grape for the bagging robot.

In the detection of fruit peduncles and obstacles, Cubero et al. (2014) presented a pedicel detection method that was implemented by detecting the connecting points of the contours between the peduncle and the fruit, which was suitable for grape berries but not grape clusters. To provide the harvesting robot with an obstacle location, Bac et al. (2014) used a support wire as a visual cue to locate the stem of a sweet-pepper. Van Henten et al. (2006) detected the peduncle of a leaf for a de-leafing cucumber harvesting robot based on two

images at wavelengths of 970 nm and 850 nm. Cai, Zhou, Wang, and Lǜ (2009) restored 3D information of obstacles for a citrus harvesting robot by extracting the skeleton of the branch image. Herrero-Huerta, González-Aguilera, Rodriguez-Gonzalvez, and Hernández-López (2015) reconstructed the 3D bunch grapes by extracting the point cloud using five images (main position, up, down, left, right) to estimate the vineyard yield. However, the method of calculating the bounding volume of the grape clusters through acquiring a point cloud requires a large amount of time for the stereo matching and points handling, and thus, the real-time performance did not meet the requirements of a harvesting robot. Liu, Whitty, and Cossell (2015) proposed a lightweight method for generating a representative 3D reconstruction of an individual grape bunch from a single image from one side of the bunch, but it was difficult to extract the spatial coordinates of the grape cluster by this method. Kicherer, Roscher, Herzog, Förstner, and Töpfer (2015) calculated the length and width of the grape cluster using image processing from a single camera image under an indoor environment.

From the above research status, we can see that most of the current studies focus on the image segmentation of grapes (Reis et al., 2012), grape recognition (Berenstein, Ben Shahar, Shapiro, & Edan, 2010) and yield estimation (Herrero-Huerta et al., 2015). However, there have been a few studies in the literature mentioning the localisation of the cutting point and the bounding volume of the grape clusters for a harvesting robot. To avoid collisions between the robot arm and the grape clusters, this paper therefore focused on a 3D information extracting method for grape clusters. A binocular stereo vision approach that had good real-time performance as the sensor device to extract the available 3D information from the grape clusters was applied. Because the radius of the grapes' peduncle was relatively small, to improve the matching score of the stereo vision (Bac et al., 2014), a baseline of 5 cm was taken. The main steps of this algorithm are as follows: (1) calibrating the cameras and rectifying the images, (2) detecting the cutting point on the peduncle of the grape cluster and the centre of the berries, (3) extracting three-dimensional spatial coordinates of those points detected by step 2, and (4) calculating the bounding volume of the grape clusters.

The novel elements of this approach included mainly two aspects. Firstly, the cutting point on the peduncle was detected by using the geometric constraint method of the shortest distance from the barycentre of the grape cluster to the extension line of the peduncle. Secondly, the maximum section of the grape cluster was acquired by using the convex polygon method based on the 3D coordinates of the front surface centre of the berries' sphere, and the bounding volume of the grape cluster was calculated by rotating the maximum section around the downward centre axis of the grape cluster that goes through the cutting point.

The objectives were as follows: (1) to determine the 3D coordinates of the cutting point on the peduncle for the end-effector in such a way that the harvesting robot can smoothly grasp the grape cluster and cut off the peduncle, and (2) to estimate the bounding volume of the grape clusters for the motion planner in such a way that the robot arm could avoid colliding with the grape clusters and injuring the berries.

2. Materials and methods

2.1. Equipment and image acquisition

The test equipment included the following: a Nikon D5200 digital camera; a binocular stereo vision system that consists of two colour cameras (CMOS, maximum resolution, 600 × 800, focus length of optical lens, 6 mm) with a 5 cm baseline; a calibration plate (the distance of the near circles centres, 25 mm, the number of points, 7 × 7) ordered from MVTec Software GmbH (Germany); a laptop (Lenovo T430) with 4G RAM and Intel(R)Core(TM)i5-3230 M CPU@2.60 GHz, a Windows 7 operating system; OpenCV2.3.1 and a Visual C++ 2008 programming environment. There were different varieties of grapes, and the summer black grapes and red grapes were selected in this paper as test samples.

The image acquisition for the recognition rate test of the cutting point of the peduncle in the vineyard was as follows: grape harvesting may take place under sunny or cloudy sky resulting in drastically different lighting conditions. Considering the characteristics of different ripe times for a different variety of grape (some varieties are ripe in July–August and some in September–October), we captured the test images on July 21, 2014 (sunny day) and October 6, 2014 (cloudy) using the Nikon D5200 digital camera in Tianjin ChaDian grapes science park, to obtain the images of all of the ripe grapes from the different varieties. The acquisition time was from 8:30 AM to 3:00 PM. The auto-exposure control mode based on the shutter speed priority was adopted, and the exposure time was fixed to 1/100 s; 300 images were recorded, most of the grapes were not covered by other objects, and these images were only used to test the success ratio of the cutting point detection.

To determine the positioning accuracy of the cutting point of the peduncle and to estimate the bounding volume of the grape clusters, the stereo images were acquired at an outdoor vineyard on July 21, 2015 through the following processes: firstly, the binocular stereo vision system was fixed in a supporting structure that was placed on the forepart of a length 1600 mm experimental platform. Secondly, a zero line was set below two cameras, and a ruler was installed on the platform. Thirdly, the experimental platform was placed in front of the tested grape cluster. Fourthly, 38 pairs of images were captured for every sample when the supporting structure was moved along the ruler. The moving range was from 350 mm to 1100 mm, which was consistent with the work space of the harvesting robots. The procedure was repeated for a total of 12 grape clusters.

2.2. Three-dimensional information acquisition method for grape clusters

The processing architecture of the 3D information extraction method for the grape clusters consists of four steps, which are sketched in Fig. 1. A complete description of the extraction algorithm is the following.

Starting with the camera calibrations, the parameters of the two cameras were obtained by the algorithm in step 1, and then, the left and right images were rectified by using those parameters to enable stereo matching. In step 2, the cutting

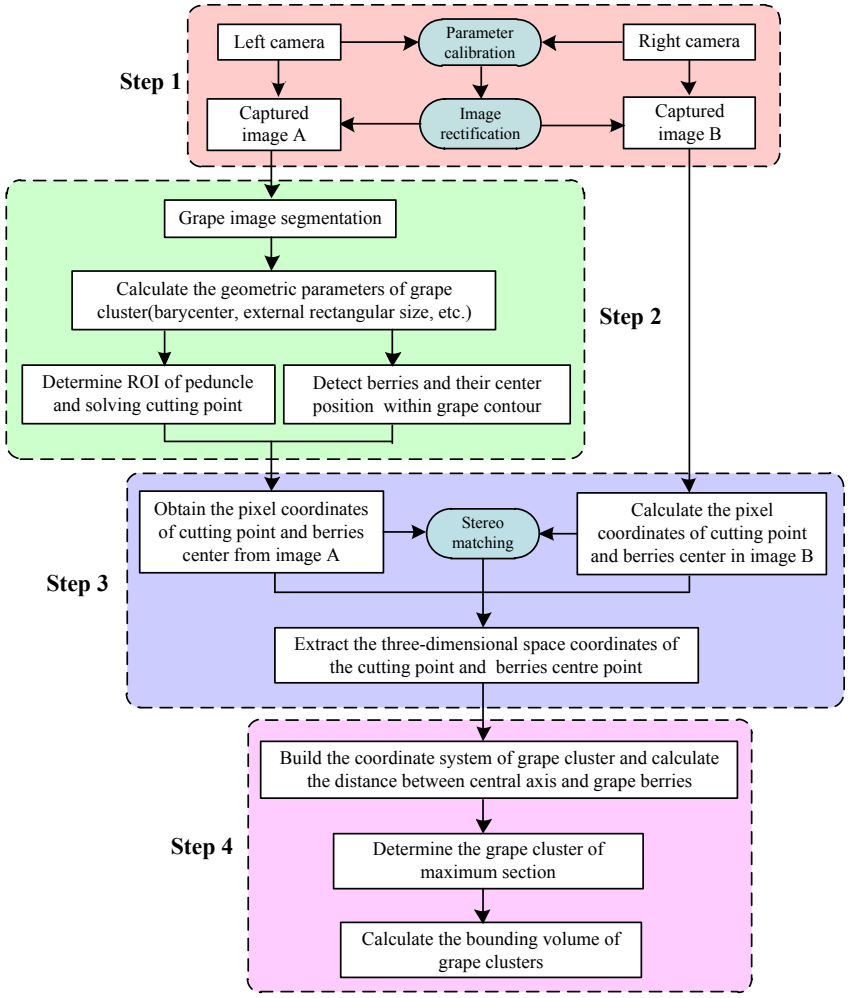

Fig. 1 – Algorithm flowchart for extracting the 3D spatial information of a grape cluster using stereo vision.

point of the grape clusters and the centres of the grape berries in the images captured by the left camera were detected through image processing and a geometric constraint method. After obtaining the pixel coordinates of the cutting point and berries, in step 3, stereo matching was performed, and the 3D coordinates of the points were extracted by a triangulation method. Finally, the bounding volume of the grapes was calculated by solving the maximum section of the grape cluster in step 4. The obtained 3D coordinates of the cutting point and the bounding volume of the grape cluster would be sent to the end-effector of the harvesting robot to accurately grasp and cut the peduncle of the grape cluster.

2.2.1. *Step 1: camera calibration and image rectification*

To extract the three-dimensional coordinates of the grape clusters, intrinsic camera parameters (e.g., focal length) and extrinsic camera parameters (e.g., pose) must be acquired firstly. These camera parameters were determined by a calibration procedure (Steger, Ulrich, & Wiedemann, 2007) in Opencv2.3.1 that used the calibration plate produced by MVTec Software GmbH (Germany). The focal length, the centre pixel of the image, the width and height of a pixel, and one parameter to compensate for any lens distortion were estimated by the procedure (Bac et al., 2014). In addition, the relative poses between the left and right images were also obtained through the calibration procedure. Subsequently, the images were rectified using intrinsic camera parameters. The aim of the image rectification was to make the optical axis parallel in the two cameras to process stereo matching and three-dimensional reconstruction. After rectification, the image rows of the same single point in the left and right image were unanimous. As a result, the images conformed with the epipolar constraints (Brown, Burschka, & Hager, 2003), and thus, in processing the stereo matching, we needed only to search its matching points on the same line, and the efficiency of the stereo matching was improved.

2.2.2. *Step 2: detecting the cutting point on the peduncle of the grape cluster and the centre of the berries*

2.2.2.1. *Image segmentation.* Image segmentation was a key step in the recognition of the grape clusters. The main methods for recognising the fruits in the images were shape-based analysis and colour-based analysis (Si et al., 2015). Considering the irregular contours of the grape clusters, colour-based analysis was adopted in this study. To find a more effective colour indicator, the colour components of

YCbCr, HSV, HIS, L*a*b were extracted, and then, the colour difference between the grapes and environment (e.g., the foliage, limb, sky) was contrasted and analysed. It was found that the H–I component was a more effective indicator for summer black grapes and the Cr component was more effective for red grapes. Figure 2(b) shows the H–I component of Fig. 2(a).

The illumination changes frequently in a vineyard environment, and to minimise the influence produced by the changing illumination on the image segmentation, an adaptive threshold method based on the Otsu algorithm (Otsu, 1979; Sezgin & Sankur, 2004) was used to segment the grape clusters. Figure 2(c) was initially obtained by segmenting Fig. 2(b) using this method. After the image segmentation, some of the noise can still be found in the images, and to eliminate this noise, post-processing was performed, which included a dilation operation, removal of small regions, and maximising the connected region extraction; then, Fig. 2(d) could be obtained. Finally, hole filling and a dilation operation were run on Fig. 2(d) to accurately acquire the set of pixels of the grape cluster, which was then denoted by S_{GC}. Figure 2(e) shows the resulting image after the image segmentation.

2.2.2.2. Calculating the geometric information of the grape clusters. Extracting the contour information of the grape clusters played an important role in solving the cutting point on the peduncle. These geometric parameters included the barycentre, extreme point and bounding rectangle of the contour, which are shown in Fig. 2(f). A binary image of the grape region can be obtained by section 2.2.2.1, i.e., the pixel value of the grape region was set to 255 (white) and the background region to 0 (black). According to the definition of the image moment (Gonzalez & Woods, 2012), the barycentre of the grape region image can be computed by the following formula:

$$\begin{cases} x_c = \sum xf(x,y) \Big/ \sum f(x,y) \\ y_c = \sum yf(x,y) \Big/ \sum f(x,y) \end{cases} \quad (1)$$

where x_c and y_c are the barycentre coordinates, x and y are the pixel coordinates, and $f(x,y)$ are the pixel values of (x,y) in the binary image.

After the barycentre was calculated, the bounding rectangle of the grape region was extracted by solving the extreme value point of up, down, left and right. The top point coordinates (x_t,y_t) and the maximum Euclidean distance between the left and right points of the grape contour L_{max} can be ultimately acquired to determine the region of interest of the peduncle.

2.2.2.3. Determining the cutting point on the peduncle. After extracting the geometric information of the grape clusters, the pixel coordinates of the cutting point on the peduncle was computed by the following steps:

a) Determining the region of interest (ROI) of the peduncle.

Due to the action of gravity, the grape clusters typically hang vertically down over the air, if they have no other upholding object in their way. Thus, the peduncle is located right

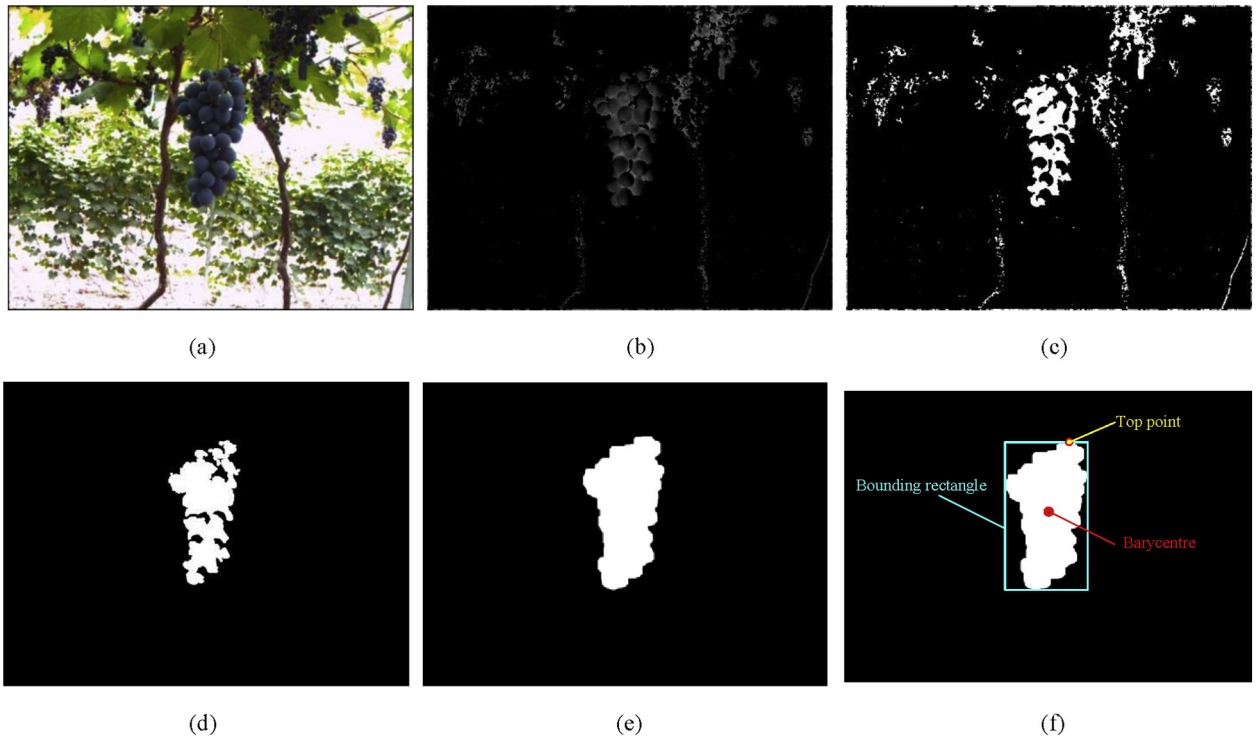

Fig. 2 – An example of image segmentation and the extraction of geometric information. (a) original image, (b) H–I component image, (c) initial segmentation image, (d) post-processing image, (e) resulting images, and (f) extracted geometric information of the grape clusters.

above the grape cluster in most cases. To minimise the influence produced by the complex orchard environment and to improve the speed of finding the cutting point on the peduncle, we considered the morphological character of the grape clusters and the required space of the cutter for clamping the grape. The ROI of the grape peduncle was determined by the barycentre (x_c, y_c), the top point coordinates (x_t, y_t) and the maximum Euclidean distance between the left and right points of the contour (L_{max}). The length of the ROI was calculated by aL_{max}, and the height of the ROI was determined by $b|y_c - y_t|$. Then, the performance of the cutting point detection was tested by adjusting the coefficients a and b between 0~1; when the length was taken to be $0.8L_{max}$ and the height was taken to be $0.5|y_c - y_t|$, the performance of the cutting point detection was the best. Thus, the sizes of the ROI can be computed as follows:

$$\begin{cases} Roi\_L = 0.8L_{max} \\ Roi\_H = 0.5|y_c - y_t| \end{cases} \quad (2)$$

where *Roi_L* is the length of the ROI, and *Roi_H* is the height of the ROI. The centre coordinates of the ROI were set to be (x_c,y_t−0.5*Roi_H*). The geometrical description of the ROI and the schematic diagram of the cutting point calculation model can be observed in Fig. 3(a).

b) Detecting lines in the ROI.

A substantial amount of noise usually exists in the images that are captured from an orchard because of the impact of shadows and varying illumination. The peduncle of the grape clusters resembled a small circular tube, and their images had a bilateral character. To reduce the noise and enhance the edge images, a bilateral filter was applied to the ROI during pre-processing, which was a non-linear, edge-preserving and noise-reducing smoothing filter for the images. This approach combines colours based on both their geometric closeness and their photometric similarity, and near values are preferred to distant values in both the domain and range (Tomasi & Manduchi, 1998). The bilateral filter was applied through calling the function 'bilateralFilter' in Opencv, and the size of the filter mask was fixed at 7, while the photometric spread in the image range was 20.0, and the geometric spread in the domain was 4.0.

Then, the edge images were extracted by using the detecting operator of the edge gradient, which took the local maximum of the gradient as the edge point. The Canny edge detector operator (Canny, 1986) was taken as the edge detecting operator. Figure 3(c) shows the edge image of the ROI of the peduncle in Fig. 3(b).

The edge of the peduncle was close to a straight line. To detect it, a Hough line detection based on the cumulative probability (Matas & Kittler, 2000) was performed. The number of maximum votes of the detection lines was set to sixteen, and the length of the line was not less than twenty pixels. All of the lines that fulfilled these conditions were selected, and then, their endpoint's pixel coordinates were recorded. Figure 3(d) shows the detected lines in the ROI.

c) Calculating the pixel coordinates of the cutting point

After obtaining the endpoint coordinates of the detected lines, the lines can be described by the following mathematical equation:

$$L_i(x, y) = (x - x_{i1})(y_{i2} - y_{i1}) + (y - y_{i1})(x_{i1} - x_{i2}) \quad (3)$$

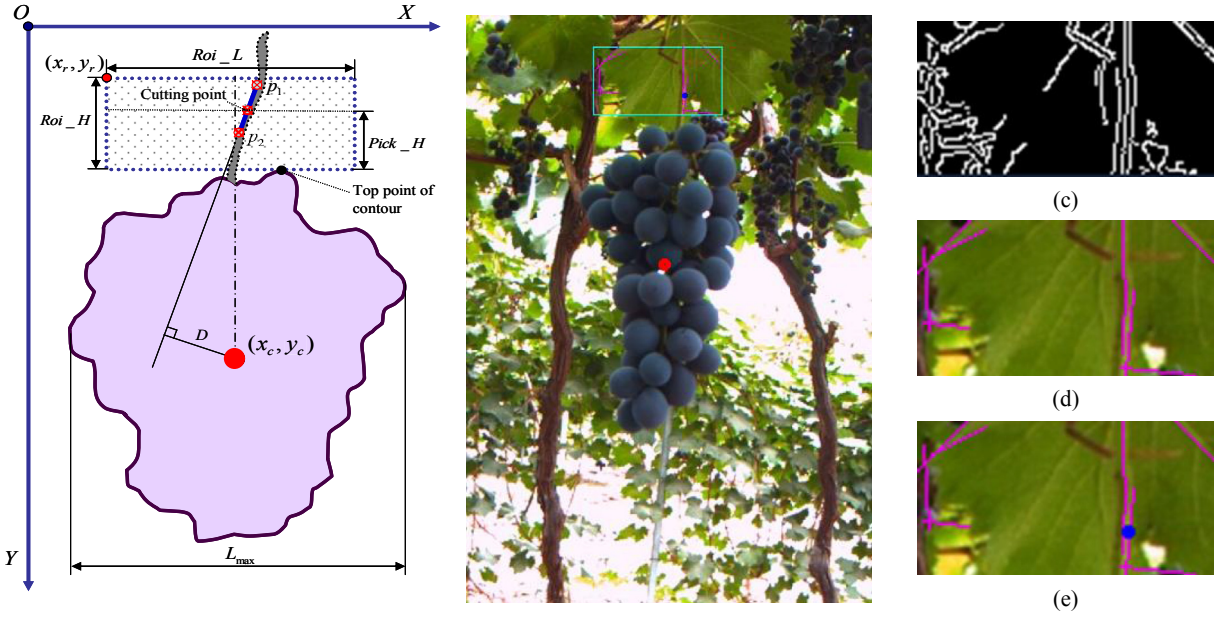

Fig. 3 – Solving the process of finding the cutting point. (a) Schematic diagram of the cutting point calculation model, (b) The detected information on the example image, (c) Edge binary image of the stem ROI, (d) Detected line segment (red lines), (e) Calculated cutting point (blue point).

where (x_{i1}, y_{i1}) and (x_{i2}, y_{i2}) are the end's pixel coordinates, and $L_i(x,y)$ is the straight line equation of i.

The peduncle of the grape clusters was located generally right above the barycentre due to the action of gravity, and thus, the line that passed through the peduncle was bound to get through the barycentre in theory. However, in a real situation, some of the influence factors such as errors in the image segmentation, irregular contours of the grapes, obstructions produced by foliage and branches, and so on could have the detected line on the edge of the peduncle, with a slight deviation from the barycentre. To determine the line that was located on the peduncle, a minimum distance restraint between the barycentre and the detected lines was adopted. The resulting line could be selected by solving the minimum distance between the detected lines and the barycentre. The equation for solving the minimum distance is as follows:

$$D_{result\_line} = min\left|\frac{(x_c - x_{i1})(y_{i2} - y_{i1}) + (y_c - y_{i1})(x_{i1} - x_{i2})}{\sqrt{(y_{i2} - y_{i1})^2 + (x_{i1} - x_{i2})^2}}\right| \quad (4)$$

Finally, the midpoint of the resulting line was chosen as the cutting point of the grape clusters. Figure 3(e) indicated that the blue point was the cutting point.

2.2.2.4. *Detecting the centre and the radius of the berries*. The appearance of the grape berries is approximately circular, and thus, the berries can be detected by using the circle regression method with four steps, as follows. Firstly, the edge image of the berries for the grape clusters was extracted. Secondly, a circle detection was performed in the edge image, and a prediction model of the berries sizes in the images was built to automatically set the parameters for the circle regression (Xiang et al., 2014). Thirdly, three regression rules were constructed to eliminate the redundant circle. Finally, the centre and radius of the berries were obtained.

Above all, to decrease the influence caused by the complex background and enhance the detection speed, the detection region was limited to be within the bounding rectangle region of the grape clusters, obtained as described in section 2.2.2.2. Then, the edge image was extracted by using a filter and Canny edge detector (Canny, 1986). Figure 4(b) shows the edge image of Fig. 4(a).

Secondly, circle detection was conducted by using the least squares method, to minimise the deviation of the squares between the edge points and the regression circle (Xiang et al., 2014). The detection circle equation was as follows (Galambos & Kittler, 2002):

$$\begin{cases} x = x_0 + rcos\theta \\ y = y_0 + rsin\theta \end{cases} \quad (5)$$

where r was the radius of the berry, x_0 and y_0 were the centre coordinates of the berry, x and y were the image coordinates, and θ was the angle, whose value range was from 0 to 360°.

In most cases, the gradient direction of the points on the circle in the images was toward the centre of the circle. To determine the possible centre position of the circle (x_0, y_0), a two-dimensional accumulator A was used to store the score of every possible centre of the circle. If the gradient direction of some of the points was directed to point (x_0, y_0), then the accumulator A would add a score. When A reached the predefined value (13) that was obtained by testing different values and choosing the optimal value, the point (x_0, y_0) was taken as a possible centre of the circle. Then, circle regression was performed within a given radius range based on the possible circle centres, which minimised the deviation of the squares between the edge points and the regression circle by using the least squares method. In the end, an adaptive radius of the circle was acquired.

In the process of circle detection, setting an ideal range for the berries radius was very important to enhance the berry detection precision and speed. However, the radius of a berry varies in the images, which were taken at different distances, and thus, it was difficult to find an invariable range of values that was adaptable to the changing environment. To solve this problem, a prediction model of the berries radius at different distances was found, as follows.

Firstly, seventeen representative berry images were captured at each interval of 50 mm within the distance of 300 mm–1100 mm. The pixel radius of a berry in the images captured at different distances were obtained through manual measurements by Photoshop software. Then, to determine the law of how the berry radius varies at different distances, a mathematical prediction model was built through fitting the discrete data by using a power multiplication method. The pixel radius of the berries in the images captured at different distances and the fitting curve of the mathematical model are illustrated in Fig. 5. The R^2 of the model was 0.9959, which indicated that it was a reasonable prediction model.

$$r = 3185 \times r_t \times z^{-1.16} \quad (6)$$

where z was the distance between the berry and the left camera (mm). Here, r_t was the real radius of the berries (mm).

The largest grape berry and the smallest berry were selected from many grape clusters, and their real diameters were obtained by manual measurement with vernier calipers. In all of the tested samples, the largest and smallest radius of the grape berries r_t were approximately 20 mm and 8 mm, respectively. Then, the maximum radius r_{min} and the minimum radius r_{max} could be calculated based on the prediction model and the real diameter of the selected largest and smallest grape berries, respectively.

Thirdly, after obtaining the prediction model, an automatic parameter setting method was conducted that can initially calculate the range of the radius of the berries based on the distance. Still, some redundant circle can be detected because some of the fragmented berries' edges existed in the images. To eliminate these redundant circles, three rules were built. The first rule was that the distance between any two circles must be greater than the double value of the minimum berry radius. When the distance was smaller than the value, the circle whose circle regression score was lower would be eliminated. The second rule was that the centre coordinate of the circle must be located within the pixels region of the grape clusters that was obtained through the image segmentation described in section 2.2.2.1. The third rule was that the radius of the circle be greater than r_{min} and less than r_{max}. These three rules can be described by the following expressions:

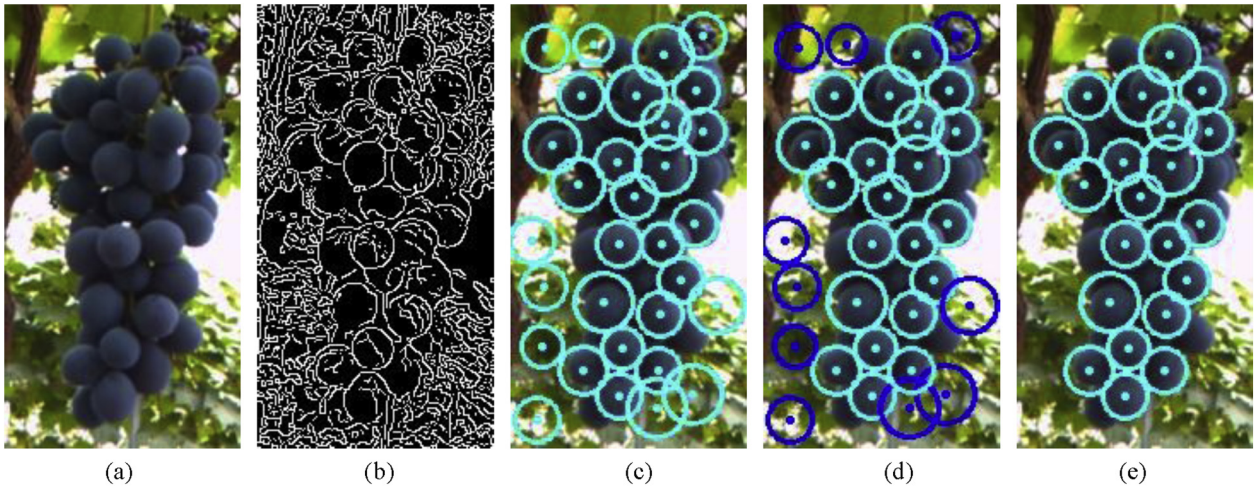

Fig. 4 – Detection process of grape berries. (a) obtained bounding rectangle region of a grape cluster from the example image, (b) edge image of Fig. 4(a), (c) initial berries detection, (d) redundancy circles (blue), (e) the detection result.

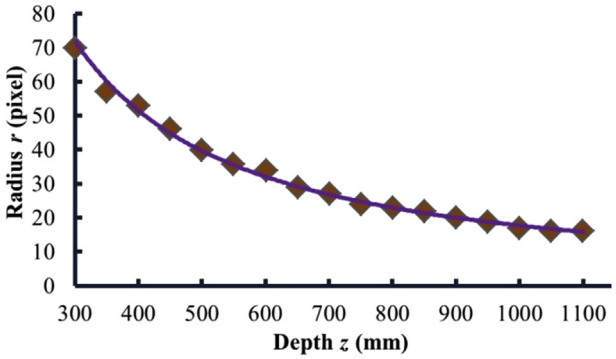

Fig. 5 – The radius of the berries in the images captured at various distances and the corresponding trend curve; the fitting curve of the mathematical model was $r = 3185 \times r_t \times z^{-1.16}$, and the R^2 of the model was 0.9959.

$$\begin{cases} d_{jk} \geq 2r_{min} \\ (x_0, y_0) \in S_{GC} \\ r_{min} \leq r \leq r_{max} \end{cases} \quad (7)$$

where d_{jk} denoted the distance between any two berries' centres.

Finally, the berries were detected by using the above circle regression method. Additionally, Fig. 4(c) was the initial detection, and the blue circle in Fig. 4(d) was a redundant circle, Fig. 4(e) showed the final detection result for the berries. The quantitative information of the detection rate at different depths is shown in Table 1.

2.2.3. *Step 3: extracting three-dimensional spatial coordinates*

2.2.3.1. Stereo matching of the cutting point and the centre of the berries. A pair of aligned images was obtained through the parameter calibration and image rectification in section 2.2.1, and then, the cutting point and the centre of the berries in left aligned images were obtained. To extract the three-dimensional coordinates of the cutting point and the centre of the berries, the disparity of these points between the right and left image must be calculated in advance.

The process of calculating a given point's disparity can be viewed as the matching process of the point region. Firstly, a rectangular window in the left image was set by using the point as the centre. Then, the most similar windows in the right image were searched along the epipolar line by using the maximum similarity measuring method. Finally, the centre of the found window was the resulting point.

The dense disparity calculation method based on the matching of the grey value (Steger et al., 2007) was adopted, of which the computational accuracy and time consumption varied with the similarity measuring methods. The measuring methods, which had been used frequently, included the sum of the absolute gray value differences, the sum of the squared gray value differences and the normalised cross correlation (*NCC*). To enhance the performance of the stereo matching under outdoor lighting, the robust method based on *NCC*, which can overcome a changing illumination better, was used to measure the similarity between the given rectangular window and the target window.

A window of size 9 × 9 pixels was taken, which was approximately a quarter of the diameter of a berry when the

Table 1 – The detection rate for the grape berries at different depths.

Depth/mm	Number of visible berries in images/mm	Number of detected berries/mm	Detection ratio/%
400	27	22	81.48
500	27	22	81.48
600	27	21	77.78
700	27	21	77.78
800	26	20	76.92
900	26	19	73.08
1000	26	18	69.23
1100	26	17	65.38

Note: The berries that were occluded partly were occluded by less than 1/3 compared with the visible berries.

distances of the images captured was approximately 800 mm because smaller windows resulted in 'salt and pepper noise', and larger windows resulted in 'foreground fattening' (Hu & Mordohai, 2012). The disparity was related to the depth, and it was decreasing along with the increasing depth. To reduce the disparity range that was searched and to improve the speed of the stereo matching, the minimum and maximum disparities were set to correspond to the depth range of 350 mm–1100 mm. To match these points accurately, the lowest correlation score between the left and right image was set to 0.92, which was gained by testing the stereo images. The stereo matching results of the cutting point and berries are shown in Fig. 6.

2.2.3.2. *Calculating the 3D coordinates of the cutting point and the berries' centre point.* After working out the disparities of the cutting point and the centre of the berries, the three-dimensional coordinates (X_w, Y_w, Z_w) can be extracted by triangulation in Eqn. (8) (Rodriguez & Aggarwal, 1990).

$$\begin{cases} X_w = \dfrac{x_l \times B}{d} \\ Y_w = \dfrac{y_l \times B}{d} \\ Z_w = \dfrac{f \times B}{d} \end{cases} \tag{8}$$

where X_w is the direction of the plant row, Y_w is the height-direction, Z_w is the depth, B (mm) is the baseline distance between the left and right camera, x_l and y_l represent the pixel coordinates of S_{GC} in the left image, f (mm) is the focal length, and d (pixels) is the disparity. The 3D grape clusters were drawn by using OpenGL according to the 3D coordinates of the cutting point and berries. Figure 6(c) shows an example of a 3D grape cluster.

2.2.4. *Step 4: calculating the bounding volume of the grape clusters*

After extracting the spatial coordinates of the cutting point and the berries, to estimate the three-dimensional bounding volume of the grape clusters, a 3D coordinate system of *XYZ* was established, which used the cutting point as the origin and the downward *Y* axis as the centre axis of the grape cluster. Figure 7(a–b) shows the 3D coordinate system. The coordinates of all of the detected berries were transformed to the *XYZ* coordinate system. Then, the distances between each of the detected berries and the centre axis were calculated by the following equation:

$$l_j = \sqrt{x_j^2 + z_j^2} \tag{9}$$

where l_j was the distance between the *j*th berry, and the centre axis *Y*, (x_j, y_j, z_j) were the coordinates of the front surface centre of the berries sphere under the *XYZ* coordinate system.

Then, the 2D coordinate system of *Y*–*L* as shown in Fig. 7(c) was established, which used the cutting point as the origin. All of the berries (l_j, y_j) are shown in Fig. 7(c). Then, the points of the maximum external border were identified by using the convex polygon method. Figure 7(c) shows the maximum section of the grapes. The bounding volume of the grape cluster was acquired by spinning 360° to the maximum section of grape clusters around the downward centre axis *Y*. Figure 7(d) shows the bounding volume of the grape cluster that was drawn by using OpenGL.

3. Experiments and results

To validate the method proposed in this paper, three experiments were performed. Firstly, the detected cutting point of the grape clusters at the vineyard was visually and quantitatively assessed. Secondly, three-dimensional localisation of the cutting point was conducted, and the localisation accuracy was determined under an outdoor vineyard. Thirdly, the bounding volume estimate of the grape clusters was implemented through comparing the measurement of the stereo-vision and the manual measurement. In addition, the time consumption of the algorithm was calculated to test the real-time performance of this method.

3.1. *Detection of the cutting point*

The working environment of the grape harvesting robot was usually in a greenhouse or outdoor vineyard, where the illumination conditions varied. To validate the robustness of this

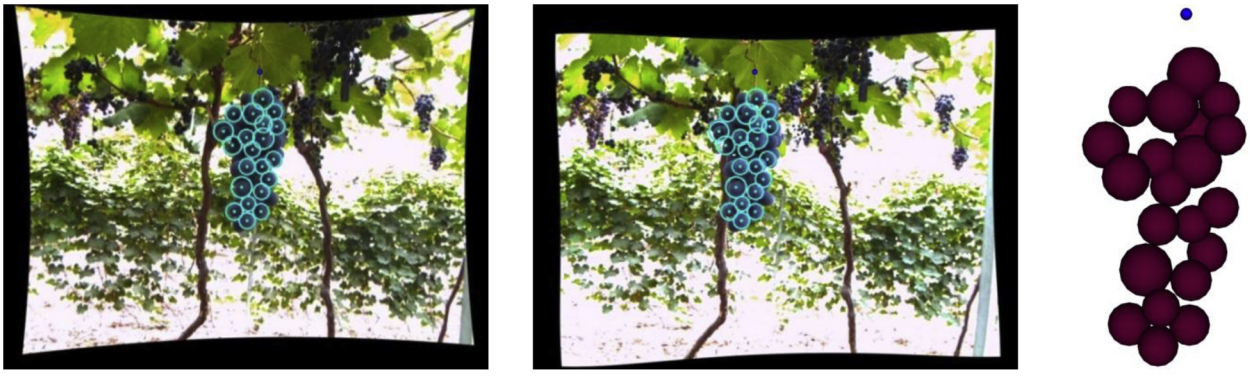

Fig. 6 – An example of stereo-matching on the cutting point and berries. (a)The detection results of the cutting point and berries in the left image, (b) the results of the stereo-matching in the right image, (c) reconstruction of the 3D grape cluster.

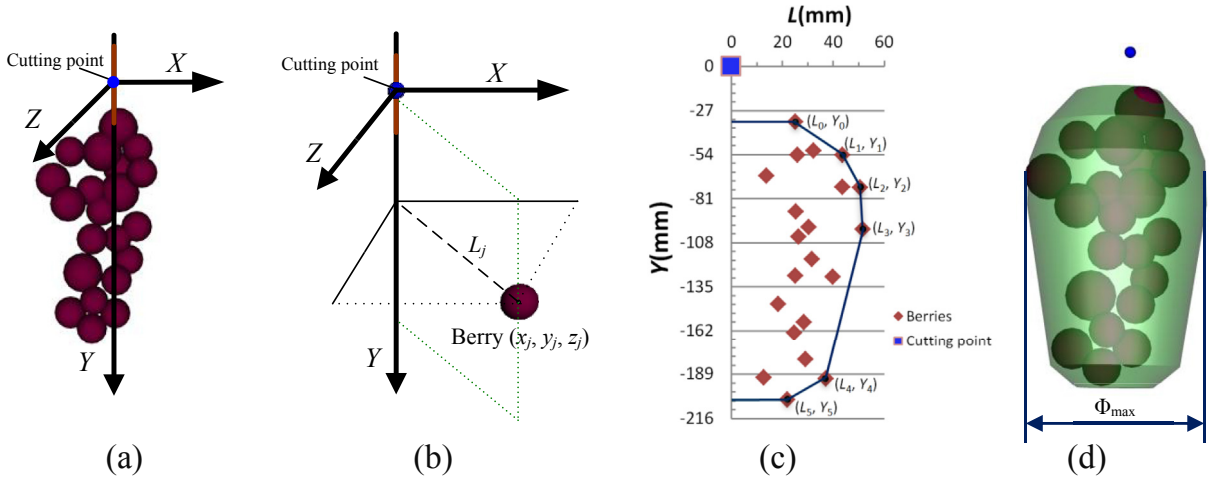

Fig. 7 – The sketch of the coordinate system and the calculation process for the bounding volume of the grape cluster. (a) An example image and its spatial coordinate system. (b) The *XYZ* coordinate system that was used to solve the distance between the centre axis and the berries. (c) The *Y*–*L* coordinate system that was used to solve the maximum section of the grape cluster. (d) The calculated bounding volume of the example.

algorithm under varying lighting conditions and complex backgrounds, 300 images that were captured from different perspectives at the vineyard were tested using the cutting point detection method proposed in Section 2.2.2. Among the 300 images, there are 150 summer black grape images and 150 red grape images, and among every 150 images, 50 were collected in overcast lighting conditions, 50 were collected in sunny overshadow, and the remaining images were collected in sunny frontlight. The grape clusters were not largely occluded by branches or leaves, and the photograph distance was approximately 350–1100 mm. The tests were performed at three different lighting conditions.

Figure 8 shows the detection results of the cutting point under three lighting conditions. The peduncle of grapes was detected by using Hough line detection first. Because the peduncle edge image usually had two or more sides, there were several lines that would be detected on the edge of the peduncle in most cases, as shown in Fig. 8. For example, there are three lines that were detected on the peduncle in Fig. 8(a), and two lines in Fig. 8(b) and (c), and eventually, the cutting point was located on one of the lines because the distance between the line and the barycentre is the smallest. The cutting point located on any one line of the peduncle would be viewed as a successful detection, and in contrast, it was a false detection. The test results are shown in Table 2. From this table, we know that the average success rate of detecting the cutting point was 87%, and the success rate was highest in sunny frontlight conditions and lowest in sunny overshadow.

The method achieved an average 92%, 82% and 83% success rate for summer black grapes on sunny frontlight, sunny overshadow and overcast light conditions, respectively, and 90%, 84% and 86% success rates for red grapes under the same three lighting conditions. The results for the highest levels of false detections on summer black and red grapes were 18% and 16%, respectively, under sunny overshadow. The main reason for the false detection was that the background was mistakenly recognised as the grapes because of the poor quality images that were captured under sunny overshadow. In addition, a strangely shaped peduncle made the lines difficult to detect and finally resulted in a cutting point false detection.

In the experiment, some of the grape clusters that were not all in the field of view or were partially occluded were tested also, under these situations; when their peduncles were not covered by other obstacles, the cutting point could still be well calculated using the proposed method. For example, Fig. 9(a) showed that the cutting point was precisely located on the peduncle although the right part of the grape cluster was not visible. When the grape clusters were partially occluded by leaves, the cutting point could still be detected, as is shown in Fig. 9(b). However, when the peduncles were largely occluded, false or failure line detections on the peduncle would appear, and it would be difficult to accurately locate the cutting point.

3.2. *Localisation precision testing of the cutting point*

After obtaining the pixel coordinates of the cutting point, extracting its three-dimensional coordinates played an important role in conducted successful grape harvesting. To test the precision of the vision position, localisation testing of the cutting point was performed under outdoor conditions, and the accuracy in the z-direction was determined because the literature indicated that most of the inaccuracy can be expected in that direction (Van Henten et al., 2002). In addition, the accuracy in the *x*-direction was also tested. To analyse the positional error of the cutting point in the direction of the *x*-direction, the calibration board as a reference was placed at the peduncle, to ensure that the calibration plane and the peduncle were in contact with each other. Moreover, the normal vector of the calibration plane was parallel to the *Z* axis, and the diagram is shown in Fig. 10. A total of 38 pairs of stereo images were captured for every sample by the binocular camera at different distances in the outdoor vineyard,

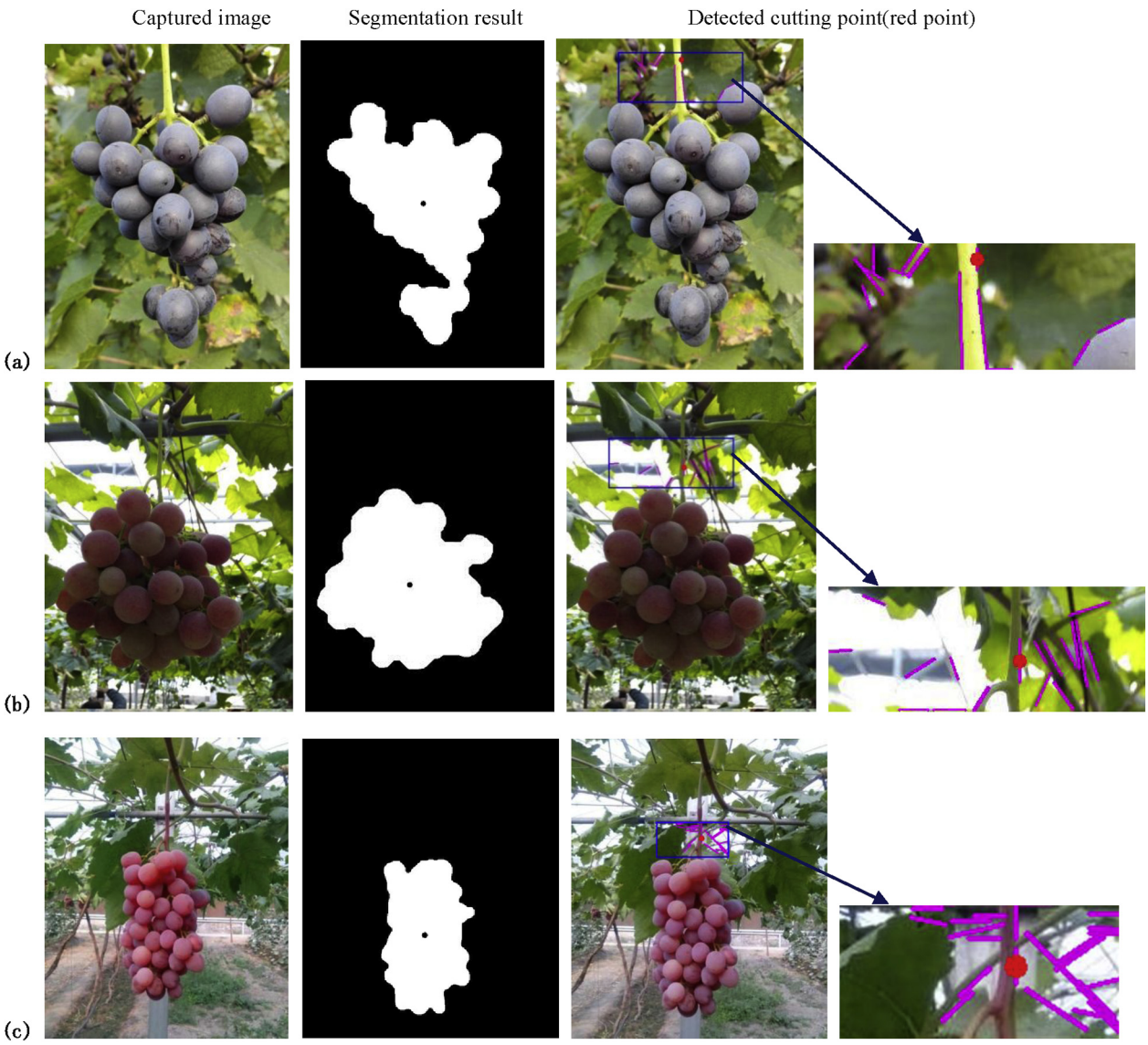

Fig. 8 – Detection results of the cutting point: (a) image captured in sunny frontlight conditions, (b) image captured in sunny overshadow conditions, and (c) image captured in overcast light conditions.

Table 2 – Test results of the cutting point detection under three different light conditions.

Grape varieties	Light conditions	Images (frame)	Amount (frame)		Rate (%)	
			Success	False	Success	False
Summer black	Sunny frontlight	50	46	4	92	8
	Sunny overshadow	50	41	9	82	18
	Overcast light	50	44	6	88	12
Red grapes	Sunny frontlight	50	45	5	90	10
	Sunny overshadow	50	42	8	84	16
	Overcast light	50	43	7	86	14
Total		300	261	39	87	13

and the photograph distance was from 350 mm to 1100 mm. The world coordinate system was set on the left camera, and then, the cutting point of the coordinates was extracted in the world coordinate system. The localisation of the average errors in z direction (e_z) and x direction (e_x) was calculated by the following two formulas:

$$e_z = \frac{\sum_{n=1}^{N} \left| z(CP)_n - z(GT)_n \right|}{N} \tag{10}$$

$$e_x = \frac{\sum_{n=1}^{N} \left| x(CP)_n - x(RP)_n - D_n \right|}{N} \tag{11}$$

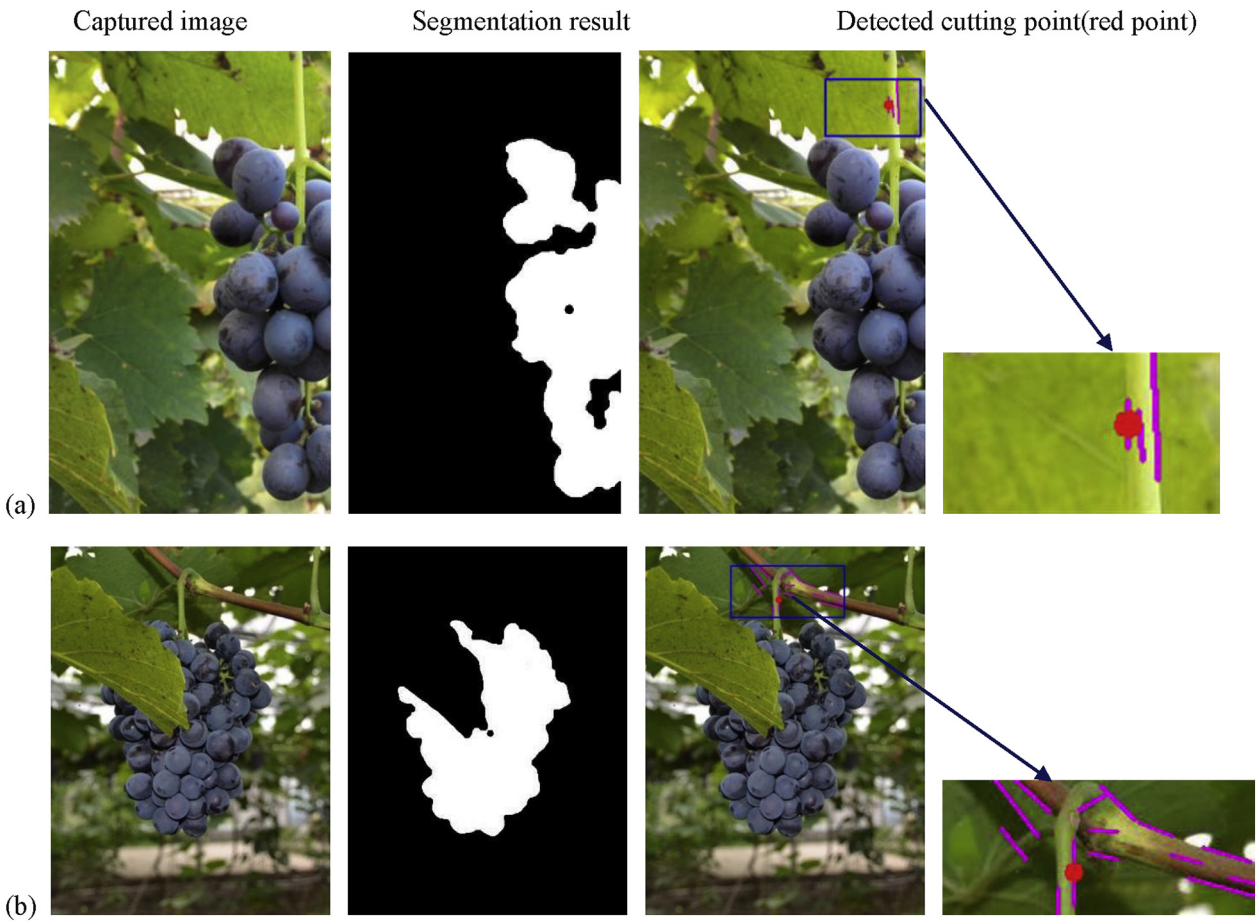

Fig. 9 – –The detection results for the cutting point under occlusion and incompletely visible grape clusters: (a) the grape cluster was not all in sight, (b) the grape cluster is occluded partially by leaves.

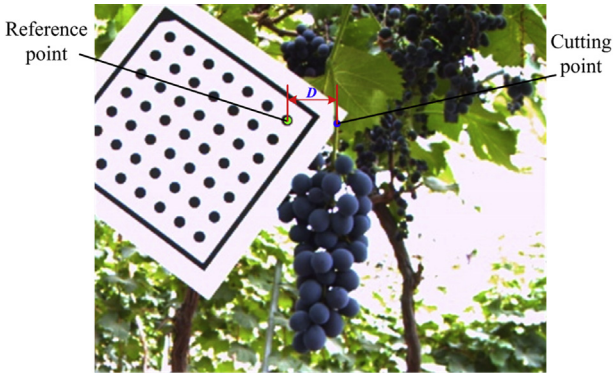

Fig. 10 – –The diagram of the reference point on the calibration board.

where $z(CP)_n$ and $x(CP)_n$ indicate the z and x coordinate values of the cutting point acquired by the stereo-vision system, $z(GT)_n$ is the depth value of the cutting point acquired by the ground-truth measurement device, $x(RP)_n$ are the x coordinate values of the reference point on the calibration board, D_n is the actual distance between the cutting point and the reference point in the x direction, which was measured by using a vernier caliper, n was the serial number of the tested grape sample, and N was the total number of samples.

Figure 11(a) shows the relationship of the actual photograph distance and the z-direction average error. As shown in Fig. 11(a), the accuracy of the z-direction was less than 12 mm; the error in the z-direction reached the smallest value when the depth was between 750 mm and 850 mm, and it would increase with increasing depth. Figure 11(b) indicated the relationship of the actual photograph distance and the x-direction average error. As shown in Fig. 11(b), the accuracy in the x-direction was less than 9 mm, and the error in the x-direction reached the smallest value when the depth was almost 800 mm; it would increase with increasing depth when the depth was greater than 850 mm. Thus, the best distance between the camera and grapes was approximately 800 mm.

3.3. *Assessment of the bounding volume and the real-time algorithm*

To accomplish harvesting with undamaged grapes, a collision-free path must be planned. Obtaining the bounding volume of the grape clusters was one of the important prerequisites for path planning. To validate the feasibility of the proposed method, an assessment test of the bounding volume of the grape clusters was conducted. Taking 12 representative grape clusters as test samples, firstly the maximum radius and height of these grape clusters were acquired by manual

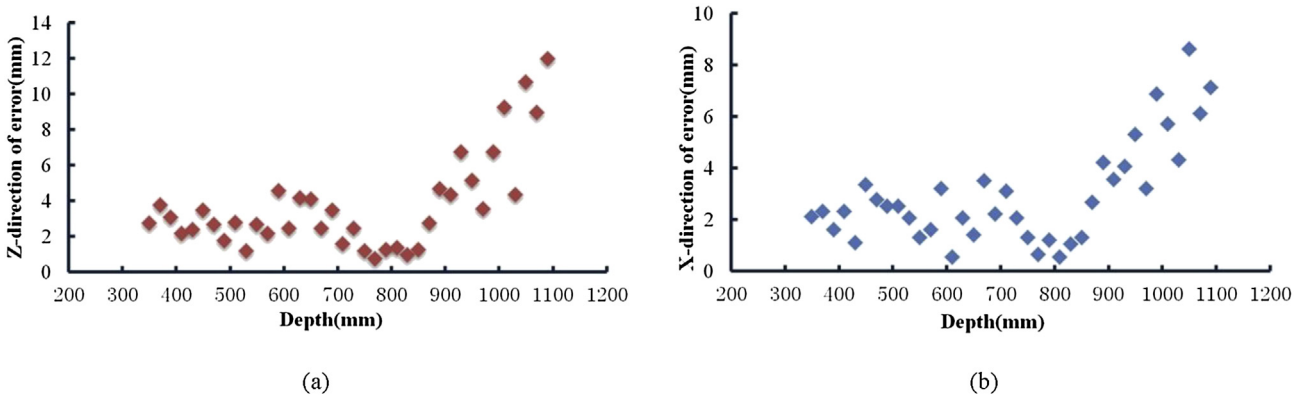

Fig. 11 – Error of the stereo-vision system, measured under outdoor conditions. (a) localisation error in the z-direction, and (b) localisation error in the x-direction.

measurement. The maximum diameter was measured by using an adjustable annulus and vernier caliper, and the height was measured using a vernier caliper. Then, the bounding volume of these samples were calculated through the proposed method, and subsequently, the maximum diameter and height of the bounding volume were extracted. The comparison data between the measurement of the stereo-vision and manual measurement showed that the extracted height was always less than the manual measurement and its maximum error was approximately 16.2 mm. The error in the maximum diameter was less than 18.6 mm.

In addition, the real-time performance of the grape clusters' spatial information extraction was very important for the robotic harvesting, and it determined the harvesting efficiency of the robots. To determine whether the proposed method meets the requirements of robotic harvesting, the elapsed time of the whole process from image segmentation to calculating the bounding volume of the grape clusters was obtained using the clock function in the C language. The results showed that the elapsed time of every pair of images was less than 0.7 s.

4. Discussion

The cutting point on the peduncle of grape clusters was detected by combining the detected peduncle edge line detection and the minimum distance restraint between the image barycentre and the detected line, which was a new and adoptable method for harvesting robots. From the test of the cutting point detection, we can know that the success rate was approximately 87%, and this method was useful to the vari-colour's peduncle, but we found that there were still some factors that affected the performance of the proposed algorithm. Firstly, if the peduncle of grapes was largely covered by leaves, then the peduncle edge line would not be detected smoothly, which would make the detection of the cutting points more difficult. Secondly, the dramatically changing illumination in the vineyard brought substantial difficulty to the image segmentation. Thirdly, when the grape clusters were overlapping each other or occluded largely by other objects, the proposed method would not be very functional. Therefore, further research should focus on the recognition of overlapping and occluded grape clusters. The accuracy of the image segmentation of detecting grapes and berries in an outdoor environment must also be further enhanced to apply the proposed method to grape harvesting robots in preference to other methods.

Errors in the cutting point localisation were mainly caused by false detected cutting points, camera calibration errors and inaccurate stereo matching (Zou, Zou, & Lu, 2012). In the localisation test, the localisation accuracy of the cutting point in the z and x directions under outdoor conditions were less than 12 mm and 9 mm, respectively. The test results indicate that the fault-tolerant range of the designed grasp-cutting mechanism in the z and x direction should be greater than 12 mm and 9 mm, respectively, to enable the harvesting robots to grasp the grapes and cut off the peduncle accurately (Zou et al., 2016).

The errors in the bounding volumes of the grape clusters in the height and maximum diameter were 16.2 mm and 18.6 mm, respectively. The height error was mainly caused by the fact that the topmost or bottom berries of the grape cluster were not detected, and the error in the maximum diameter was caused by certain factors, such as false berries detection and inaccurate stereo matching. The colour of adjoining berries was very similar, and as a result, sometimes the stereo matching of the berries was not satisfactory. Further research would be needed to improve the accuracy of the grape berries detection and their stereo matching a vineyard environment. Calculating the bounding volume for the grape clusters could also be used to yield an estimation of the vineyard and a weight estimation.

In addition, from the real-time algorithm in the test, the elapsed time at each localisation was less than 0.7 s; thus, the real-time performance can meet the requirement of robotic harvesting well. However, when more overlapping or occluded grape clusters are in the line-of-sight of the camera, the image processing time might be longer.

5. Conclusions

A successful robotic and undamaged harvesting method for grapes firstly requires a vision system that acquires the spatial coordinates of the cutting point and the bounding volume of the grape clusters accurately to enable the robot arm to plan a collisions-free path when it approaches a candidate grape for

harvesting. This paper proposed an effective localisation method for the cutting point of the peduncle and a new estimation method for the bounding volume of grape clusters based on binocular stereo vision. Three tests were performed to validate the proposed methods. The following conclusions can be obtained based on these test results:

(1) When the grape clusters and their peduncle were not seriously occluded by other objects, the average success rate of the cutting point recognition was approximately 87%.
(2) In the range of 350–1100 mm, the accuracy of the 3D localisation of the cutting point in the *z* and *x* direction under outdoor conditions was 12 mm and 9 mm, respectively. The results can be used to guide a fault-tolerant design of the end-effector for grape harvesting robots.
(3) The height and the maximum diameter of the bounding volume of the grape clusters were tested by comparing the stereo-vision measurement with manual measurements, and the accuracy of the height and the maximum diameter were approximately 17 mm and 19 mm, respectively. The real-time performance was investigated, and the elapsed time of each localisation was less than 0.7 s.

In conclusion, within the depth range from 350 mm to 1100 mm, this method can meet the need of the harvesting robots when the grape clusters were not overlapped or occluded. Moreover, the bounding volume of the grape clusters can also be used to estimate the weight of the grape clusters and then to guide the end-effector to allocate an appropriate grasping force. Moreover, it can be useful to other cluster fruits, such as the longan and litchi. However, the localisation accuracy was not satisfactory when multiple grape clusters overlapped one another or were largely occluded. In further studies, we will focus on the localisation of multiple overlapping grape clusters and the recognition of grape clusters that are occluded by other obstacles.

Acknowledgements

The authors thank Xiping Song and Cong Zhang for their contributions to test unit set-up and images capture. We are grateful to executive Na Liu from Tianjin Chadian grape science park who allowed us to perform experiments in their vineyard. This work was supported by a grant from the National Natural Science Foundation of China (No. 31571568), the Project of Province Science and Technology of Guangdong (No. 2014A010104011, 2015A020209111) and a grant from Tianjin University of Technology and Education (KJ15-04).

REFERENCES

Bac, C. W., Hemming, J., & van Henten, E. J. (2014). Stem localization of sweet-pepper plants using the support wire as a visual cue. *Computers and Electronics in Agriculture, 105*, 111–120.

Berenstein, R., Ben Shahar, O., Shapiro, A., & Edan, Y. (2010). Grape clusters and foliage detection algorithms for autonomous selective vineyard sprayer. *Intelligent Service Robotics, 3*, 233–243.

Brown, M. Z., Burschka, D., & Hager, G. D. (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 993–1008.

Cai, J., Zhou, X., Wang, F., & Lǜ, Q. (2009). Obstacle identification of citrus harvesting robot. *Transactions of the Chinese Society of Agricultural Engineering, 40*(11), 171–175 (in Chinese with English abstract).

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(6), 679–698.

Cubero, S., Diago, M. P., Blasco, J., Tardáguila, J., Millán, B., & Aleixos, N. (2014). A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis. *Biosystems Engineering, 117*, 62–72.

Font, D., Pallejà, T., Tresanchez, M., Teixidó, M., & Martinez, D. (2014). Counting red grapes in vineyards by detecting specular spherical reflection peaks in RGB images obtained at night with artificial illumination. *Computers and Electronics in Agriculture, 108*, 105–111.

Galambos, C., Kittler, J., & Matas, J. (2002). Gradient-based progressive probabilistic hough transform. *IEEE Vision Image and Signal Processing, 148*(3), 158–165.

Gonzalez, R. C., & Woods, R. E. (2012). *Digital image processing* (3rd ed.). China: Publish House of Ecectronics Industry (Chapter 3).

Hayashi, S., Ganno, K., Kurosaki, H., Arima, S., & Monta, M. (2003). Robotic harvesting system for eggplants trained in V-shape (part 2). *Journal of Society of High Technology in Agriculture, 15*(4), 211–216.

Hayashi, S., Shigematsu, K., Yamamoto, S., Kobayashi, K., Kohno, Y., Kamata, J., et al. (2010). Evaluation of a strawberry-harvesting robot in a field test. *Biosystems Engineering, 105*, 160–171.

Herrero-Huerta, M., González-Aguilera, D., Rodriguez-Gonzalvez, P., & Hernández-López, D. (2015). Vineyard yield estimation by automatic 3D bunch modelling in field conditions. *Computers and Electronics in Agriculture, 110*, 17–26.

Hu, X., & Mordohai, P. (2012). A quantitative evaluation of confidence measures forstereo vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*, 2121–2133.

Jimenez, A. R., Ceres, R., & Pons, J. L. (2000). A vision system based on a laser range-finder applied to robotic fruit harvesting. *Machine Vision and Applications, 11*, 321–329.

Kai, Z., Lining, Z., & Zhe, S. (2013). Design and experiment of intelligent grape bagging robot. *Applied Mechanics and Materials, 389*, 706–711.

Kicherer, A., Roscher, R., Herzog, K., Förstner, W., & Töpfer, R. (2015). Image based evaluation for the detection of cluster parameters in grapevine. *Acta Hortic, 1082*, 335–340. http://dx.doi.org/10.17660/ActaHortic.2015.1082.46.

Kondo, N., Monta, M., & Hisaeda, K. (2001). Harvesting robot for strawberry grown on annual hill top (part 2). *Journal of Societyof High Technology in Agriculture, 13*(4), 231–236.

Kondo, N., Shibano, Y., & Mohri, K. (1994). Basic studies on robot to work in vineyard (part 2). *Journal of the Japanese Society of Agricultural Machinery, 56*(1), 45–53.

Liu, S., Whitty, M., & Cossell, S. (2015). A lightweight method for grape berry counting based on automated 3D bunch reconstruction from a single image. In *Workshop on Robotics in Agriculture, Seattle, USA*.

Matas, C. G., & Kittler, J. (2000). Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision Image Understanding, 78*, 119–137.

Mehta, S. S., & Burks, T. F. (2014). Vision-based control of robotic manipulator for citrus harvesting. *Computers and Electronics in Agriculture, 102*, 146–158.

Monta, M., Kondo, N., Ting, K. C., Giacomelli, G. A., Mears, D. R., Kim, Y., et al. (1998). Harvesting end-effector for inverted single truss tomato production systems. *Journal of the Japanese Society of Agricultural Machinery, 60*(6), 97–104.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man and Cybernetics, 9*(1), 62–66.

Reis, M. J. C. S., Morais, R., & Peres, E. (2012). Automatic detection of bunches of grapes in natural environment from colour images. *Journal of Applied Logic, 10*, 285–290.

Rodriguez, J. J., & Aggarwal, J. K. (1990). Stochastic analysis of stereo quantization error. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 467–470.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging, 13*(1), 146–165.

Si, Y., Liu, G., & Feng, J. (2015). Location of apples in trees using stereoscopic vision. *Computers and Electronics in Agriculture, 112*, 68–74.

Steger, C., Ulrich, M., & Wiedemann, C. (2007). *Machine vision algorithms, machine vision algorithms and applications*. Weinheim, Germany: Wiley-VCH (Chapter 3).

Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and colour images. In *Proceedings of the IEEE International Conference on Computer Vision, Bombay, India* (pp. 839–846).

Van Henten, E. J., Hemming, J., Van Tuijl, B. A. J., Kornet, J. G., Meuleman, J., Bontsema, J., et al. (2002). An autonomous robot for harvesting cucumbers in greenhouses. *Autonomous Robots, 13*, 241–258.

Van Henten, E. J., Van Tuijl, B. A. J., Hemming, J., Kornet, J. G., Bontsema, J., & Van Os, E. A. (2003). Field test of an autonomous cucumber picking robot. *Biosystems Engineering, 86*(3), 305–313.

Van Henten, E. J., Van Tuijl, B. A. J., Hoogakker, G. J., Van Der Weerd, M. J., Hemming, J., Kornet, J. G., et al. (2006). An autonomous robot for de-leafing cucumber plants grown in a high-wire cultivation system. *Biosystems Engineering, 94*(3), 317–323.

Wang, C., Zou, X., Tang, Y., Luo, L., & Feng, W. (2016). Localisation of litchi in an unstructured environment using binocular stereo vision. *Biosystems Engineering, 145*, 39–51.

Xiang, R., Jiang, H., & Ying, Y. (2014). Recognition of clustered tomatoes based on binocular stereo vision. *Computers and Electronics in Agriculture, 106*, 75–90.

Yang, Q., Liu, C., Xun, Y., Bao, G., Wang, Z., & Huang, P. (2013). Target recognition for grape bagging robot. *Transactions of the Chinese Society for Agricultural Machinery, 44*(8), 234–239 (in Chinese with English abstract).

Zhao, D., Lv, J., Ji, W., Zhang, Y., & Chen, Y. (2011). Design and control of an apple harvesting robot. *Biosystems Engineering, 110*, 112–122.

Zou, X., Ye, M., Luo, C., Xiong, T., Luo, L., Chen, Y., et al. (2016). Fault-tolerant design of a limited universal fruit-picking end-effector based on vision positioning error. *Applied Engineering in Agriculture, 32*(1), 5–18.

Zou, X., Zou, H., & Lu, J. (2012). Virtual manipulator-based binocular stereo vision positioning system and errors modeling. *Machine Vision and Applications, 23*(1), 43–63.