Available online at www.sciencedirect.com

ScienceDirect

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

Research Paper

Localisation of litchi in an unstructured environment using binocular stereo vision

Chenglin Wang [a], Xiangjun Zou [a,*], Yunchao Tang [b], Lufeng Luo [a], Wenxian Feng [b]

[a] College of Engineering, South China Agricultural University, Guangzhou 510642, China
[b] School of Civil and Transportation Engineering, Guangdong University of Technology, Guangzhou 510006, China

ARTICLE INFO

Article history:
Received 2 October 2015
Received in revised form
17 January 2016
Accepted 9 February 2016
Published online 22 March 2016

Key words:
Binocular stereo vision
Wavelet transform
Litchi recognition
Stereo matching

The major constraints for a litchi harvesting robot were to recognise and locate litchi in an unstructured environment with varying illumination and random occlusion. A rapid and reliable method based on binocular stereo vision was developed with the aim of effectively recognising and locating litchi in the natural environment. The method involved the application of wavelet transform to a pair acquired images of litchi to normalise illumination of an object surface. A litchi recognition algorithm based on K-means clustering was presented to separate litchi from leaves, branches and background. A matching algorithm to locate litchi based on a label template was discussed. Litchis with a similar label template were matched according to the preset threshold by traversing a litchi label template of a left image in a right image to find optimal matching. The experimental results showed that the proposed recognition method could be robust against the influences of varying illumination and precisely recognising litchi, the highest average recognition rate for unoccluded and partially occluded litchi was 98.8% and 97.5% respectively. From 100 pairs of tested images of unoccluded and partially occluded litchis 98% and 94% were successfully matched, respectively. Errors had no significant difference and they were less than 15 mm when the measuring distance was between 300 mm and 1600 mm under varying illumination and partially occluded conditions.

1. Introduction

Litchi is a kind of much-favoured fruit. The average annual production of litchi in the world is more than 200 million tonnes, about 80% of which is produced from China. However, with ageing population of China, problems of labour shortages and high labour-costs have been serious and they may even affect the smooth completion of litchi harvesting in the near future. In order to solve this problem, the research and development of a litchi harvesting robot is essential. Many researchers in other fields have developed fruit harvesting robots. Monta, Kondo, and Shibano (1995) developed an agricultural robot using four end-effectors for grape harvesting, berry thinning, spraying and bagging. The experimental

* *Corresponding author*. College of Engineering, South China Agricultural University, Wushan Road, Guangzhou 510642, China. Tel.: +86 020 85280066.
E-mail addresses: wangchenglin055@163.com (C. Wang), xjzou1@163.com (X. Zou).
http://dx.doi.org/10.1016/j.biosystemseng.2016.02.004
1537-5110/

Nomenclature

PC	personal computer
CCD	charge coupled device
DOF	degree of freedom
R, G, B	red, green, and blue colour value
C(x,y)	R colour channel of original litchi colour image
x,y	the coordinate of the pixel point
k,l	independent variable that are both integers
g,h	high-pass filter and low-pass filter
c_j	the original image
R	the radius
d^D_{j+1}, c_{j+1}	diagonal high-frequency and low-frequency
x_0,y_0	coordinates of the centre of the circle
d^V_{j+1},d^H_{j+1}	vertical high-frequency and horizontal high-frequency
$P(x_0,y_0,R)$	an accumulator matrix
$P_{max}(x_0,y_0,R)$	the maximum accumulator matrix
A	the ratio of area
T	threshold for measuring partially occluded region size
b	the value of baseline
O (x_1,y_1,z_1)	world coordinates of litchi feature point
u_0, v_0, a_x, a_v	camera internal parameters
a,b	two portions that litchi C was segmented into
d	the value of $u_1 - u_2$, namely disparity
NCC	the normalised cross-correlation
$M \times N$	the size of the template
$I_1(u + i, v + j)$	the grey value of point $(u + i, v + j)$
Sig.	difference significance test value
$\overline{I_1}(u,v)$	the average value of grey values of label template
$\overline{I_2}(u-d,v)$	the average value of grey values of label template translation
z_1	distance measurement value between the litchi and cameras
(u,v)	the diagonal intersection point coordinates of litchi label template
$(u + i - d, v + j)$	the translation result of $(u + i, v + j)$ in right image

results showed that it could work efficiently. An orange picking robot constructed by Muscato and Prestifilippo (2005) involved a totally autonomous robot for fruit picking and handling crates. The picking time for the robotic fruit picker was at 8.7 per orange. Baeten, Donné, Boedrij, Beckers, and Claesen (2008) proposed an autonomous fruit picking machine (AFPM) for robotic apple harvesting. Its necessary components were assembled. It demonstrated the feasibility and functionality of the AFPM so that the picking cycle period was an average of 9. A litchi harvester designed by Liu, Zeng, and Ke (2011) was able to speed up litchi picking in the long term and picking by litchi harvester was significantly faster than hand picking tall litchi trees. An apple harvesting robot (De-An, Jidong, Wei, Ying, & Yu, 2011) consisted of a manipulator, end-effector and image-based vision servo control system autonomously performed its harvesting task using a vision-based module and picked the apples by controlling a manipulator with 5 DOF structure. The success rate of apple harvesting was 77%, and the average harvesting time was approximately 15 per apple. A strawberry harvesting robot system (Feng, Liu, Wang, Zeng, & Ren, 2012; Feng, Zheng, Qiu, Jiang, & Guo, 2012) based on machine vision, sonar technology and an independent navigation system was found to harvest strawberry from both sides and distinguish and locate fruit by a vision system. A successful harvesting rate of 86% was achieved, every successful harvesting operation on average took 31.3, and the average error for fruit localisation was <4.6 mm. Robotic fruit harvesting was reviewed by Bac and van Henten (2014).

Recognition and localisation of fruit is an important basic procedure for developing a fruit harvesting robot (Mehta & Burks, 2014). In most reported studies, the camera is a key element of the vision system of the robot playing a crucial role in producing images for the robot to recognise and locate fruits. Some studies concerning the ability of the vision system of the robot to recognise fruits have been reported. In Slaughter and Harrell (1987), over 75% oranges were detected through enhancing the contrast between orange fruits and others by a colour camera with a filter of 675 nm wavelength, but some misclassified non-fruit part were presented. Grey level fruit images of peach and apple were obtained by a single CCD camera with filters, and up to 92% fruits were detected by Sites Peter and Delwiche Michael (1988). Zhao, Tow, and Katupitiya (2005) analysed a single colour image to segment apples using both colour and texture, the correct rate was about 90%. Although these methods were able to recognise fruits, they did not have sufficient ability to adapt to the fruit harvesting because the analysis was still based on two-dimensional images which could not determine the distance from the camera to the image scene.

The binocular stereo vision is a more favourable method to carry out the three-dimensional localisation of fruit. It includes two cameras using different views to capture two images, thus the distance between the fruit and the camera can be calculated by triangulation. In reported studies, images have been captured by two parallel cameras or cameras at specified angles, thus the distance from the camera to the image scene could be calculated by combining the left and right images. An early stereo vision system developed for an apple harvesting robot involved the use of two cameras positioned at an angle with 40% of visible apples recognised in the experiment (Kassay, 1992). A similar scheme (Grasso & Recce, 1996) was implemented so that oranges could be recognised by a stereo matching algorithm using two cameras to capture images from different view of angles. Another binocular stereo vision system for robotic apples harvesting was developed using two cameras placed in parallel (Teruo, Shuhuai, & Hiroshi, 2002). The rate of fruit discrimination was about 90% for red fruits and 65%–70% for yellow-green apples. This method was also used in several studies for recognition and localisation of fruits. Xiang, Jiang, and Ying (2014) proposed a binocular vision system for recognising clustered tomatoes. The depth images of clustered tomatoes were acquired through the two colour cameras placed in parallel and the recognition accuracy rate of clustered tomatoes was 87.9% when the leaf or branch occlusion rate was

less than 25%. The same binocular scheme was implemented by Si, Liu, and Feng (2015) and over 89.5% of fruits were successfully recognised from 160 tested images. These above mentioned studies, although designed to meet the needs of robotic fruit harvesting, had two problems that influenced the ability of the fruit harvesting robot to successfully recognise and locate fruits. For example, a vision system of recognising clustered tomatoes (Xiang et al., 2014) presented a low rate of recognising tomatoes occluded by leaves and branches and did not satisfactorily segment tomato images under varying illumination. The vision system developed by Si et al. (2015) had high rate of recognition for apples, but the researchers did not discuss occluded apples. Cameras as the main sensing devices with this kind of binocular vision system are often affected by variations in illumination. Different vision systems often presented a low rate for recognising occluded fruits and were often not able to locate occluded fruits.

The active triangulation method has been presented by some researchers as a method to reduce the influence of varying illumination. Active triangulation can include a laser projector and position sensitive devices. A laser projector has been used to project two laser beams with different wavelengths on the fruits and other objects. Each laser beam had a different spectral-reflection characteristic. When the reflected laser beams were received by position sensitive devices, the beams and fruits were distinguished from the background by the characteristic differences between the two laser beams. Researchers have often used the range sensors to locate fruits in robotic fruit harvesting. Ceres, Pons, Jimenez, Martin, and Calderon (1998) accomplished this using a laser range finder and promising results were obtained. Bulanon, Kataoka, Okamoto, and Hata (2005) designed a cylindrical manipulator for picking apples, in which was installed a colour camera and a laser sensor. The distance between the target fruit centre and the visual servo was measured by the laser range sensor. Although the accuracy of this system in estimating distance to fruit was ±3 mm, it increased complexity during data acquisition and was bulky and slow. A novel three-dimensional vision system developed by Feng, Liu, et al. (2012), Feng, Zheng, et al. (2012) effectively enhanced the overall performance of a robotic fruit harvester, which used a laser range finder to collect distance information around it when it moved horizontally. Range sensors methods for robotic fruit harvesting were reviewed by Gongal, Amatya, Karkee, Zhang, and Lewis (2015). However, range sensors not only are not able to exactly locate all the points in the harvesting scene because they only return a distance of a single point, but also this method cannot provide three-dimensional localisation of fruit when objects are far from the robot because the reflected laser intensity from the distant objects may be weakened (Si et al., 2015).

Litchi harvesting has a real-time requirement. Among the techniques mentioned earlier, binocular vision appears to be the only technology which analyses two images of the recognised fruit and does not have other mechanical time-consuming problems, thus it has great potential for developing a litchi harvesting robot. This study uses binocular stereo vision to develop a vision system for a litchi harvesting robot. The litchi binocular stereo vision localisation system presented in this study was tested in a litchi orchard to demonstrate its ability to recognise and locate litchis in an unstructured environment (i.e. natural light and natural litchi growth). The highlights of this study were: (1) a robust algorithm for recognising litchi was against varying illumination developed; and (2) a label template-based matching algorithm was proposed. The experiment showed that these algorithms enabled our litchi localisation system to successfully recognise and locate unoccluded and partially occluded litchis in their real natural environment.

2. Materials and methods

2.1. System components

As shown in Fig. 1, the proposed binocular stereo vision system consisted of a PC, two CCD colour cameras, a calibration board and localisation software system. The cameras (model MV-VD120SC) produced by Xian Microvision company in China had a digital video output of 1280 by 960 effective pixels, and they were placed in parallel and mounted on the manipulator that had six DOF in order to obtain images flexibly. The distance between centres of the two camera lenses could be adjusted but was fixed at 200 mm in this study. The focal length of the cameras were selected as 6 mm. The localisation software system composed by cameras calibration module was programmed in OpenCV 3.0 (supplied by Intel Corporation at Santa Clara in California, USA), litchi image segmentation module and litchi matching module programmed in Matlab 8.3 (supplied by MathWorks Corporation at Nedik in Massachusetts, USA) ran in the PC with 4 GB RAM, an Intel Core i5-2500 CPU, a Windows 7 operating system.

2.2. Calibration of cameras and image acquisition

Camera calibration is the crucial process in stereo vision, the aim of which is to determine the parameters of the cameras and obtain real-world co-ordinates through establishing the geometry relationship between the coordinates of a point on a spatial surface of an object and its mapping coordinate in the image plane model by experiment and calculation. This was achieved by adjusting the distance between the cameras and the calibration board, and by making sure that each corner of the calibration board was in the field of view of the camera. Multiple images were obtained by changing the calibration board pose relative to the cameras. Thus, the cameras internal parameters and external parameters could be established by calculating the collected corner position set. More comprehensive cameras calibration was described in the previous article by our research group (Zou, Zou, & Lu, 2012).

After calibrating the cameras, litchi images from the field under sunny and cloudy day light conditions were obtained by adjusting the distance between cameras and litchis by controlling the manipulator shown as Fig. 2. The distance between cameras and litchis was in the range of 300 mm–1600 mm. The acquired litchi images were uniformly cropped into 640 × 480 pixels.

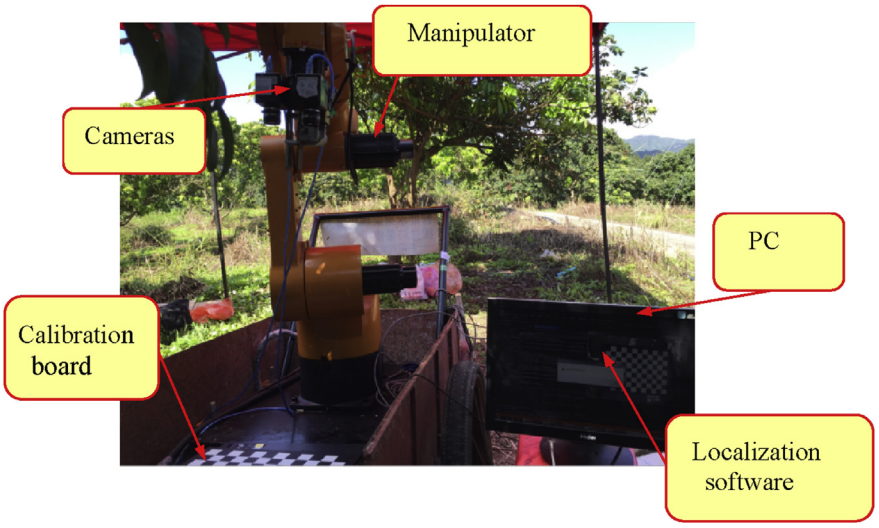

Fig. 1 – The binocular stereo vision system.

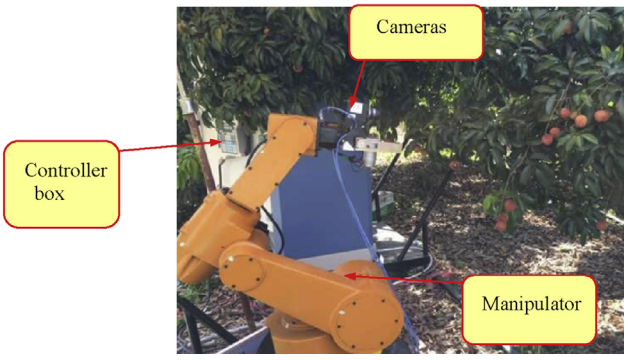

Fig. 2 – System for obtaining litchi images.

2.3. *Litchi recognition algorithm*

Normally, shape based analysis and colour based analysis are used for recognition fruits (Si et al., 2015). A Hough transform algorithm was applied to fruit image segmentation to determine circular fruit objects, such as tomatoes (Dale Whittaker, Miles, Mitchell, &, Gaultney, 1987), citruses (Pla, Juste, & Ferri, 1993) and cherries (Tanigaki, Fujiura, Akase, & Imagawa, 2008). In space, the litchi shape is basically conical. The shape is irregular round and sometimes approximates to a triangle in two-dimensions, therefore it is difficult to apply simple algorithm for recognising a random litchi shape. Therefore, shape-based analysis should not be used for recognising litchi. In the natural environment, the ripe litchi colour is red and that is different from other objects on a litchi tree. However, varying light makes uneven illumination on the surface of objects. Under these conditions, light spots or shadows can easily form on the surface of the litchis and other objects in the litchi tree, even in the background. If the acquired litchi image directly is segmented based on colour analysis, serious errors in segmentation will be caused.

2.3.1. *Wavelet-based illumination normalisation*

In order to reduce the effects of varying illumination, colour indicators have been proposed to segment fruits. In shadows, the indicator 2R–G–B (Yaqin & Hua, 2004) was found not be able to present ideal fruit segmentation. Although the indicator (R–G)/(G–B) (Si et al., 2015) successfully segmented apples in the acquired apple image, holes formed by varying illumination still existed on the surface of segmented apples.

As shown in Fig. 3a, apparent light spot occurs on the surface of litchis, leaves and background. The changes of image caused by varying illumination are mainly due to the change of brightness and contrast, thus the low-frequency components of the spectrum will also change. High-frequency components of the spectrum represents the texture, edge and other detail information, and in litchi image these components reflect the position of litchi, leaves and fruit peduncles, the shape information and the information of colour mutation. However, light mainly affects the low-frequency components of image. Wavelets are a function of short-term volatility with compact support, which is a potential tool for extraction details and other approximate composition of image because it is able to combine space domain and frequency domain to analyse.

Considering R colour channel (Fig. 3b) of original litchi colour image as C(x,y), the two-dimensional discrete wavelet transform (two-dimensional Mallat algorithm) in the first layer was adopted to decompose it into low-frequency component and high-frequency component. Its quick decomposition formula is as below:

$$\begin{cases} d^{V}_{j+1}(x,y) = \sum_{k}\sum_{l} g(k-2x)h(l-2y)c_j(k,l) \\ d^{H}_{j+1}(x,y) = \sum_{k}\sum_{l} g(k-2x)h(l-2y)c_j(k,l) \\ d^{D}_{j+1}(x,y) = \sum_{k}\sum_{l} g(k-2x)h(l-2y)c_j(k,l) \\ c_{j+1}(x,y) = \sum_{k}\sum_{l} g(k-2x)h(l-2y)c_j(k,l) \end{cases} \quad (1)$$

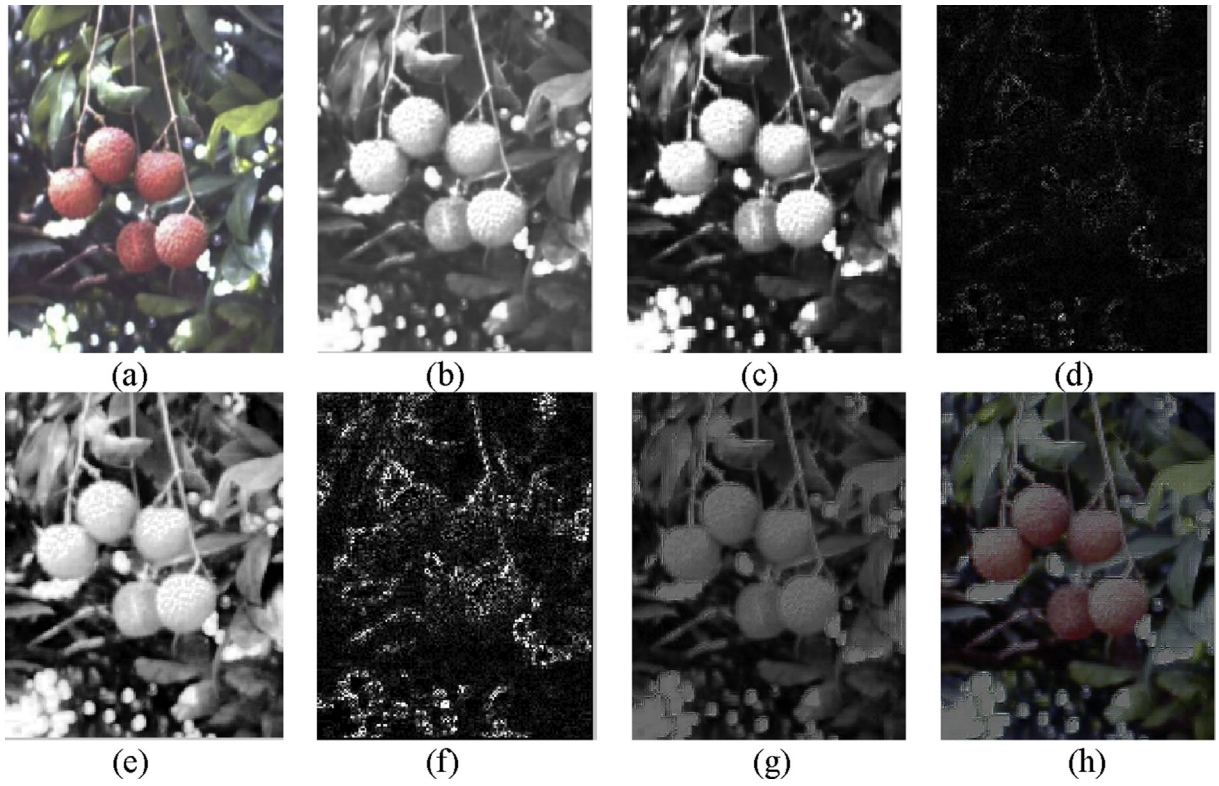

Fig. 3 – Illumination normalisation (a) Original image, (b) R colour channel, (c) low-frequency of R, (d) high-frequency of R, (e) low-frequency histogram equalisation, (f) high-frequency contrast enhancement, (g) wavelet reconstruction of R and (h) original image illumination normalisation.

Where x, y presented the coordinate of the pixel point, k, l were both integers, and g and h were high-pass filter and low-pass filter respectively. c_j was the original image, d^V_{j+1} was vertical high-frequency component, d^H_{j+1} was horizontal high-frequency component and d^D_{j+1} was diagonal high-frequency component. c_{j+1} was low-frequency component. Thus, R colour channel was decomposed into low-frequency component (Fig. 3c) and high-frequency component (Fig. 3d) by Eq. (1). Histogram equalisation and contrast enhancement were then used as the low-frequency component and the high-frequency component respectively (Fig. 3e and f). Illumination normalisation was accomplished when the low-frequency component and the high-frequency component were reorganised into a grey scale image (Fig. 3g) using two-dimensional wavelet reconstruction using Eq. (2). The parameters in the Eq. (2) were the same as in the Eq. (1). Programming was carried out using Matlab. Biorthogonal wavelet was selected as the wavelet function from the Matlab wavelet tool box.

$$\begin{aligned} c_j(x,y) = &\sum_l \sum_k h(k-2x)h(l-2y)c_{j+1}(k,l) \\ &+ \sum_l \sum_k g(k-2x)h(l-2y)d^H_{j+1}(k,l) \\ &+ \sum_l \sum_k h(k-2x)g(l-2y)d^V_{j+1}(k,l) \\ &+ \sum_l \sum_k g(k-2x)g(l-2y)d^D_{j+1}(k,l) \end{aligned} \quad (2)$$

The same method was applied to other two colour channels, G and B channels, of a litchi colour image for illumination normalisation and contrast enhancement. A wavelet-based illumination normalisation was then carried out (Fig. 3h) by reorganisation of the three illumination normalised colour channels of the litchi colour image.

2.3.2. *K-means clustering-based litchi segmentation*

Litchi fruit is red, which is different from branches, leaves and background. Thus, colour analysis has potential for litchi segmentation. Litchi segmentation is simply the procedure that finds the red components and discards other colour components. Clustering is an unsupervised learning that can automatically divide data set into the same class and different class, in which has greater similarity and a large degree of difference respectively. *K*-means clustering is a commonly used clustering algorithm. That determines *K* divisions which reach a minimum squared error. When a cluster is dense the distinction between classes is obvious, the algorithm has better results. For dealing with large data sets, the algorithm is relatively scalable and efficient. *K*-means clustering needs to specify the number of clusters to be partitioned and a distance metric to quantify how near the two classes are to each other. Litchi fruit is red, the leaves are green and the peduncles and branches are other colours, thus the number of clusters should be determined as three. Figure 4a and d presented original litchi colour images under sunny and cloudy days respectively, and their illumination normalised image are shown as Fig. 4b and e. After illumination normalisation, from

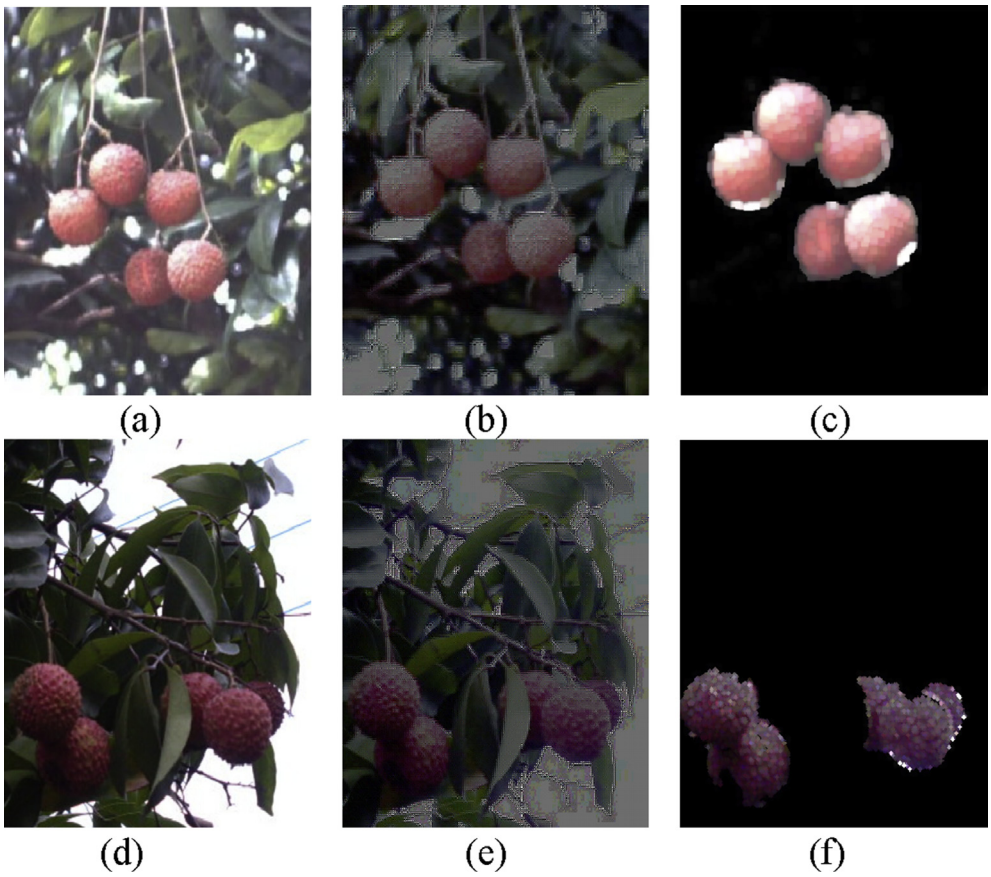

Fig. 4 – Litchi segmentation (a) and (d) original image under sunny and cloudy days, (b) and (e) illumination normalisation images, (c) and (f) segmented litchis.

Fig. 4c and f, it can be seen that the fruits of unoccluded and partially occluded litchis were segmented by *K*-means clustering.

2.3.3. *Implementation of litchi recognition algorithm*

In order to evaluate the effect of the proposed litchi recognition algorithm, 300 images under outdoor conditions captured in a litchi orchard in Guangzhou were tested. The image acquisition conditions were as follows: distance of the image acquisition system, 300 mm–1600 mm; dates, June 20, 2015 to June 25, 2015; time, 7:00 AM to 5:00 PM; weather, cloudy and sunny; and lighting conditions, front and back. Images were captured by our vision system every 2 h, 100 images in sunny front lighting conditions, 100 images in sunny back lighting conditions and 100 images under cloudy conditions. The images were processed and both unoccluded and partially occluded litchi fruits were segmented using the proposed recognition algorithm. Meanwhile, the images were also segmented using image editing software Photoshop 13.0 (supplied by Adobe Corporation at San Jose in California, USA) by manually labelling litchi for artificial criteria. The remaining litchi portion in the difference image between the manual segmentation image and the image segmented by algorithm was labelled manually, and the pixels were calculated. If the pixel rate of the remaining portion to that manually labelled was less than 0.05, then the litchi segmented by the proposed recognition algorithm could be considered as successful and its segmentation rate could be calculated. The litchis in our experimental orchard all belonged to the same cultivar (Guiwei litchi), thus each litchi had almost the same size and obscured around 16,284 pixels in the image. Therefore, in the difference image, the pixel count of the remaining portion of litchi was counted. It was found that the pixel count of the remaining portion of successfully segmented litchi was about 100 pixels, which could be ignored and it was also the minimum size that could be detected.

2.4. *Matching method for segmented litchi*

After segmenting the litchi, the same litchis from the left and right images needed to be matched for localisation.

2.4.1. *Edge detection-based litchi label extraction*

In some reported studies, fruit contour was detected by circle fitting using circular Hough transform, and fruits were matched by matching the feature points (circular centre) of the detected fruits. The basic idea of a circular Hough transform (Mukhopadhyayab & Chaudhuria, 2015) is to map the edge points of the image space into the parameter space. The size and centre of a circle were judged by calculating the cumulative value of all the coordinate points in the parameter space. The circle function in Cartesian coordinate system is shown as Eq. (3):

$$(x - x_0)^2 + (y - y_0)^2 = R^2 \quad (3)$$

where x_0, y_0 are the horizontal and vertical coordinates of the centre of the circle respectively, and R is the radius. The three parameters that need to be identified form an accumulator matrix $P(x_0,y_0,R)$. By increasing x_0, y_0, the solution R of the Eq. (4) is solved. The corresponding accumulator matrix $P(x_0,y_0,R)$ is updated, and at last the coordinates of the maximum $P_{max}(x_0,y_0,R)$ in the accumulator matrix $P(x_0,y_0,R)$ is just the coordinates of (x_0,y_0).

$$R = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (4)$$

Dale Whittaker et al. (1987) applied the circular Hough transform to the location of tomatoes. Results showed that the circular Hough transform was an acceptable method for locating and identifying tomatoes in an image of natural field condition. However, the algorithms were computationally intensive on a serial processor and were not able to be performed in real time. Recognition of partially occluded tomatoes did not present a promising result using the circular Hough transform. Plebe and Giorgio (2001) pointed out the drawbacks of the circular Hough transform for detecting oranges and that it was time consuming. It was found empirically that the Hough transform did not work well with clusters of oranges, or when a substantial part of an orange's surface was partially occluded. Figure 5b shows a circular Hough transform result from two overlapping litchis where surface of litchi A was partially occluded by litchi B (Fig. 5a). The centre and the radius of the circular Hough transform was (42, 96) and 21.783 pixels, respectively. However, the arc portion detected by the circular Hough transform may not be real portion occluded by litchi B. Therefore, the accuracy of centre and radius of circular Hough transform of litchi A cannot be proved, and the matching error would increase by matching this centre and radius.

In this study, an edge detection algorithm was used to determine litchi contour. As can be seen in Fig. 5c, edge detection was based on a Canny operator of Fig. 5a. By adjusting threshold to 0.3, the internal noise points of litchi contour were removed. However, under this condition, there were many points with the same value of X or Y coordinates at the edge, which caused some parts of the edge to be in the form of a line. Figure 5d was edge thinning result of Fig. 5a based on the interpolation of grey scale. In this binary image, the coordinates (x_0,y_0) of all points of '1' were stored. Two straight lines existing in the image which were parallel to the Y axis were made as the tangent lines of litchi contours at the maximum and minimum values of x_0, similarly, two straight lines parallel to the X axis were also found at the maximum and minimum values of y_0. The four straight lines were moved and intersected with the two contours, and if there were only three points and the distance between the three points and one line of four lines was '0', then the lines were determined as the lines of the tangent to the middle contour and the algorithm stopped. Thus, the rectangle tangent to litchi contours was determined and defined using the litchi label. If some labels had four points of tangency, the litchis

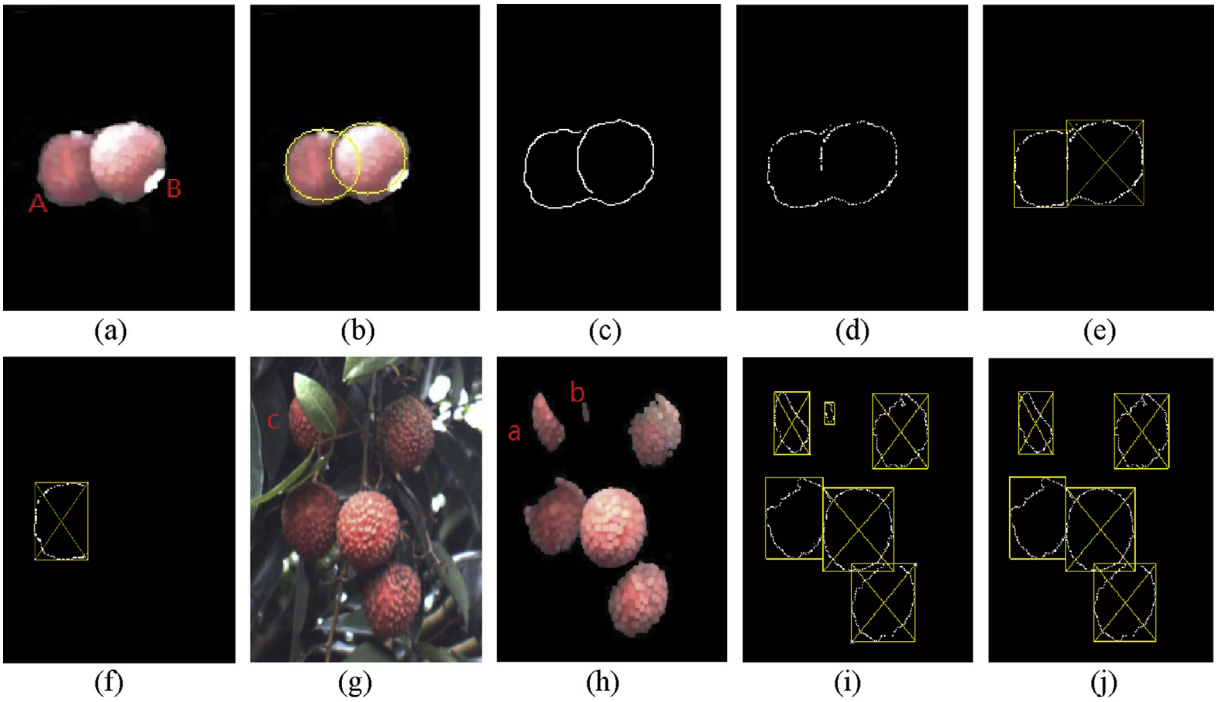

Fig. 5 – Image of litchi label and feature point extraction (a) two overlapping litchis fruits, (b) circular Hough transform result, (c) Canny edge detection result, (d) edge thinning result, (e) label and feature point extraction of former litchi, (f) Label and feature point extraction of partially occluded litchi, (g) partially occluded litchi image by leaves, (h) litchi segmentation image, (i) portion size determination of partially occluded litchi, (j) label and feature point extraction of partially occluded litchi.

represented by these labels would be determined as the litchis closer to the cameras. These litchis would be firstly matched by matching their feature points which were diagonal intersection points of the labels (Fig. 5e). After matching this kind of litchi, the whole region of the label was set to '0'. Next, the label which had three points of tangency was considered as partially occluded contour and would be matched (Fig. 5f). By a large number of simulation experiments, the ratio A of area of the partially occluded litchi label to the image area was selected as threshold T for measuring partially occluded region size. If T was less than 0.01, the partially occluded region was abandoned and not matched. One extreme position of two overlapping litchis was that the two litchis were separated and each contour had four points of tangency. The other extreme was that one litchi image was inside of the other or completely covered the other. Obviously, the above algorithm was suitable for all positions of two overlapping litchis. However, for the position of complete covering, only the label of the former litchi can be extracted for matching, the contour covered by former litchi contour is invisible so not able to be determined. For three litchis, or more litchis, their position relationship can be considered as the arrangement between one litchi and two litchis or more litchis. So their labels and feature points can also be extracted by the above algorithm.

As shown in Fig. 5g, litchi C was partially occluded by leaves and it was segmented into two portions, a and b as shown in Fig. 5h. The ratio of area of label of portion b to the image area was 0.003 and less than 0.01 (Fig. 5i). As a result, portion b was abandoned. The result of label and feature point extraction of Fig. 5h is shown in Fig. 5j. If a litchi is segmented into two portions and each rate of area of label of portion contour to the image area is larger than 0.01, then the label and feature point of each portion will be extracted for matching.

2.4.2. *Label template-based litchi matching*

For locating litchi, the coordinates O (x_1,y_1,z_1) of a litchi feature point should be determined. Equation (5) is the conversion formula between the world coordinate system and image coordinate system (Zou et al., 2012).

$$\begin{cases} x_1 = \dfrac{b(u_1-u_0)}{(u_1-u_2)} \\ y_1 = \dfrac{ba_x(v_1-v_0)}{a_y(u_1-u_2)} \\ z_1 = \dfrac{ba_x}{(u_1-u_2)} \end{cases} \tag{5}$$

where u_0, v_0, a_x, a_y are the camera internal parameters and they were obtained by cameras calibration, b is the value of baseline, 200 mm. Thus, if $u_1 - u_2$ is determined, O (x_1,y_1,z_1) can be calculated. The value of $u_1 - u_2$ is defined as disparity d and determined by the matching procedure. Matching based on binocular stereo vision is the process that searches for the corresponding point of left image in the same image scene in the right image. This search process can be seen as a template matching process. As discussed in the previous section, litchi feature point which is the diagonal intersection point of litchi label has been extracted. Therefore, the litchi label of left image can be considered as the template and it is traversed into right image along the epipolar line to find the most similar window, the diagonal intersection point of the found window is the matching point of litchi feature point of the left image. The similarity measure method was selected the normalised cross-correlation (NCC) based on the grey value matching shown as Eq. (6), which was invariant to all linear illumination changes, and hence was suitable for natural environments (Steger, Ulrich, & Wiedemann, 2008

$$NCC(d) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_1(u+i,v+j)-\bar{I}_1(u,v)\right]\left[I_2(u+i-d,v+j)-\bar{I}_2(u-d,v)\right]}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_1(u+i,v+j)-\bar{I}_1(u,v)\right]^2}\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I_2(u+i-d,v+j)-\bar{I}_2(u-d,v)\right]^2}} \tag{6}$$

Here, (u,v) was the diagonal intersection point coordinates of litchi label template in left image, the size of the template was $M \times N$. $I_1(u+i,v+j)$ was the grey value of point $(u+i,v+j)$, and $\overline{I_1}(u,v)$ was the average value of grey values of label template. The coordinates $(u+i-d,v+j)$ was the translation result of the coordinates $(u+i,v+j)$ in right image, and similarly, $I_2(u+i-d,v+j)$ was the grey value of point $(u+i-d,v+j)$. $\overline{I_2}(u-d,v)$ was the average value of grey values of a $M \times N$ size window that its diagonal intersection point coordinates was $(u-d,v)$. By solving Eq. (6), obtain the solution d which makes $NCC(d)$ reach the maximum value. Thus, the disparity in d was calculated, the matching point coordinates of (u,v) was obtained and the coordinates O (x_1,y_1,z_1) of litchi feature point were determined by Eq. (5).

2.4.3. *Implementation of litchi matching algorithm*

It is difficult for irregular graphs to select the feature points of matching. A frame is the smallest enclosing rectangle of litchi contour, and the sides of the rectangle are parallel to the x and y axes of the image. Therefore, the frame as the label of litchi is a regular graph, which can be easy to match. The label of the single litchi in the left image was considered as a template, which was to be traversed in the right image for searching the matching object. When the match of two labels reached a preset threshold, the match ended. In this paper, when 0.6 was chosen as the threshold, matching achieved a satisfactory result. In order to obtain the correct matching, some matching constraints need to be noted. By the ordering constraint, if a litchi is on the left side of another in the left image, it should also be on the left side of it in right image. If there were not litchi fruits on the left side of the litchi in right image, the corresponding position of the litchi could not be determined. Under this condition, this kind of litchi was ignored. If some portion of a litchi needed to be matched was in left image, and the whole litchi or some part of the litchi was in right image,

the preset threshold constrainted the correct matching. The epipolar constraint stated that for the mapping point on an image, its match point must fall on another image of epipolar online. In binocular stereo vision system, this epipolar constraint may hypothesis that the matching points were in the same horizontal line, that was Y values of coordinates of the matching points in the two images were equal. So, matching in the different rows should be discarded. Furthermore, the length between the centres of the two camera lens, defined as baseline, should be adjusted to a suitable distance. By triangulation, if the baseline is shorter, both the position of litchi in left and right image will be more consistent, which makes the matching of features be easy. The shorter baseline will also make distance measurement errors become larger. In this study the baseline was fixed at 200 mm.

Therefore, litchi matching result was shown as Fig. 6 through the discussion of the above sections. And 100 pairs of images were used in the matching experiment. A total of around 600 pairs of litchis in both the left and right images were tested in the matching experiment, in which there were around 200 pairs of partially occluded litchis and 400 pairs of unoccluded litchis, respectively. The proposed algorithm and circular Hough transform were applied to these images for matching. The success rate of the matching was calculated.

2.5. *Distance measurement*

After recognition and matching of litchi were carried out, distance measurements were carried out under various lighting conditions. Distance measurement was based on triangulation using Eq. (5) where z_1 was the distance measurement between the litchi and cameras. A laser range finder, MileSeey S9 (supplied by Step Test Technology Co., Ltd. in Shenzhen, China) whose measurement error was ±1.5 mm within the range of 60,000 mm, was taken as measuring the true distance. The experiment was carried out when the distance between litchi and camera was in the range of 300 mm–1600 mm. For each condition, fifteen litchis samples in the field of vision were selected for the distance measurement experiment. At every 100 mm, the distance between each litchi and the cameras was measured using the algorithm and the measurement error with respect to the true (laser range finder) value was recorded.

3. Results and discussions

3.1. *Performance of litchi segmentation algorithm*

Figure 7 shows the litchi segmentation result of a representative image (Fig. 7a) composed of leaves, branches, soil, light spot, which is compared with the results of the algorithm (R–G)/(G–B) and *K*-means clustering. It can be seen from Fig. 7b that the litchi fruits were completely segmented from the leaves, branches and background. Figure 7c and d showed that the segmented litchis were not complete and there many holes in the surface of the segmented litchi fruits due to the uneven lighting. Figure 7c and d also reveal that the fruits located in shadowed areas and partially occluded by other litchis were not successfully recognised. The result showed the proposed algorithm was more suitable for segmenting litchi in the real natural environment than the algorithm (R–G)/(G–B) and *K*-means clustering.

The segmentation data compared with artificial criteria was recorded to illustrate the performance of the recognition algorithm proposed in segmenting the litchi fruits under different lighting and partially occluded conditions. As shown in Table 1, the proposed image segmentation method yielded very satisfactory results. For unoccluded litchis, the average fruit recognition rate (hit rate) under front light conditions was the highest (98.8%), and declined to 94.6% under back light conditions. However, the rate was the lowest (only 90.5%), occurred under cloudy conditions. Wavelet-based illumination normalisation algorithm made the light uniformly distributed on the surface of the objects. Light could be compensated under front light and back light conditions better differentiating litchi and other objects. However, for cloudy days, the colour of the background (i.e. soil) was similar to the litchi colour. It was difficult to distinguish the litchi and the soil based on colour analysis. The rate of background mistakenly recognised as fruit (i.e. false positive rate) under this condition was (4.9%) much higher than under the other two conditions, front light (0.9%) and back light (1.1%). The rate of fruit mistakenly recognised as background (false negative rate) was much lower than the false positive rate and under every condition was 0.8%, 1.2% and 1.5% respectively because the litchi, as the main part of the image, could easily

Fig. 6 – Litchi matching based on the proposed algorithm (a) litchi left image, (b) litchi right image.

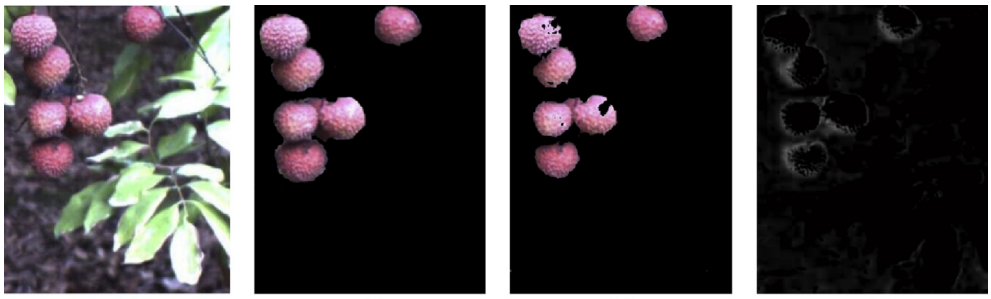

Fig. 7 – Comparison of segmentation results (a) original image, (b) proposed algorithm, (c) K-means clustering and (d) (R–G)/(G–B) algorithm.

Table 1 – Performance of fruit segmentation algorithm.

		Hit rate (%)	False positive rate (%)	False negative rate (%)
Unoccluded	Front light	98.8	0.9	0.8
	Black light	94.6	1.1	1.2
	Cloudy day	90.5	4.9	1.5
Occluded	Front light	97.5	0.8	2.1
	Black light	95.5	0.8	2.3
	Cloudy day	89.4	1.1	5.6

be recognised by the colour analysis based algorithm. For the partially occluded litchis, the fruit hit rate achieved on 97.5%, 95.5% and 89.4% under front light conditions, back light conditions and cloudy day conditions respectively. Except for the effect of light, due to the occlusion by leaves, branches and other litchis, litchi was divided into many portions randomly. As long as any portion of litchi was recognised, partially occluded litchi recognition was just completed. As mentioned in Section 2.3.3, 100 pixels is the minimum size to be detected, calculation in turn, the minimum size was 6% of Guiwei litchi total pixels, so litchi of occluded rate less than 94% could be recognised by the algorithm. However, a partially occluded litchi including one area or more areas which are composed of few red pixels was too little (<100 pixels) to be detected and was not recognised. This resulted in higher levels of false negatives under cloudy days conditions (5.6%), and much higher than under the other two conditions, 2.3% and 2.1%. While false positive rate for partially occluded litchi were very low, it only achieved 0.8%, 0.8% and 1.1% because that background area was large enough not to be recognised as the portion of partially occluded litchi. Furthermore, it was found according to statistics that most of false negatives and false positives appeared between 11:00 AM to 1:00 PM. The light was very strong during this time in summer of 2015 in Guangdong, China. Light spots distributed on the surfaces of litchis, leaves and soil, which made white bright hole on surface of these objects. Under front light conditions, the stronger the light was, the brighter white bight hole was. Under back light conditions, the stronger the light was, the more similar the colour between litchi and background was. This led to the inability of the proposed segmentation algorithm to carry out this function. On other occasions, light intensity did not reach the degree that colour of the litchi and the background could not be differentiated. Using the proposed algorithm, satisfactory segmentation results could be obtained and the proposed image segmentation algorithm was effective.

3.2. *Performance of litchi matching algorithm*

The litchi matching result was based on circular Hough transform and it showed mistaken matching as seen from Fig. 8. Litchi A was wrongly matched with litchi B in right image. The reason was that litchi A was partially occluded by other litchi in left image that had similar radii and coordinates

Fig. 8 – Litchi matching based on circular Hough transform (a) litchi left image, (b) litchi right image.

Table 2 – Performance of distance measurement.

Conditions	Litchis	Distance measurement error with respect to the true value (mm)													
		300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600
Cloudy and unoccluded	Sample1	−1.49	5.68	9.84	0.47	−10.77	1.30	−9.25	10.89	3.22	0.32	13.25	−10.01	−1.80	11.02
	Sample2	−1.54	5.23	10.60	0.55	−10.80	1.05	−10.05	10.21	4.18	0.46	13.89	−9.77	−1.62	9.22
	Sample3	−1.61	5.77	10.66	0.53	−11.00	1.39	−8.88	9.82	2.88	0.33	14.02	−9.77	−1.53	10.00
	Sample4	−1.49	5.82	10.50	0.43	−10.74	1.00	−9.17	9.33	3.12	0.47	12.56	−9.53	−1.92	10.12
	Sample5	−1.51	5.18	10.89	0.55	−10.54	1.21	−10.32	11.29	4.56	0.31	12.99	−10.12	−2.07	9.34
	Sample6	−1.46	5.32	10.21	0.57	−10.54	1.18	−9.68	11.18	3.44	0.51	12.89	−9.08	−1.81	10.24
	Sample7	−1.55	5.44	9.93	0.45	−11.00	1.12	−9.99	9.67	4.13	0.49	13.78	−9.42	−1.80	10.20
	Sample8	−1.53	5.56	10.97	0.41	−10.88	0.95	−11.01	10.62	3.32	0.33	13.33	−10.46	−1.79	10.12
	Sample9	−1.60	5.89	9.52	0.50	−10.77	1.05	−10.56	10.38	4.01	0.37	13.28	−9.67	−1.90	11.12
	Sample10	−1.58	5.81	10.48	0.52	−10.66	1.20	−8.47	9.26	3.04	0.54	12.98	−9.77	−1.86	10.10
	Sample11	−1.64	5.59	11.08	0.46	−11.04	1.28	−10.97	10.85	4.21	0.43	14.93	−9.87	−1.70	10.12
	Sample12	−1.44	5.41	10.11	0.59	−10.91	0.93	−9.11	11.22	3.78	0.39	14.02	−9.56	−1.80	9.12
	Sample13	−1.56	5.50	9.81	0.51	−10.63	1.07	−9.89	9.78	3.01	0.40	13.88	−9.77	−1.80	10.13
	Sample14	−1.48	5.82	10.89	0.55	−10.50	1.26	−10.23	10.98	3.51	0.49	13.03	−9.96	−1.74	10.12
	Sample15	−1.62	5.23	11.11	0.41	−10.77	0.96	−10.77	12.02	2.99	0.56	14.44	−9.77	−1.80	10.12
Cloudy and occluded	Sample1	10.48	−4.00	−10.80	−11.80	10.60	−14.22	6.00	3.45	9.84	−12.00	−9.34	10.48	13.56	15.00
	Sample2	11.13	−3.30	−11.20	−8.00	10.52	−14.31	4.60	3.53	10.60	−12.04	−8.00	10.86	14.44	14.20
	Sample3	11.00	−4.40	−10.20	−10.80	10.96	−14.56	5.20	2.97	10.66	−12.01	−8.24	10.10	14.45	14.10
	Sample4	10.42	−2.20	−11.40	−9.00	10.16	−14.10	5.40	4.01	10.50	−13.04	−8.14	10.74	16.01	13.20
	Sample5	9.90	−3.40	−11.00	−11.70	11.56	−13.66	6.10	3.98	10.89	−12.03	−9.10	10.22	16.03	16.00
	Sample6	10.50	−3.20	−9.40	−8.10	9.56	−15.00	4.50	3.00	10.21	−11.00	−9.20	10.75	14.21	14.25
	Sample7	10.45	−2.60	−12.00	−11.60	11.00	−13.44	6.20	3.96	9.93	−13.01	−9.30	10.21	13.79	14.75
	Sample8	10.54	−3.00	−9.60	−8.20	10.12	−15.22	4.60	3.02	10.97	−11.03	−8.66	10.23	14.51	14.42
	Sample9	10.36	−3.60	−11.80	−11.40	10.44	−13.22	6.30	3.52	9.52	−13.05	−8.68	9.46	13.49	14.58
	Sample10	9.45	−4.00	−10.70	−8.40	10.68	−15.44	4.30	3.46	10.48	−10.99	−8.64	11.50	14.52	14.32
	Sample11	10.46	−4.50	−10.40	−8.10	10.55	−12.66	4.20	3.10	11.08	−13.07	−8.70	9.36	13.48	14.60
	Sample12	11.45	−3.10	−10.00	−11.70	10.57	−16.00	6.40	3.88	10.11	−10.97	−8.62	10.48	14.00	12.81
	Sample13	10.40	−3.50	−10.70	−11.90	10.58	−14.33	5.30	3.20	9.81	−13.09	−8.72	11.60	13.55	16.19
	Sample14	10.44	−2.60	−10.50	−7.90	10.54	−14.44	5.50	3.78	10.89	−10.95	−8.67	10.48	11.99	15.10
	Sample15	9.77	−3.10	−10.60	−9.90	10.56	−14.35	5.10	3.49	11.11	−12.02	−8.04	10.73	11.97	14.68
Sunny and unoccluded	Sample1	5.11	13.72	9.32	12.35	1.05	9.36	−9.89	−3.04	−1.56	9.50	9.45	8.95	−9.43	−1.52
	Sample2	4.01	13.68	10.02	12.33	0.75	9.34	−9.86	−4.00	−1.54	9.10	9.37	9.34	−9.30	−1.38
	Sample3	5.09	13.73	9.33	12.36	1.03	9.38	−9.91	−3.52	−1.57	9.30	9.80	9.98	−9.80	−1.68
	Sample4	4.03	13.67	10.01	12.37	0.77	9.40	−9.85	−3.44	−1.59	9.60	9.45	9.54	−9.56	−1.40
	Sample5	5.21	13.75	9.34	12.31	1.21	9.32	−9.93	−3.60	−2.00	10.80	9.80	9.30	−9.06	−1.39
	Sample6	3.91	13.65	10.00	12.32	0.59	9.30	−10.00	−3.50	−1.55	10.70	9.45	9.65	−9.15	−1.40
	Sample7	4.57	13.80	11.33	11.34	1.30	9.37	−10.10	−3.51	−1.56	9.40	9.10	8.62	−9.41	−1.12
	Sample8	4.55	13.60	8.01	11.24	0.50	9.33	−9.88	−4.03	−1.20	11.20	9.53	9.30	−9.43	−1.40
	Sample9	4.58	13.90	11.34	13.34	1.10	9.31	−10.20	−3.03	−1.10	8.20	9.45	9.26	−9.45	−1.63
	Sample10	4.54	13.50	8.00	14.68	0.70	10.00	−9.66	−4.01	−1.51	8.00	9.45	9.32	−9.61	−1.61
	Sample11	4.59	12.10	11.54	10.00	1.11	8.70	−9.87	−4.02	−1.58	9.80	9.10	9.06	−9.71	−1.19
	Sample12	4.53	15.30	7.80	14.32	0.69	9.35	−9.76	−3.00	−1.53	9.00	9.45	9.30	−9.34	−1.40
	Sample13	4.60	13.40	9.66	10.36	1.12	9.60	−9.56	−3.02	−1.54	9.70	9.45	9.30	−9.25	−1.17
	Sample14	4.52	14.30	9.68	2.34	0.68	9.10	−9.83	−3.54	−1.90	9.20	9.28	9.30	−9.43	−1.51
	Sample15	4.56	13.70	9.67	13.44	0.90	9.39	−9.90	−3.53	−1.51	8.10	9.62	9.28	−9.52	−1.40
Sunny and occluded	Sample1	11.25	−9.08	−9.30	−3.90	−10.32	−10.11	10.38	−9.26	−12.33	−9.51	−11.98	10.57	6.78	−11.23
	Sample2	11.09	−9.98	−9.45	−4.15	−9.83	−9.89	10.99	−9.97	−11.98	−8.89	−10.00	10.56	6.80	−11.25
	Sample3	11.23	−9.11	−10.42	−4.62	−8.88	−9.81	11.32	−9.86	−13.06	−8.92	−10.01	10.50	6.77	−11.30
	Sample4	11.90	−9.09	−8.89	−3.88	−8.76	−9.11	10.06	−9.61	−12.67	−10.33	−11.97	10.40	6.76	−11.17
	Sample5	11.41	−9.14	−10.11	−3.96	−9.44	−10.08	10.14	−10.32	−12.98	−10.56	−10.99	10.41	6.82	−11.31
	Sample6	11.27	−9.13	−8.92	−4.19	−9.52	−9.48	12.02	−10.59	−13.22	−8.99	−10.06	10.42	6.73	−11.24
	Sample7	10.60	−9.11	−10.30	−4.52	−9.69	−8.52	10.63	−9.69	−12.87	−9.12	−11.92	10.43	6.79	−11.18
	Sample8	11.25	−9.11	−10.21	−4.21	−8.67	−9.97	11.76	−8.87	−12.52	−9.24	−10.05	10.73	6.86	−11.19
	Sample9	11.34	−8.24	−9.82	−4.18	−9.27	−8.84	9.67	−8.97	−12.93	−9.68	−11.93	10.58	6.83	−11.28
	Sample10	11.25	−9.87	−10.99	−4.96	−10.12	−9.60	9.98	−9.22	−13.67	−9.70	−11.94	10.74	6.85	−11.20
	Sample11	11.16	−9.11	−12.00	−5.02	−8.39	−9.66	10.32	−9.63	−13.14	−8.89	−11.96	10.75	6.84	−11.22
	Sample12	11.46	−8.88	−9.33	−3.46	−10.22	−9.50	10.84	−9.66	11.86	−10.02	−10.04	10.76	6.74	−11.26
	Sample13	11.25	−9.11	−8.96	−4.20	−9.69	−9.89	12.11	−9.31	−11.68	−9.43	−10.05	10.66	6.81	−11.27
	Sample14	11.04	−8.35	−10.11	−4.11	−9.67	−9.21	10.48	−10.59	−12.45	−9.56	−10.02	10.60	6.74	−11.21
	Sample15	11.25	−9.34	−9.69	−3.63	−10.03	−8.93	11.00	−9.50	−12.54	−10.41	−10.02	10.59	6.72	−11.29

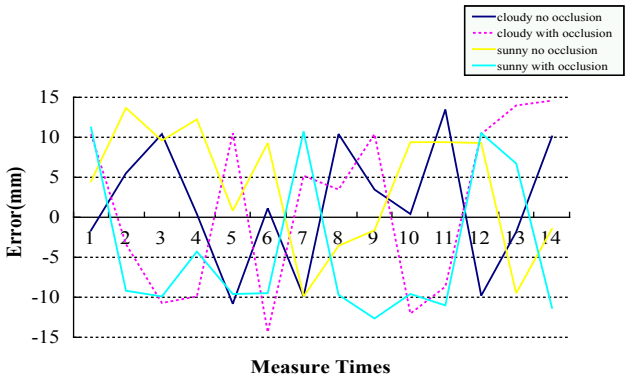

Fig. 9 – Distance measurement average error result.

of the centre with litchi B in right image after circle fitting was carried out on the same row. It was seen that the circle produced after circle fitting was different to the actual litchi contour causing a matching error. After the litchi matching experiment, the statistics showed the matching success rate was very low using the circular Hough transform. Only 81.5% and 74.2% of unccoluded litchis and partially occluded litchis were correctly matched, respectively.

By comparison, occluded litchis matched using the actual contour by applying our proposed method produced 98% correctly matched. For the partially occluded litchis, correct matching rate was 94%. The main reason of the incorrect matching for the unoccluded litchis was that the two label templates of the same litchi in both left and right images were different and one of them was similar to the label template of another litchi close to it on the same row. This is also the reason for the mistaken matching of partially occluded litchis, meanwhile, the label template, which was tangent to all portion of the partially occluded litchi, may be similar to the label templates of the other litchis on the same row. These results indicated that the proposed label template-based matching algorithm was more suitable for litchi matching than the Hough transformation as shown in Fig. 6.

3.3. *Error analysis of distance measurement*

Distance measurement error from the 15 litchis samples per 100 mm with respect to the true value under four conditions were recorded in Table 2. As can be seen, the maximum error and the minimum error were 16.19 mm and 0.31 mm, respectively. Average value of distance measurement errors at each measurement point all fell within ±15 mm as shown in Fig. 9. These four sets of data were divided into six pairs of data, and analysed by SPSS 19.0 (supplied by SPSS Corporation in Chicago, USA) for T-test. The corresponding values of Sig., 0.978, 0.416, 0.204, 0.621, 0.080 and 0.056 were all larger than 0.05, and there was no significant difference for all conditions. Using the distance measurement Eq. (5), the value of $u_1 - u_2$ defined as disparity d was determined by the matching procedure. Once litchis in both left and right images were correctly matched, this disparity value would be considered to be correct. In the actual operation, accurate determination of the position relationship between the two cameras is very difficult. Assuming that the coordinate system of the left camera is completely accurate, the coordinate system of the right camera inevitably has a certain deviation in the position and direction with respect to the left camera, which is bound to lead to error in the determination of the camera internal parameter $a_x$by cameras calibration. These contributed to generation of distance measurement error. The above result indicated that the proposed algorithm was sufficiently robust under variable illumination and partially occluded conditions.

3.4. *Real time performance of the proposed algorithm*

The time consumed by the proposed algorithm mainly occurred during litchi image segmentation and label template-based matching. Wavelet-based illumination normalisation required three colour channels of a litchi colour image and each colour channel to be decomposed into four different frequency images. Although there are already packaged wavelet functions in Matlab, the average time of processing achieved was 1186 ms. Moreover, traversing the litchi label template of left image into right image to search for the optimal matching window based on the normalised cross-correlation similarity measure function cost much time; the average processing time for a pair of images was 1892 ms. However, the propose label extraction method operating on the binary image was fast. The average time consumed image processing from litchi segmentation to litchi location was 3213 ms, which should be sufficient to meet the needs of litchi robotic picking controlled in real-time.

4. Conclusions

Binocular stereo vision-based system was developed for locating litchi in its natural environment. Two algorithms were proposed to recognise and match litchis. According to the results, some conclusions can be made: (1) The proposed litchi recognition algorithm was very satisfactory under sunny front light conditions from 7:00 AM to 5:00 PM except during the period 11:00 AM to 1:00 PM. Even on cloudy days, the average hit rates for unoccluded and occluded litchis were 90.5% and 89.4%, respectively. Litchi of occluded rate less than 94% could be recognised by the algorithm; (2) The proposed label template-based matching algorithm extracted matching feature points and labelled based on litchi contour. Matching results were superior to circle Hough transformation and partially occluded litchi could be matched precisely. 98% and 94% of unoccluded and partially occluded litchis were successfully matched from 600 pairs of litchis respectively; (3) Errors inevitably exist due to the positioning relationship of two cameras, however, average errors of all four conditions were <±15 mm when the measurement was implemented in the range of 300 mm–1600 mm. This showed that the proposed algorithm was robust to variable illumination and partially occluded conditions. Future research will be into an adaptive algorithm for segmentation of litchi and other fruit under variable illumination and improving the localisation accuracy of litchi via hardware and software improvements. The proposed litchi matching algorithm could be applied to other fruits as it has simple direct advantages.

Acknowledgements

This project was supported by a grant from the National Natural Science Foundation of China (No. 31571568, 51578162).

REFERENCES

Bac, C. W., & van Henten, E. J. (2014). Harvesting robots for high-value crops: state-of-the-art review and challenges ahead. *Journal of Field Robotics, 31*(6), 888–911.

Baeten, J., Donné, K., Boedrij, S., Beckers, W., & Claesen, E. (2008). Autonomous fruit picking machine: a robotic apple harvester. *Field and Service Robotics, 42*(3), 531–539.

Bulanon, D., Kataoka, T., Okamoto, H., & Hata, S. (2005). *Feedback control of manipulator using machine vision for robotic apple harvesting*. ASAE Paper Number: 053114.

Ceres, R., Pons, J. L., Jimenez, A. R., Martin, J. M., & Calderon, L. (1998). Agribot: a robot for aided fruit harvesting. *Industrial Robot, 25*(5), 337–346.

Dale Whittaker, A., Miles, G. E., Mitchell, O. R., & Gaultney, L. D. (1987). Fruit location in a partially occluded image. *American Society of Agricultural Engineers, 30*(3), 591–597.

De-An, Z., Jidong, L., Wei, J., Ying, Z., & Yu, C. (2011). Design and control of an apple harvesting robot. *Biosystems Engineering, 110*(2), 112–122.

Feng, J., Liu, G., Wang, S., Zeng, L., & Ren, W. (2012). A novel 3D laser vision system for robotic apple harvesting. In *ASABE Annual International Meeting Presentation*. Paper Number: 12–1341025.

Feng, Q., Zheng, W., Qiu, Q., Jiang, K., & Guo, R. (2012). Study on strawberry robotic harvesting system. In *Paper presented at the Computer Science and Automation Engineering (CSAE)* (Vol. 1, pp. 320–324).

Gongal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: a review. *Computers and Electronics in Agriculture, 116*, 8–19.

Grasso, G., & Recce, M. (1996). Scene analysis for an Orange Picking Robot. In *International Congress for Computer Technology in Agriculture*. Wageningen, Netherlands: VIAS.

Kassay, L. (1992). *Hungarian robotic apple harvester* (pp. 1–14). American Society of Agricultural Engineers 92.

Liu, T. H., Zeng, X. R., & Ke, Z. H. (2011). Design and prototyping a harvester for litchi picking. In *Paper presented at the 4th International Conference on Intelligent Computation Technology and Automation, ICICTA 2011, Shenzhen, China* (Vol. 2, pp. 39–42).

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., & Shibano, Y. (1995). Agricultural robot in grape production system. In *Paper presented at the IEEE International Conference on Robotics and Automation, Nagoya, Japan* (Vol. 3, pp. 2054–2509).

Mukhopadhyayab, P., & Chaudhuria, B. B. (2015). A survey of Hough Transform. *Pattern Recognition, 48*(3), 993–1010.

Muscato, G., & Prestifilippo, M. (2005). A prototype of an orange picking robot: past history, the new robot and experimental results. *Industrial Robot: An International Journal, 32*(2), 128–138.

Pla, F., Juste, F., & Ferri, F. (1993). *Feature extraction of spherical objects in image analysis: An application to robotic citrus harvesting*. 8 pp. 57–72). B.V. Amsterdam: Elsevier Science Publishers.

Plebe, A., & Giorgio, G. (2001). Localization of spherical fruits for robotic harvesting. *Machine Vision and Applications, 13*, 70–79.

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

Sites Peter, W., & Delwiche Michael, J. (1988). Computer vision to locate fruit on a tree. *American Society of Agricultural Engineers, 31*(1), 257–263.

Slaughter, D. C., & Harrell, R. C. (1987). Color vision in robotic fruit harvesting. *American Society of Agricultural Engineers, 30*(4), 1144–1148.

Steger, C., Ulrich, M., & Wiedemann, C. (2008). *Machine vision algorithms and applications*. Darmstadt, Germany: Wiley-VCH GIT Verlag Press.

Tanigaki, K., Fujiura, T., Akase, A., & Imagawa, J. (2008). Cherry-harvesting robot. *Computers and Electronics in Agriculture, 63*(1), 65–72.

Teruo, T., Shuhuai, Z., & Hiroshi, F. (2002). *Measurement of 3-D locations of fruit by binocular stereo vision for apple harvesting in an orchard*. American Society of Agricultural Engineers. no.021102.

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

Yaqin, W., & Hua, G. (2004). Study on the segmentation and orientation of fruit image under natural environment. *Computer Engineering, 30*(13), 128–129 (in Chinese).

Zhao, J., Tow, J., & Katupitiya, J. (2005). On-tree fruit recognition using texture properties and color data. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

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.