



# Fruit detection in natural environment using partial shape matching and probabilistic Hough transform

Guichao Lin<sup>1,2</sup> · Yunchao Tang<sup>3</sup> · Xiangjun Zou<sup>1</sup> · Jiabing Cheng<sup>1</sup> · Juntao Xiong<sup>1</sup>

© Springer Science+Business Media, LLC, part of Springer Nature 2019

## Abstract

This paper proposes a novel technique for fruit detection in natural environments which is applicable in automatic harvesting robots, yield estimation systems and quality monitoring systems. As most color-based techniques are highly sensitive to illumination changes and low contrasts between fruits and leaves, the proposed technique, conversely, is based on contour information. Firstly, a discriminative shape descriptor is derived to represent geometrical properties of arbitrary fragment, and applied to a bidirectional partial shape matching to detect sub-fragments of interest that match parts of a reference contour. Then, a novel probabilistic Hough transform is developed to aggregate these sub-fragments for obtaining fruit candidates. Finally, all fruit candidates are verified by a support vector machine classifier trained on color and texture features. Citrus, tomato, pumpkin, bitter melon, towel melon and mango datasets were provided. Experiments on these datasets demonstrated that the proposed approach was competitive for detecting most type of fruits, such as green, orange, circular and non-circular, in natural environments.

**Keywords** Fruit detection · Shape descriptor · Partial shape matching · Probabilistic Hough transform · Support vector machine

## Introduction

The automatic detection of fruit is of great importance for a wide range of applications in the agriculture field. Robots equipped with automatic fruit detection can detect and locate products to harvest automatically, which may significantly reduce labor tensions in

---

✉ Yunchao Tang  
ryan.twain@gmail.com

✉ Xiangjun Zou  
xjzou1@163.com

<sup>1</sup> Key Laboratory of Key Technology on Agricultural Machine and Equipment, Ministry of Education, South China Agricultural University, 483 Wushan Road, Guangzhou 510642, China

<sup>2</sup> College of Mechanical and Electrical Engineering, Chuzhou University, 1 Huifeng West Road, Chuzhou 239000, China

<sup>3</sup> College of Urban and Rural Construction, Zhongkai University of Agriculture and Engineering, Guang Xin Road, Guangzhou 510225, China

countries such as China (Xiang et al. 2014; Zou et al. 2012, 2016; Wang et al. 2016). Automatic fruit detection may also assist in smart quality monitoring during the fruit growing process, as the appearance of the product is a valuable cue—such information can help farmers make useful decisions regarding fertilization or harvesting (Herold et al. 2005). It can also provide appropriate number of fruits in an orchard, which can be used to estimate yield as a necessary step to plan the harvest schedule (Font et al. 2014).

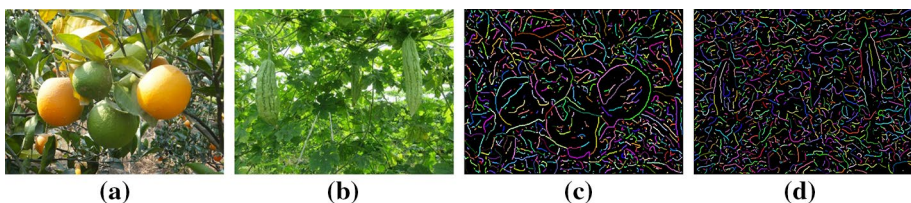
There currently exists a wide variety of color-based fruit detection methods, but challenges, such as low contrast between fruits and leaves, unstable illumination condition and occlusion, probably render them ineffective (Luo et al. 2016; Barnea et al. 2016). As the contour of an object is invariant to color and lighting conditions and remains stable to intra-category shape variation, contour is often employed for object detection (Muriillo-Bracamontes et al. 2012; Lu and Sang 2015; Li et al. 2016). The aim of this study was to exploit contour rather than color information for automatic fruit detection. Figure 1 shows the contours marked in different colors. As can be seen, there are two issues to contour-based methods: (1) target fruit contours are often broken into inconsistent fragments which may be connected to non-target fragments. An example is given in Fig. 1c, where the edge fragment of each citrus either contains part of the contour of the leaf or shadow on the fruit surface. (2) A huge proportion of the input image is occupied by non-target fragments, which thus makes valid detection extremely difficult (see Fig. 1d). Inspired by Donoser et al. (2009) who presented partial shape matching (PSM) to identify parts of an edge fragment matching parts of a given reference contour from cluttered background, PSM is applied to detect sub-fragments of interest, then a novel probabilistic Hough transform (PHT) is investigated to aggregate these sub-fragments to obtain fruit candidates, and finally a support vector machine (SVM) classifier is built to exclude false candidates.

## Related work

Automatic fruit detection with color cues has been extensively studied over recent decades (Wang et al. 2013; Payne et al. 2014; Bargoti and Underwood 2017; Puttemans et al. 2016; Zhao et al. 2016). Interested readers are recommended to read recent reviews by Kapach et al. (2012) and Gongal et al. (2015). Here, the background material of sub-fragment detection and aggregation, and fruit classification is discussed.

## Sub-fragment detection and aggregation

PSM is a commonly used, generalized algorithm (Donoser et al. 2009) to detect sub-fragments matching parts of a given reference contour. Riemenschneider et al. (2010) invented



**Fig. 1** Illustration of the challenges in fruit detection. **a** Citrus image. **b** Bitter melon image. **c, d** Contours obtained from **a** and **b**, respectively, and marked in different colors

an angular descriptor to represent geometrical properties of the fragment and applied this descriptor to PSM to detect candidate sub-fragments, followed by employing PHT to group these sub-fragments into objects. Ma and Latecki (2011) first utilized lengths and orientations cues to increase the discrimination power of the descriptor, then used PSM to obtain sub-fragments of interest, and finally aggregated these sub-fragments into possible objects by finding maximum weight subgraphs in a graph using the graph clustering method proposed by Liu et al. (2010). Such a length-based descriptor is not invariant to scale change, thus limiting the application of this method. Fan et al. (2015) proposed a variation of PSM by breaking both the edge fragments and reference contour into shorter and continuous sub-fragments between any two neighboring curvature extreme points (CEP), then calculating the similarity between each pair of sub-fragments, and selecting the top 10% best matches as final results. PHT was used to merge these sub-fragments. Unfortunately, this method is not suited to circular fruits as no CEPs exist on their contours.

Generative model, which leverages a learned reference contour, is another popular method (Fergus et al. 2003; Opelt et al. 2006; Lu et al. 2009). Shotton et al. (2008) first learned parts of the reference contour using  $k$ -medoids from training samples, then detected them in the input image by applying the oriented chamfer matching algorithm (OCM), and finally used a cascaded classifier to combine fragments. Yu et al. (2017) also utilized OCM to locate sub-fragments of interest and formulated the sub-fragment aggregation problem as minimizing a Markov random field energy solved by the loopy belief propagation algorithm (Pearl 1982). However, fruit contours are too simple to be learned. To avoid complex learning, Su et al. (2015) directly divided a template into overlapped parts, maximized the posterior probability for the fragments in the image and the locations of object center given parts of a template to obtain sub-fragments of interest, and finally applied dynamic programming to assemble these sub-fragments. This method was effective, though the length and overlapping area of the parts must be determined carefully.

Alternative approaches utilize color or texture features of the target fruit to extract fragments of interest, and use circular fitting or circular Hough transform (CHT) to obtain fruit targets (Lu and Sang 2015; Linker 2017; Lu et al. 2018). Lu et al. (2018) applied CHT on hierarchical contours of an input image to detect citrus fruits. Linker et al. (2012) used a  $k$ -nearest neighbor (KNN) classifier to detect seed areas, segmented the contours of the seed areas into the arc and non-arc sub-fragments, then grouped adjacent arc sub-fragments into apple targets. This method was modified by Qureshi et al. (2016) to detect mango fruits. Circular models render these methods only applicable, naturally, to circular fruits.

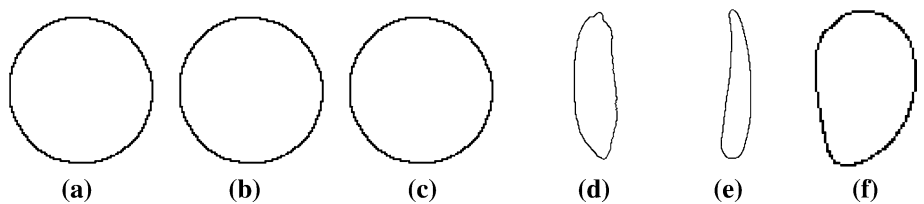
In this research, a generalized algorithm that could deal with most type of fruits, such as circular, and non-circular, in the natural environment was investigated. PSM is first utilized to detect sub-fragments of interest. PHT, which has been well-established and successfully applied in previous studies (Maji and Malik 2009; Ommer and Malik 2009), is then used to assemble candidate sub-fragments.

## Fruit classification

A classifier can be used to remove false positives after fragment aggregation. Several methods have been investigated to classify fruits, such as  $K$ -means clustering (Wachs et al. 2010), Bayesian classifier (Slaughter and Harrell 1989), KNN clustering (Kurtulmus et al. 2014), AdaBoost classifier (Luo et al. 2016), and SVM (Ji et al. 2012). Ji et al. (2012) had demonstrated the advantage of SVM in classification, therefore SVM is used in the final verification stage.

**Table 1** The proposed datasets

Category	Number of images	Number of fruits
Citrus	109	212
Tomato	86	205
Pumpkin	59	84
Bitter gourd	100	171
Towel gourd	62	105
Mango	34	330
Total	450	1107



**Fig. 2** Reference models. **a–f** are reference models of citrus, tomato, pumpkin, bitter gourd, towel gourd and mango, respectively

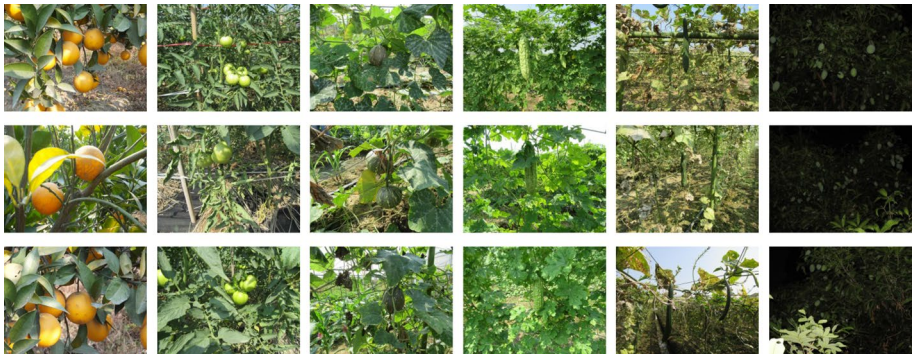
## Materials and methods

### Image acquisition

The datasets encompassed six categories of fruits: citrus, tomato, pumpkin, bitter gourd, towel gourd and mango, and comprised a total of 450 images (Table 1). Citrus images were gathered from an orchard in Guangzhou, China on December 2016, tomato, pumpkin, bitter gourd and towel gourd images were taken from the Nansha Base at the Guangzhou Academy of Agricultural Science on November 2017, and mango images were captured in an orchard located in South China Agricultural University, China, by using a consumer digital camera (DSC-W800, Sony Inc., Japan). All images were saved as JPG format at a resolution of 5184×3888 pixels. The imresize function with a bilinear interpolation method in MATLAB was performed to resize the images to 480×360 pixels to improve the computational efficiency. Ground-truth bounding boxes for each image were provided manually. A reference model for each category was also supplied (see Fig. 2). Examples of the datasets are shown in Fig. 3, and the acquisition scene of the datasets is shown in Fig. 4.

### Algorithm overview

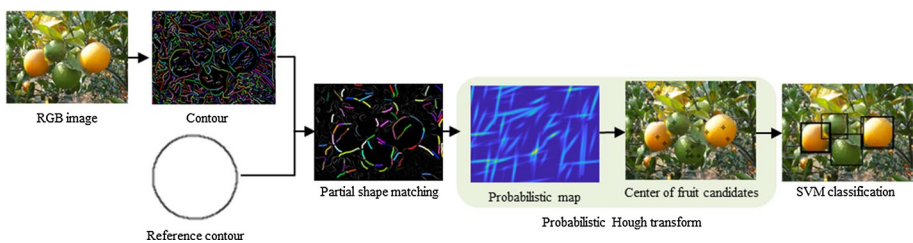
The flow chart of the proposed algorithm is shown in Fig. 5, which includes four steps. Edges of an input image are first detected by the gPb detector (Maire et al. 2008) and grouped to form fragments (Kovesi. 2000). Then, a novel shape descriptor is presented to measure the similarity between arbitrary pairs of fragments and applied to PSM for



**Fig. 3** Examples of the datasets



**Fig. 4** Acquisition scene of our datasets



**Fig. 5** The proposed contour-based fruit detection method

detecting candidate sub-fragments that are similar to parts of a given reference contour. Thirdly, a novel PHT is investigated to aggregate these sub-fragments, thus obtaining fruit candidates. Finally, an SVM classifier trained on color and texture features is used to remove false positives. More details are provided in the following subsection.

It should be noted that as gPb outperforms most edge detectors, such as Canny, Pre-witt, and Sobel, in terms of precision and recall (Arbelaez et al. 2011), gPb is used to extract edges from the input image.



## Partial shape matching

The aim is to detect candidate sub-fragments that are similar to parts of a reference contour. A rotation and scale invariant shape descriptor is first presented to capture the geometric property of a given fragment and used to measure the similarity between any pair of fragments. This descriptor is then applied to PSM to detect candidate sub-fragments.

### Shape descriptor

Let  $f$  denote an edge fragment with a sequence of edge pixels  $f = \{p_1, p_2, \dots, p_N\}$ , where  $p_i = (x_i, y_i)^T$  is the  $i$ 'th edge pixel on the fragment and  $N$  is the number of edge pixels on the fragment. Edge fragment  $f$  is smoothed by convolving with a Gaussian kernel with standard deviation  $\sigma$  (He and Yung 2008), which was set to 3 in the experiments. Next, the tangent vector  $t_i$  and normal vector  $n_i$  of each edge pixel  $p_i$  are calculated as follows:

$$\begin{cases} t_i = (\Delta x_i, \Delta y_i)^T / \sqrt{(\Delta y_i)^2 + (\Delta x_i)^2} \\ n_i = (-\Delta y_i, \Delta x_i)^T / \sqrt{(\Delta y_i)^2 + (\Delta x_i)^2} \end{cases} \quad (1)$$

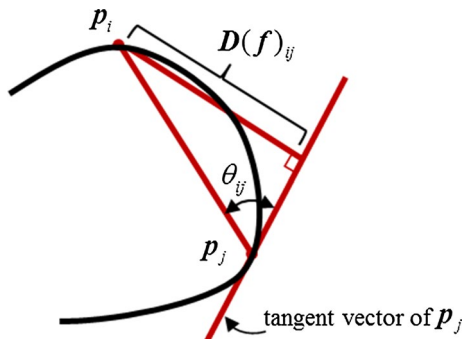
where  $\Delta x_i, \Delta y_i$  are the first-order partial derivatives in the  $x$  and  $y$  directions, respectively, and can be approximated by finite differences. Based on the tangent and normal vector cues, two  $N \times N$  matrixes, distance matrix  $D(f)$  and angle matrix  $\theta(f)$ , are defined and used as descriptors to encode the geometric structure of fragment  $f$ . The element of the  $i$ th row and  $j$ th column of distance matrix  $D(f)$  is defined as the distance from edge pixel  $p_i$  to the tangent vector of edge pixel  $p_j$  (Fig. 6):

$$D(f)_{ij} = \frac{|(p_i - p_j)^T n_j|}{p_i - p_{j2}} \quad (2)$$

where  $i, j = 1, \dots, N$ . As depicted in Eq. (2), because the distance descriptor here utilizes the distance and normal vector properties of a fragment, it is intuitively more discriminative than the descriptors developed by Ma and Latecki (2011) and Fan et al. (2015) that only used the distance property.

The element of the  $i$ th row and  $j$ th column of angle matrix  $\theta(f)$  is defined as:

**Fig. 6** Geometric structure of edge fragment



$$\theta(f)_{ij} = \min\{\theta_{ij}, \theta_{ji}\} \quad (3)$$

where  $\theta_{ij} \in [0, 2\pi]$  refers to the relative angle at which tangent vector  $t_j$  is rotated anticlockwise to be parallel to vector  $(p_i - p_j)$  (Tico and Kuosmanen 2003). Because  $\theta_{ij}$  considers the signed angle between two orientations, it is more robust than that of Ma and Latecki (2011) which directly uses the absolute angle of the orientations. Another advantage is that  $\theta(f)$  is invariant to scale and rotation.

Given another edge fragment of length  $N$ , denoted as  $g$ , an  $N \times N$  affinity matrix measuring the similarities between the corresponding pairs of elements of the two descriptors is drawn as follows (Ma and Latecki 2011):

$$A(f, g) = \exp\left(-\frac{(D(f) - D(g))^2}{\sigma_D^2}\right) + \exp\left(-\frac{(\theta(f) - \theta(g))^2}{\sigma_\theta^2}\right) \quad (4)$$

where  $\sigma_D$  and  $\sigma_\theta$  are hyper-parameters, which not only play the role of normalization, but also can be used to balance the contributions of the distance and angle descriptors. These two values were set to 0.2 and  $\pi/6$ , respectively, in the experiments. The similarity between edge fragments  $f$  and  $g$  is defined as:

$$s(f, g) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N A(f, g)_{ij} \quad (5)$$

## Principle of partial shape matching

The purpose of PSM is to identify parts of a fragment matching parts of a given reference contour.

Let  $M = \{q_1, q_2, \dots, q_M\}$  denote the edge pixel sequence of a given reference contour, where  $M$  is the number of edge pixels on this contour; let  $E = \{e^1, \dots, e^K\}$  be the edge fragments obtained from the input image, where  $e^k$  denotes the  $k$ 'th edge fragment of  $E$  with a sequence of edge pixels as  $e^k = \{p_1^k, p_2^k, \dots, p_{N_k}^k\}$ , and  $N_k$  is the number of edge pixels on the  $k$ 'th edge fragment.

To determine all sub-fragments of interest, let  $e^k(e, l) = \{p_e^k, \dots, p_{e+l}^k\} \subseteq e^k$  denote a sub-fragment of  $e^k$ , where  $l$  is the length. Similarly,  $M(m, l) = \{q_m, \dots, q_{m+l}\} \subseteq M$  denotes a sub-fragment of  $M$ . The similarity between  $e^k(e, l)$  and  $M(m, l)$  can then be calculated by using Eq. (5), which yields a 3D matrix defined by Ma and Latecki (2011):

$$\Gamma_1^k(e, m, l) = s(e^k(e, l), M(m, l)) \quad (6)$$

Note that the edge-linking algorithm (Kovesi. 2000) arranged edge pixels into fragments from either clockwise or counterclockwise directions. Consequently, if the order of a candidate sub-fragment is different from that of the given reference contour, there may be a low similarity score in the 3D matrix. Therefore, another 3D matrix is established to resolve this problem:

$$\Gamma_2^k(e, m, l) = s(e^k(e, -l), M(m, l)) \quad (7)$$

These two matrixes are computed in an efficient manner using integral image (Donoser et al. 2009), which only took a few milliseconds per edge fragment in the experiments. Because triplets  $\{e, m, l\}$  with low similarity scores or short lengths may be related to

non-target fragments,  $\Gamma_1^k(e, m, l)$  and  $\Gamma_2^k(e, m, l)$  are set to zero if their scores or lengths are below a user-defined threshold.

Any tuple  $\{e, m\}$  with length  $l$ , which has a large similarity value, is highly possible to belong to a valid match. Therefore, the above 3D matrices are compressed into 2D matrices along length  $l$  to maximize the similarity:

$$\begin{cases} L_1^k(e, m) = \operatorname{argmax}_l \Gamma_1^k(e, m, l) \\ L_2^k(e, m) = \operatorname{argmax}_l \Gamma_2^k(e, m, l) \end{cases} \quad (8)$$

where the value of  $L_1^k(e, m)$  or  $L_2^k(e, m)$  is the length  $l$ . Based on Eq. (8), a function assigning a unique length for tuple  $\{e, m\}$  is defined as:

$$L^k(e, m) = \begin{cases} L_1^k(e, m), & \text{if } \Gamma_1^k(e, m, L_1^k(e, m)) \geq \Gamma_2^k(e, m, L_2^k(e, m)) \\ -L_2^k(e, m), & \text{else} \end{cases} \quad (9)$$

where  $L^k(e, m)$  is a positive value when the order of sub-fragment  $e^k(e, l)$  is the same as the order of  $M(e, l)$ , otherwise it is a negative value. Function  $L^k$  can be visualized by a map, as shown in Fig. 7. This map reveals a few isolated regions in  $L^k$ , each of which contains a certain number of matches that have similar start points but different lengths. A match with greater length can average out irregularities, so the extremum  $l$  as well as its location  $(e, m)$  from each region is detected as a candidate match recorded in a list denoted as  $R = \{\{e_1, m_1, l_1\}, \{e_2, m_2, l_2\}, \dots\}$ . Another list  $I$  with the same length as  $R$  is used to record the index of the edge fragment on which a candidate sub-fragment is located. An example of sub-fragments detected by PSM is given in Fig. 8.

### Probabilistic Hough transform

As candidate sub-fragments detected by PSM are scattered in the input image, and some false candidate sub-fragments may exist, a reliable fragment aggregation method to form

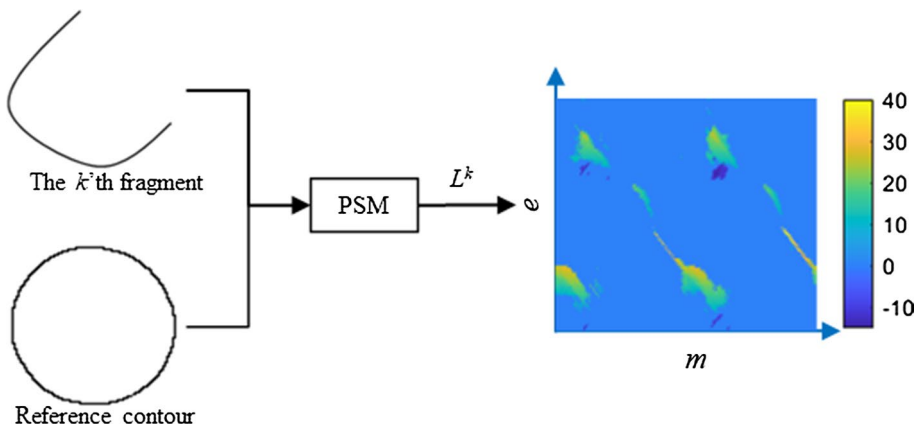
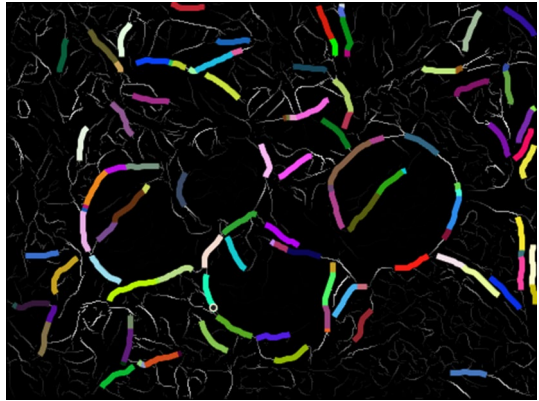


Fig. 7 Visualization of function  $L^k$



**Fig. 8** Example illustrating candidate sub-fragments of Fig. 1c detected by PSM. Each sub-fragment is marked in random color



fruit candidates is required. In this study, a new PHT was developed. Unlike traditional PHT that needs to generate codebooks from training samples (Maji and Malik 2009; Chakraborty et al. 2013), the developed PHT casts probabilistic votes directly for the locations of possible fruits based on all the observations, i.e., candidate matches.

Let  $S(\mathbf{x}, O)$  denote the score of fruit  $O$  at location  $\mathbf{x}$ . Assume that the candidate matches are independent, then  $S(\mathbf{x}, O)$  can be obtained by accumulating the probabilities  $p(\mathbf{x}, O, \{e_k, m_k, l_k\})$ :

$$\begin{aligned} S(\mathbf{x}, O) &= \sum_k p(\mathbf{x}, O, \{e_k, m_k, l_k\}) = \sum_k p(\mathbf{x}, O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \\ &= \sum_k p(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \times p(\mathbf{x}, O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \end{aligned} \quad (10)$$

As done by Maji and Malik (2009), the prior probability  $p(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  is assumed to be uniformly distributed, meaning that  $p(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) = \frac{1}{|I|}$ , with  $|I|$  the number of candidate matches. In this way,  $S(\mathbf{x}, O)$  can be reduced to:

$$\begin{aligned} S(\mathbf{x}, O) &\propto \sum_k p(\mathbf{x}, O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \\ &= \sum_k p(\mathbf{x} | O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \times p(O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \end{aligned} \quad (11)$$

where  $p(O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  refers to the probability that  $\{\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)\}$  is a true match. In the experiments,  $p(O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  was set to 1 as every candidate match  $\{\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)\}$  was obtained quite confidently by PSM (It would be possible to let  $p(O | \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  reflect the confidence of a candidate match according to its length, such as  $p(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) = \frac{|l_k|}{\sum_j |l_j|}$ , as a longer candidate is more likely to belong to a valid match). Thus:

$$S(\mathbf{x}, O) \propto \sum_k p(\mathbf{x} | O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) \quad (12)$$

where  $p(\mathbf{x} | O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  is the probability of a fruit center location, which can be estimated by accumulating the dissimilarity between the distance from each point on an

observed sub-fragment to location  $\mathbf{x}$  and that from the corresponding point on a reference contour to the center of the reference contour. Due to the fact that the point-to-line distance is more robust than Euclidean distance (Low 2004), the following point-to-line based dissimilarity metric is used:

$$d_{k,|j|} = \log \left( 1 + \left| \left( \mathbf{p}_{e_k+j}^{I\{k\}} - \mathbf{x} \right)^T \mathbf{n}_{e_k+j}^{I\{k\}} \right| \right) - \log \left( 1 + \left| \left( \mathbf{q}_{m_k+|j|} - \mathbf{c}_M \right)^T \mathbf{n}_{m_k+|j|}^M \right| \right) \quad (13)$$

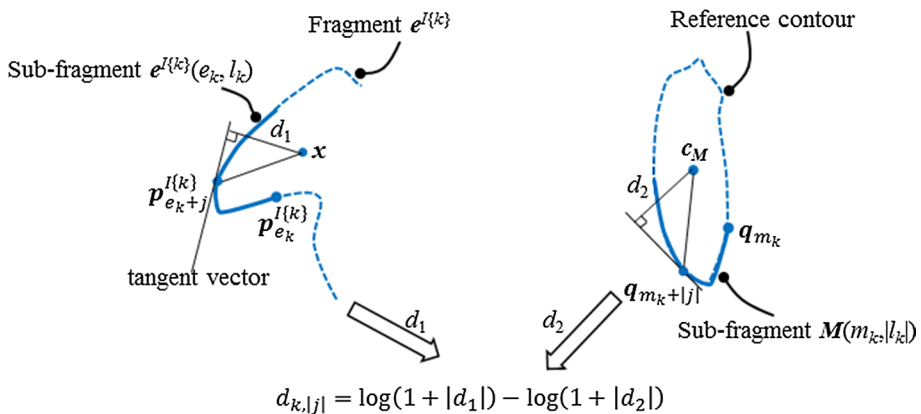
where  $d_{k,|j|}$  refers to the dissimilarity between the  $|j|$ 'th point pair of the  $k$ 'th match;  $\mathbf{p}_{e_k+j}^{I\{k\}}$  and  $\mathbf{n}_{e_k+j}^{I\{k\}}$  are the location and normal vector of the  $(e_k+j)$ 'th point on the  $I\{k\}$ 'th edge fragment, respectively;  $\mathbf{q}_{m_k+|j|}$  and  $\mathbf{n}_{m_k+|j|}^M$  are the location and normal vector of the  $(m_k+|j|)$ 'th point on the reference contour, respectively;  $\mathbf{c}_M$  is the center of the reference contour. Figure 9 shows the schematic diagram of the developed dissimilarity metric.

Based on Eq. (13), probability  $p(\mathbf{x}|O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  is defined as follows:

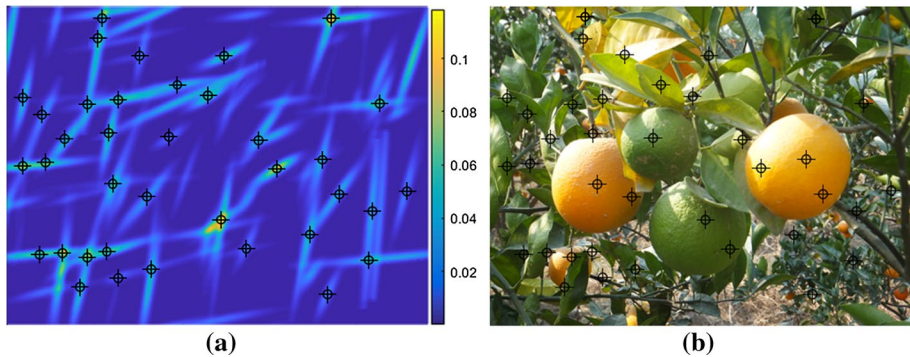
$$p(\mathbf{x}|O, \mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|)) = \exp \left( - \frac{\mathbf{d}_k^T \mathbf{d}_k}{|\mathbf{l}_k| \sigma^2} \right) \quad (14)$$

where  $\mathbf{d}_k = (d_{k,1}, d_{k,2}, \dots, d_{k,|l_k|})^T$ ;  $\sigma$  is an important constant. If  $\sigma$  is too large, the resulting probabilities will be relatively large and similar, thus losing discrimination; if  $\sigma$  is too small, some true-positive matches will have low probabilities due to disturbance. In the experiments,  $\sigma$  was set to 0.1.

To minimize the computation time, the range of  $\mathbf{x}$  for each match  $(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  is limited to a window twice the size of the reference model and centered at the center of  $\mathbf{e}^{I\{k\}}(e_k, l_k)$ . A map describing the probability distribution of the locations of fruit centers is obtained after the PHT process is completed (see Fig. 10a). This map is segmented by a given threshold, and a non-maximal suppression procedure is performed to remove repetitive candidates (see Fig. 10b).



**Fig. 9** Schematic diagram of the developed point-to-line based dissimilarity metric, where  $(\mathbf{e}^{I\{k\}}(e_k, l_k), \mathbf{M}(m_k, |l_k|))$  refers to the  $k$ 'th match



**Fig. 10** Example illustrating the result of PHT. **a** Probabilistic map of Fig. 8, where the black circle refers to the center of the possible fruit; **b** possible fruit centers in the RGB image

### Features extraction and classification

Because some false positives exist after PHT, a feature vector for each candidate is extracted and classified via an SVM classifier for removing false positives. Where there is low-contrast between fruit and leaves, such as bitter melon, texture features are more discriminative than color features. Where there is high-contrast with leaves, such as citrus, color features are preferable. In this context, a color and texture-based descriptor, named color-HOG, is proposed to describe all kinds of fruits. The color component of color-HOG comprises the mean and covariance of the HSV values in a window centered at a possible fruit center. As the covariance is a symmetric matrix, duplicate elements are not included in the color-HOG features. The texture component of color-HOG is histograms of oriented gradients (HOG) (Dalal and Triggs 2005), of which each cell is set to  $8 \times 8$  pixels and each block is  $16 \times 16$  pixels.

As a comparison, Linker et al. (2012) proposed a similar color and texture-based descriptor comprised of the statistics of RGB values in a square window centered at the pixel of interest. This descriptor is sensitive to illumination change, because varying illuminations would alter the RGB values of objects. The proposed descriptor utilizes the HOG features extracted from the gradient image which is invariant to lighting conditions, and thus relatively discriminative.

### Results and discussion

Two quantitative experiments were performed, one for verifying the performance of the color-HOG descriptor, and one for examining the detection performance of the proposed algorithm. All codes were implemented using MATLAB (R2014a) on a computer with 4 GB RAM.

#### Evaluation of color-HOG descriptor

The color-HOG descriptor was evaluated by comparing two state-of-the-art descriptors, HOG (Dalal and Triggs 2005) and LBP (Ahonen et al. 2009). These three descriptors

were utilized to train the SVM classifier, and precision and recall were used to evaluate the classification performance. Both positive and negative samples were obtained manually from one-third of the images of each dataset, among which 75% of samples were used as a training set, and the others were a test set. Note that 10-fold cross validation and a grid search method were utilized to optimize the parameters of SVM. Table 2 lists the experiment results of the three descriptors on the test set, which clearly showed that color-HOG outperformed HOG and LBP in terms of precision and recall.

## Evaluation of detection performance of the proposed algorithm

The performance of the proposed algorithm (named “PSMPHT”) was verified and compared with two methods, one was a similar contour-based detection method (Ma and Latecki 2011), and one was a modified version of the developed method by replacing PHT with a graph clustering method (Bulò and Pelillo 2013), named “PSMGC”. A detection was considered correct if the intersection of the detected and the ground-truth bounding box over their union was larger than 50% (Everingham et al. 2015). Precision and recall were used to assess the detection performance. Note that since one-third of the images per category were used to train the SVM, only the remaining images were subject to this experiment.

Tables 3 and 4 list the precision and recall of the three methods. PSMPHT had two best precisions and four best recalls, PSMGC obtained three best precisions and two best recalls, and Ma and Latecki’s method only had two best precisions, i.e., both PSMPHT and PSMGC obviously outperformed Ma and Latecki’s method. The main reason is that PSMPHT and PSMGC use the proposed bidirectional PSM, which searches the sub-fragments of interest along the input fragment from both the clockwise and anticlockwise directions, thus obtaining more confident candidate sub-fragments than the unidirectional PSM used by Ma and Latecki (2011). The developed shape descriptor encodes the tangent and normal properties of the fragment, which also contributes to the favorable performance of PSMPHT and PSMGC.

PSMPHT performed better than PSMGC in terms of the overall performance in this study. Because both PSMPHT and PSMGC use the proposed PSM to detect sub-fragments of interest, the performance difference between the two methods mainly lies in the aggregation method as-applied. As each match independently casts votes for the locations of fruit candidates, PHT can generate a multi-peak probabilistic map from which the local maxima are detected as fruit candidates. This characteristic allows PHT to obtain more confident candidates than GC and therefore enabled PSMPHT with larger precision and recall than GC. However, there was an exception. In the pumpkin category, as the pumpkin and its leaves have similar shapes, PHT would detect many false-positive candidates (Fig. 11a). This issue decreased the precision and recall of PSMPHT over the pumpkin category. A failure example is shown in Fig. 11. In general, PSMPHT has better overall performance than PSMGC in natural environments.

The mango dataset results showed that these three algorithms achieved high precision, as the nighttime environment inhibited the effects of illumination changes and thus improved the SVM classifier performance. Additionally, PSMPHT obtained a higher recall than the other algorithms. The main reason was that only one sub-fragment on the mango contour was detected as the mangoes appeared very small in the captured image, so GC could not find reliable maximum weight subgraphs from the constructed weighted graph,

**Table 2** Precision and recall of the color-HOG, HOG and LBP descriptors

Feature descriptor	Precision					Recall				
	Citrus	Tomato	Pumpkin	Bitter gourd	Towel gourd	Mango	Citrus	Tomato	Pumpkin	Bitter gourd
Color-HOG	0.923	<b>1</b>	<b>1</b>	0.967	<b>0.968</b>	<b>1</b>	<b>0.909</b>	<b>0.947</b>	<b>1</b>	<b>0.938</b>
HOG	<b>1</b>	0.985	0.954	<b>1</b>	0.898	<b>1</b>	0.904	0.896	0.969	0.929
LBP	0.833	0.896	0.954	0.893	0.939	0.833	0.300	0.776	0.308	0.926

Bold values indicate the best values

**Table 3** Precision of PSMPHT, PSMGC, and Ma and Latecki’s method

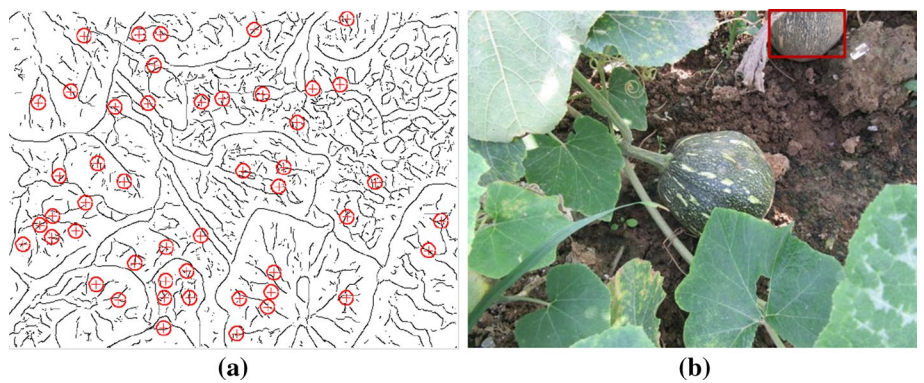
	PSMPHT	PSMGC	Ma and Latecki’s method
Citrus	<b>0.783</b>	0.762	0.756
Tomato	<b>0.848</b>	0.675	0.681
Pumpkin	0.745	<b>0.754</b>	0.750
Bitter gourd	0.762	0.676	<b>0.766</b>
Towel gourd	0.807	<b>0.862</b>	0.811
Mango (at night-time)	0.919	<b>1</b>	<b>1</b>

Bold values indicate the best values

**Table 4** Recall of PSMPHT, PSMGC, and Ma and Latecki’s method

	PSMPHT	PSMGC	Ma and Latecki’s method
Citrus	<b>0.802</b>	0.729	0.453
Tomato	0.582	<b>0.594</b>	0.368
Pumpkin	0.678	<b>0.768</b>	0.267
Bitter gourd	<b>0.791</b>	0.476	0.343
Towel gourd	<b>0.754</b>	0.409	0.492
Mango (at night-time)	<b>0.825</b>	0.118	0.113

Bold values indicate the best values

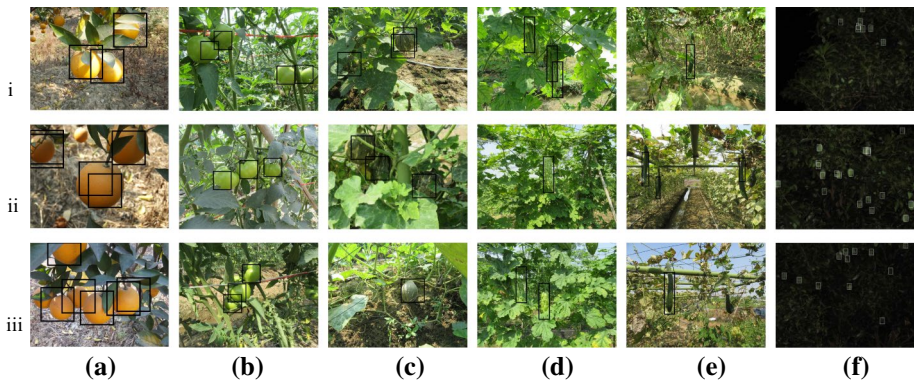


**Fig. 11** Failure example. **a** Shows possible locations of pumpkins detected by PHT. **b** Shows final results by SVM classifier

but PHT could still cast reliable votes for the fruit location. This result again confirmed that PSMPHT is highly effective.

Several detection results are shown in Fig. 12. These examples intuitively validated that PSMPHT is effective. Some advantages are presented: (a) the proposed algorithm could detect any type of fruits, such as green, orange, circular, non-circular, etc., i.e., it was a generalized algorithm, and (b) the proposed algorithm was robust under complex conditions,





**Fig. 12** Example illustrating detection results of PSMPT

such as occlusion and illumination changes (see Fig. 12ai), low contrast between fruits and leaves (see Fig. 12d) and cluttered background. Also, a disadvantage was uncovered: the developed method would obtain overlapping detections (see Fig. 12aii) or miss some true positives (see Fig. 12eiii). A possible reason is that the proposed PHT uses a scale-variant dissimilarity metric to cast probabilistic votes, so PHT would generate multiple local maximums for large-sized fruits or is unable to detect fruits with large scale changes. Therefore, future research will focus on improving the scale invariance of PHT.

## Conclusions

This paper presented a framework for detecting a wide array of fruit types—green, orange, circular and non-circular were tested here—in a natural environment. Detection of sub-fragments of interest was performed using a bidirectional PSM technique, which traces the input fragment from two opposite directions to find possible sub-fragments matching parts of a given reference model. These sub-fragments of interest were grouped as fruit candidates using PHT. False positives were excluded by an SVM classifier.

The proposed algorithm was evaluated on image datasets captured in the natural environment. The datasets comprised citrus, tomato, pumpkin, bitter gourd, towel gourd and mango images, i.e., several different types of fruits. For citrus, tomato, pumpkin, bitter gourd, towel gourd and mango categories, the precision of the proposed algorithm was 0.783, 0.848, 0.745, 0.762, 0.807 and 0.919, respectively; the recall was 0.802, 0.582, 0.678, 0.791, 0.754 and 0.825, respectively. The results are comparable to a similar contour-based detection method. In other words, the proposed algorithm has competitive advantages.

PSM and PHT were used for sub-fragment detection and aggregation without necessitating the painstaking design of specific features for each type of fruit. This makes the proposed algorithm a generalized method. PHT utilizes a scale-variant dissimilarity metric to determine the probability value of a vote, so it may fail to detect fruits with large scale change. In the future, a scale-invariant PHT will be established.

**Acknowledgements** This work was funded by a grant from the National Natural Science Foundation of China (No. 31571568), and the Project of Province Science and Technology of Guangdong (No. 2017A030222005).

## References

- Ahonen, T., Matas, J., He, C., & Pietikainen, M. (2009). Rotation invariant image description with local binary pattern histogram fourier features. In *Proceedings of the 16th Scandinavian Conference on Image Analysis* (pp. 61–70). Heidelberg, Germany: Springer.
- Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 898–916.
- Bargoti, S., & Underwood, J. P. (2017). Image segmentation for fruit detection and yield estimation in apple orchards. *Journal of Field Robotics*, 34(6), 1039–1060.
- Barnea, E., Mairon, R., & Ben-Shahar, O. (2016). Colour-agnostic shape-based 3d fruit detection for crop harvesting robots. *Biosystems Engineering*, 146, 57–70.
- Bulò, S. R., & Pelillo, M. (2013). A game-theoretic approach to hypergraph clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6), 1312–1327.
- Chakraborty, B., González, J., & Roca, F. X. (2013). Large scale continuous visual event recognition using max-margin Hough transformation framework. *Computer Vision and Image Understanding*, 117(10), 1356–1368.
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 886–893). Los Alamitos, USA: IEEE Computer Society Press.
- Donoser, M., Riemenschneider, H., & Bischof, H. (2009). Efficient partial shape matching of outer contours. In *Proceedings of the 9th Asian Conference on Computer Vision* (pp. 281–292). Heidelberg, Germany: Springer.
- Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal Visual Object Classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1), 98–136.
- Fan, H., Yang, C., & Tang, Y. (2015). Object detection based on scale-invariant partial shape matching. *Machine Vision and Applications*, 26(6), 711–721.
- Fergus, R., Perona, P., & Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 264–271). Los Alamitos, USA: IEEE Computer Society Press.
- Font, D., Pallejà, T., Tresanchez, M., Teixidó, M., Martínez, D., Moreno, J., et al. (2014). Counting red grapes in vineyards by detecting specular spherical reflection peaks in RGB images obtained at night with artificial illumination. *Computers & Electronics in Agriculture*, 108(C), 105–111.
- Gongal, A., Amatya, S., Karkee, M., Zhang, Q., & Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. *Computers & Electronics in Agriculture*, 116(C), 8–19.
- He, A. X., & Yung, N. H. C. (2008). Corner detector based on global and local curvature properties. *Optical Engineering*, 47(5), 057008.
- Herold, B., Truppel, I., Zude, M., & Geyer, M. (2005). Spectral measurements on ‘elstar’ apples during fruit development on the tree. *Biosystems Engineering*, 91(2), 173–182.
- Ji, W., Zhao, D., Cheng, F., Xu, B., Zhang, Y., & Wang, J. (2012). Automatic recognition vision system guided for apple harvesting robot. *Computers & Electrical Engineering*, 38(5), 1186–1195.
- Kapach, K., Barnea, E., Mairon, R., Edan, Y., & Ben-Shahar, O. (2012). Computer vision for fruit harvesting robots—State of the art and challenges ahead. *International Journal of Computational Vision & Robotics*, 3(1/2), 4–34.
- Kovesi, P. D. (2000). *Matlab and octave functions for computer vision and image processing*. <http://www.peterkovesi.com/matlabfns>. Accessed 25 August 2017.
- Kurtulmus, F., Lee, W. S., & Vardar, A. (2014). Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. *Precision Agriculture*, 15(1), 57–79.
- Li, H., Lee, W. S., & Wang, K. (2016). Immature green citrus fruit detection and counting based on fast normalized cross correlation (FNCC) using natural outdoor colour images. *Precision Agriculture*, 17(6), 678–697.
- Linker, R. (2017). A procedure for estimating the number of green mature apples in night-time orchard images using light distribution and its application to yield estimation. *Precision Agriculture*, 18(1), 59–75.
- Linker, R., Cohen, O., & Naor, A. (2012). Determination of the number of green apples in RGB images recorded in orchards. *Computers and Electronics in Agriculture*, 81(1), 45–57.
- Liu, H., Latecki, L. J., & Yan, S. (2010). Robust clustering as ensembles of affinity relations. *Advances in Neural Information Processing Systems*, 23, 1414–1422.

- Low, K. (2004). *Linear least-squares optimization for point-to-plane ICP surface registration*. Technical report, TR04-004, University of North Carolina, Chapel Hill, USA.
- Lu, C., Latecki, L. J., Adluru, N., Yang, X., & Ling, H. (2009). Shape guided contour grouping with particle filters. In *Proceedings of the IEEE 12th International Conference on Computer Vision* (pp. 2288–2295). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Lu, J., Lee, W. S., Gan, H., & Hu, X. (2018). Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis. *Biosystems Engineering*, 171, 78–90.
- Lu, J., & Sang, N. (2015). Detecting citrus fruits and occlusion recovery under natural illumination conditions. *Computers and Electronics in Agriculture*, 110(C), 121–130.
- Luo, L., Tang, Y., Zou, X., Wang, C., Zhang, P., & Feng, W. (2016). Robust grape cluster detection in a vineyard by combining the adaboost framework and multiple color components. *Sensors*, 16(12), 2098.
- Ma, T., & Latecki, L. J. (2011). From partial shape matching through local deformation to robust global shape similarity for object detection. In *Proceedings of the 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1441–1448). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Maire, M., Arbelaez, P., Fowlkes, C., & Malik, J. (2008). Using contours to detect and localize junctions in natural images. *Computer Vision and Pattern Recognition*, 11, 1–8.
- Maji, S., & Malik, J. (2009). Object detection using a max-margin Hough transform. In *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1038–1045). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Murillo-Bramamontes, E. A., Martinez-Rosas, M. E., Miranda-Velasco, M. M., Martinez-Reyes, H. L., Martinez-Sandoval, J. R., & Cervantes-De-Avila, H. (2012). Implementation of Hough transform for fruit image segmentation. *Procedia Engineering*, 35(12), 230–239.
- Ommer, B., & Malik, J. (2009). Multi-scale object detection by clustering lines. In *Proceedings of the IEEE 12th International Conference on Computer Vision* (pp. 484–491). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Opelt, A., Pinz, A., & Zisserman, A. (2006). A boundary-fragment-model for object detection. In *Proceedings of the 9th European Conference on Computer Vision* (pp. 575–588). Heidelberg, Germany: Springer.
- Payne, A., Walsh, K., Subedi, P., & Jarvis, D. (2014). Estimating mango crop yield using image analysis using fruit at ‘stone hardening’ stage and night time imaging. *Computers and Electronics in Agriculture*, 100, 160–167.
- Pearl, J. (1982). Reverend bayes on inference engines: a distributed hierarchical approach. In *Proceedings of the Second AAAI Conference on Artificial Intelligence* (pp. 133–136). Cambridge, UK: AAAI Press.
- Puttemans, S., Vanbrabant, Y., Tits, L., & Goedeme, T. (2016). Automated visual fruit detection for harvest estimation and robotic harvesting. In *Proceedings of the Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1–6). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Qureshi, W. S., Payne, A., Walsh, K. B., Linker, R., Cohen, O., & Dailey, M. N. (2016). Machine vision for counting fruit on mango tree canopies. *Precision Agriculture*, 17(3), 1–21.
- Riemenschneider, H., Donoser, M., & Bischof, H. (2010). Using Partial Edge Contour Matches for Efficient Object Category Localization. In *Proceedings of the 11th European Conference on Computer Vision* (pp. 29–42). Heidelberg, Germany: Springer.
- Shotton, J., Blake, A., & Cipolla, R. (2008). Multiscale categorical object recognition using contour fragments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7), 1270–1281.
- Slaughter, D. C., & Harrell, R. C. (1989). Discriminating fruit for robotic harvest using color in natural outdoor scenes. *Transactions of the ASAE*, 32(2), 0757–0763.
- Su, Y., Liu, Y., Cuan, B., & Zheng, N. (2015). Contour guided hierarchical model for shape matching. In *Proceedings of the 2015 IEEE International Conference on Computer Vision* (pp. 1609–1617). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Tico, M., & Kuosmanen, P. (2003). Fingerprint matching using an orientation-based minutia descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8), 1009–1014.
- Wachs, J. P., Stern, H. I., Burks, T., & Alchanatis, V. (2010). Low and high-level visual feature-based apple detection from multi-modal images. *Precision Agriculture*, 11(6), 717–735.
- Wang, Q., Nuske, S., Bergerman, M., & Singh, S. (2013). Automated crop yield estimation for apple orchards. In J. Desai, G. Dudek, O. Khatib, & V. Kumar (Eds.), *Experimental robotics. Springer Tracts in Advanced Robotics* (Vol. 88). Heidelberg, Germany: Springer.
- 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.
- Yu, Q., Wei, H., & Yang, C. (2017). Local part chamfer matching for shape-based object detection. *Pattern Recognition*, 65, 82–96.
- Zhao, C., Lee, W. S., & He, D. (2016). Immature green citrus detection based on colour feature and sum of absolute transformed difference (SATD) using colour images in the citrus grove. *Computers and Electronics in Agriculture*, 124, 243–253.
- Zou, X., Ye, M., Luo, C., Xiong, J., Luo, L., Wang, H., 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 modelling. *Machine Vision and Applications*, 23(1), 43–63.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.