Available online at www.sciencedirect.com

ScienceDirect

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

Research Paper

A coarse-to-fine leaf detection approach based on leaf skeleton identification and joint segmentation

Liankuan Zhang [a], *Chunlei Xia* [b,c,**], *Deqin Xiao* [a], *Paul Weckler* [d], *Yubin Lan* [e], *Jang M. Lee* [c,*]

[a] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
[b] Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, China
[c] Department of Electronics Engineering, Pusan National University, Busan 46241, South Korea
[d] Department of Biosystems and Agricultural Engineering, Oklahoma State University, Stillwater 74078, USA
[e] College of Engineering, South China Agricultural University, Guangzhou 510642, China

ARTICLE INFO

Article history:
Received 8 October 2018
Received in revised form
1 February 2021
Accepted 31 March 2021
Published online 20 April 2021

Keywords:
Leaf extraction
Leaf pose measurement
Leaf shape estimation
Occlusion detection
Active shape model
Plant image analysis

Plant leaf detection and segmentation are challenging tasks for *in-situ* plant image analysis. Here, a novel leaf detection scheme is proposed to detect individual leaves and accurately determine leaf shapes in natural scenes. A leaf skeleton-extraction method was developed by analysing local image features of skeleton pixels. Approximate positions of individual leaves were determined according to the main leaf skeleton. Sub-images containing only single target leaves were extracted from whole plant images according to position and size of the main skeleton. Accurate leaf analysis was conducted on the sub-images of individual leaves. Leaf direction was calculated by examining the structure of the main leaf skeleton. Joint segmentation by combining region and active shape model was presented to accurately elucidate leaf shape. Leaf detection was implemented using deep learning approach, Faster R–CNN. A plant leaf image dataset containing four types of leaf images of different complexity was built to evaluate detection algorithms. Plant leaves with occlusions and complex backgrounds were effectively detected and their shapes accurately determined. Detection accuracy of the proposed method was 81.10%–100%, and 86.75%–100% for Faster R–CNN. The method demonstrated a comparable detection ability to that of Faster R–CNN. Furthermore, the rates of success to determine leaf direction by our method ranged between 89.06% and 100%, while the average measurement difference was 1.29° compared with manual measurement. The accuracy of shape measurement was 75.95%–100% for all types of plant images. Therefore, this method is accurate and stable for precise leaf measurements in agricultural applications.

* *Corresponding author.*
** *Corresponding author.* Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, 264003, China.
E-mail addresses: clxia@yic.ac.cn (C. Xia), jmlee@pusan.ac.kr (J.M. Lee).

https://doi.org/10.1016/j.biosystemseng.2021.03.017
1537-5110/

Nomenclature

ACM	Active Contour Models
ASM	Active Shape Models
$\boldsymbol{b}$	Vector of weights
CNN	Convolutional neural networks
D	Square region
DSLR	Digital Single Lens Reflex
EM	Expectation-Maximization
GMM	Gaussian Mixture Models
HSI	Hue, Saturation and Intensity colour space
k-NN	k-Nearest Neighbour
n	Number of landmark points
N	Size of a square region D
N_{asm}	Total number of leaf shapes in the training set
NIR	Near-infrared
p	A pixel in a plant image
P	Matrix of first t eigenvectors
q	A pixel from a square region D
RANSAC	Random sample consensus
RGB	Red, Green and Blue colour space
RGB-D	Red, Green and Blue colour space and Depth map
$Smooth(\,)$	Smooth degree
SVM	Support Vector Machine
t	Number of eigenvectors
T	Transpose operator
$\overline{\mathbf{x}}$	Mean shape
$\mathbf{x}_i$	A vector of 2D points that representing leaf shapes
$x_{i,j}$	x coordinate of jth landmark in the ith training shape
$y_{i,j}$	y coordinate of jth landmark in the ith training shape
$\psi()$	Intensity value of a pixel

1. Introduction

Over the last decades, studies of plant image analysis under field conditions have increased due to advances in image processing and pattern recognition techniques. Plant image analysis is a fundamental task for achieving precision automation systems in agricultural practises, such as micro-spray, deleafing, or plant inspection (Ota et al., 2007; Slaughter et al., 2008).

Plant leaf analysis is a crucial issue for agricultural automation systems because plant leaves contain abundant information regarding plant growth, health status, and disease, among other biological phenomena. Leaf image analysis is also a challenging task because of shape and position variations, especially under field conditions. Most previous studies have focused on species identification, disease recognition, leaf segmentation, and leaf detection. Species identification or disease recognition of plant leaves focuses the best distinguishable feature descriptors and optimised classifiers to produce accurate predictions of plant species and diseases (Mata-Montero & Carranza-Rojas, 2016; Wäldchen & Mäder, 2018). In order to obtain accurate shape and texture features, much of the recognition work of plant leaves or diseases is usually carried out on individual leaf images in a white background under laboratory conditions (Horaisová & Kukal, 2016; Zhao et al., 2015). Leaf recognition against natural backgrounds has also been reported after the introduction of computational features and accurate leaf segmentation approaches (Olsen et al., 2015; Wang et al., 2008). Recently, deep learning has greatly increased the accuracy and the number of target species for plant leaf recognition (Barré et al., 2017; Dyrmann et al., 2016).

Numerous segmentation approaches have been developed to extract individual leaf images from field complex images. Thus, for example, a watershed segmentation of the HSI colour space was presented to extract an individual leaf image from a natural background (Tang et al., 2009). Similarly, Wang et al. (2013) applied Otsu and Canny operators to segment the region of a leaf in a sub-image containing a single leaf. They used a set of Boolean, morphology operations, and a shape-identification algorithm to obtain accurate edges of target leaves. In turn, Viaud et al. (2017) transformed the foreground of a plant image into a Euclidean distance image and used a watershed-based approach to obtain a set of segments; then, the second operation was implemented based on ellipsoid-shaped leaves to refine the segmentation.

The above methods were proposed based on low-level image features. No structural constraints evolved in these methods; therefore, over-segmentation might occur. Active contour models (ACM) were employed to segment individual leaves of various shapes (De Vylder et al., 2011). ACM is rather flexible in fitting the shape of a variety of leaves. The contour of ACM is constrained to ensure that the segmentation results are maintained within regular shapes. To that effect, Cerutti et al. (2011) proposed a parametric active polygon defined by 10 points and four numeric parameters to model the general shape of a leaf. They initialised a region in the middle of the image and generated a colour distance map based on a 2-component GMM estimated in the initial region. Then, the parameters of the leaf model were adjusted within an authorised range to produce the largest region with few colour-distant pixels. Active shape models (ASM) for the extraction and classification of crops using field images were reported by Persson and Åstrand (2008). The ASM was trained using the image samples of weed leaves. Three ASMs were constructed using different images with different degrees of complexity. The ASM-extracted leaves were further classified using a k-NN classifier.

All the studies above targeted plant images containing only one leaf. Their purpose was to segment individual leaves from the background. However, plant images in agricultural fields usually contain multiple leaves. The detection and segmentation of individual leaves from natural plant images with multiple leaves is a practical requirement of agricultural automation systems. Severe occlusions and position variations in plant leaves frequently occur in natural plant images. Model-based leaf detection and segmentation schemes have been widely adopted. For instance, Manh et al. (2001) developed a parametric deformable leaf model to describe shape variations. The leaf model was applied to the segmentation of

individual weed leaves. Recently, modified ASMs were proposed by integrating a leaf boundary classifier to detect individual leaves under field conditions (Xia et al., 2013). ASM learns shape variation from training samples and can determine the shape of the occluded parts of the leaves under study. In addition to 2D plant image analysis, depth information has been introduced for plant leaf detection. Thus, for example, Teng et al. (2011) applied an optical flow estimation algorithm and a camera self-calibration algorithm to recover 3D scene points. The RGB-D camera was used to capture colour and depth images of the plants. The mean-shift was conducted to remove background on the depth image, and ACM was applied to segment individual leaves. However, additional instrumentation is required to achieve depth images in these studies (Xia et al., 2015).

It is very important to detect the position of individual leaves in whole-plant images. Although the sophisticated leaf segmentation approaches have shown excellent performance in complicated environments, these approaches greatly rely on initialisation accuracy, such as position, size, or orientation. Leaf tips, boundaries, and skeletons have been adopted to estimate the approximate positions of leaves (Manh et al., 2001; Xia et al., 2013; Zhang et al., 2016). Recently, deep learning approaches have achieved promising performance in counting plant leaves. A convolutional neural network (CNN) can extract robust image features for describing leaf images. Many types of CNNs have been applied to solve leaf counting and segmentation problems. Thus, for instance, Dobrescu et al. (2017) presented a modified deep residual network for counting rosette leaves. Similarly, Giuffrida et al. (2018) developed a multimodal architecture of a deep neural network for combining RGB, NIR, and fluorescence features. The accuracy of this counting work was 88.5% for rosette leaves. In turn, Kumar and Dominic (2020) applied the orthogonal transform to plant region segmentation and improved the accuracy for leaf counting by fine-tuning AlexNet and VGG net. These studies were successful in detecting and counting individual leaves on a public dataset; however, estimating leaf position under natural conditions has not been widely studied.

In this study, a coarse-to-fine leaf estimation scheme is proposed to detect individual leaves and determine leaf shape under field conditions. The approximate position of individual leaves was determined by detecting the main leaf skeletons. Sub-images containing single leaves were subsequently extracted from the whole plant image to reduce noise and complexity for leaf analysis. Leaf direction was estimated by examining the structural characteristics of the main leaf skeleton. A joint segmentation scheme combining low-level segmentation and model-based object detection was proposed to accurately determine the leaf shape, including occluded leaves. A novel in situ leaf detection scheme was proposed to extract information from individual leaves in complex natural scenes. The proposed leaf detection scheme effectively determined the precise shape of individual occluded leaves and predicted the shape of the occluded part of the leaves. The method effectively detected individual leaves from plant leaf images containing a large number of leaves without need for an additional training process for leaf detection.

2. Materials

Sweet potato (*Ipomea batatas* L.) plants were selected for leaf extraction. Plant images were captured from a field located in Guangzhou, Guangdong province, China. The plant images were captured by two types of high-resolution digital imaging devices: a professional digital single-lens reflex (DSLR) camera (Cannon EOS-700D) with 18 M pixel resolution (maximal image resolution: 3456 × 5184 pixels), and a Samsung smartphone (SCH-P729) with an 8 M pixel camera with a maximum image resolution of 3264 × 2448 pixels. All plant images containing multiple leaves in the dataset were scaled to 1556 × 1037 pixels to reduce the computational cost. Single leaf images were scaled to 805 × 790 to maintain the leaves of a similar size. Automatic focussing and automatic parameter settings of the cameras were selected to capture the plant images. The distance between the camera and the plant leaves was varied from 0.1 to 1.8 m. Therefore, various scales and different numbers of plant leaves could be captured in the image set. An example of a sweet potato image collected from the field is shown in Fig. 1a. The leaf skeleton in this study indicates the main leaf skeleton and branches, as shown in Fig. 1b.

Plant leaf images were divided into four types according to the complexity of the specific natural scene (Fig. 2): single-leaf images with non-green background, single-leaf images with green background, images with multiple leaves, and images with numerous leaves. A single-leaf image with a non-green background contained only one leaf, and most of the background consisted in the ground surface; further, a small part of the leaf overlaps with the plant stem and other young leaves. A single-leaf image with a green background was generally a close-up photograph of an individual leaf above the neighbouring leaves. Many partially occluded leaves occur in the background. Multiple-leaf images captured an isolated individual sweet potato plant from the top view. In this case, the leaves are positioned in various directions and overlap among them, thereby occluding one another. The number of leaves was less than 10 in all single-plant images. Numerous-leaf images contained more than 10 leaves from different plants. The leaves were crowded in the images of numerous leaves. Leaves were of various sizes and positioned in different directions, frequently giving raise to occlusions to different extents.

Plant images were collected in the spring of 2018 and 2019 when plants were young and most leaves maintained regular shapes. Leaf images were acquired on April 12 to 15, 2018 and on April 16 to 18, 2019 by two groups of staff. To avoid shadow interference, leaf images were collected either on cloudy days or early in the morning, or late in the afternoon. Images were captured from randomly selected sweet potato plants in the field. In our test dataset, 190 images were obtained. In all, 2219 sweet potato leaves were selected and labelled as detection targets. The number of target leaves in a single image ranged from one to 63. The details of our dataset are listed in Table 1. The number of single-leaf images was 25 for both non-green background images and green background images. There were 344 target leaves in 60 images with multiple leaves, and 1825 in 80 images with numerous leaves. The mean number of

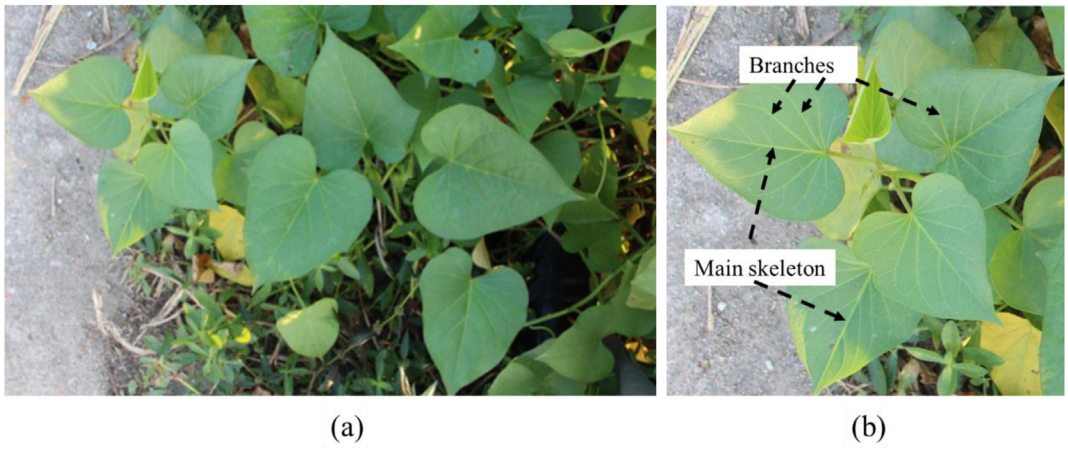

Fig. 1 – Plant images and skeletons, (a) original leaf image, (b) leaf skeleton and branches.

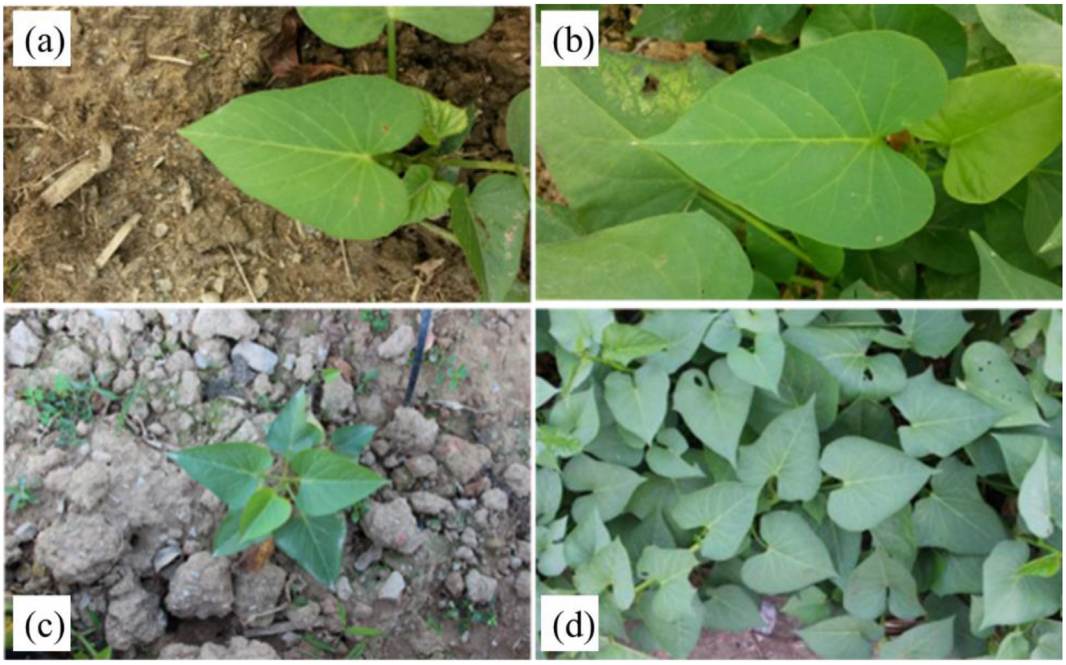

Fig. 2 – Categories of plant leaf images; (a) single leaf image with non-green background, (b) single leaf image with green background, (c) multiple-leaf image, and (d) numerous-leaf image. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1 – Description of leaf-image dataset.

Image type	No. of images	No. of target leaves	Max	Min	Mean	Collection Date	Operators
Single-leaf with non-green background	25	25	1	1	1.00	Apr, 2018	Group 1
Single-leaf with green background	25	25	1	1	1.00	Apr, 2018	Group 1
Multiple leaves	60	317	9	3	5.7	Apr, 2018	Group 1
						Apr, 2019	Group 2
Numerous leaves	80	1609	63	10	21.4	Apr, 2018	Group 1
						April, 2019	Group 2

leaves was 5.7 and 21.4 for multiple-leaf images and numerous-leaf images, respectively. Plant leaves with a regular shape and proper size were identified as detection targets. Small leaves and damaged leaves that could not show the main leaf vein were not considered as detection targets. Weeds were excluded from the study.

3. Methods

3.1. Overview of the leaf segmentation system

A coarse-to-fine leaf shape determination process was proposed based on leaf skeleton analysis and sub-image segmentation. The overall procedure consists of three main steps: pre-processing, main-leaf skeleton analysis and sub-image extraction, and joint segmentation (Fig. 3). The background is initially removed from the plant images for noise reduction. In leaf skeleton analysis, candidate skeleton pixels are detected, and leaf skeletons are extracted. Then, the leaf main skeleton is identified from the skeleton to represent leaf direction and approximate leaf size. The individual leaf image is extracted from the whole-plant image according to the size and position of the main leaf skeleton. The leaf direction is determined by examining the structure of the main skeleton. Accurate segmentation of individual leaves is conducted on a sub-image that contains only the target leaf. Region growing segmentation and active shape models are utilised to accurately extract the shape of individual leaves from the plant images.

3.2. Background removal from plant images

Background removal is a common process in plant leaf-image analysis. The complexity of plant images can be largely reduced by removing the background. In particular, background images in agricultural fields can produce numerous noises that reduce the accuracy of plant image analysis. In this work, the background was initially removed from the plant image. A leaf surface extraction method developed in our previous work was utilised to achieve background removal. Leaf surfaces are usually present in large areas with a homogeneous colour. This method examines the smoothness of the neighbouring area of each pixel. For a given pixel p, the smoothness degree of p is defined as:

$$Smooth(p) = \frac{1}{N}\sum_{q\in D}(|\psi(p) - |\psi(q)||) \quad (1)$$

where $\psi()$ represents the intensity value of a pixel in the G channel in the RGB colour space; q is a pixel from the square region D, which is centred at p. The size of D is defined as $N \times N$ ($N = 9$). Accordingly, a smoothness threshold is applied to determine whether the pixels belong to the leaf surface. The smoothness threshold was set to 12 in this study. The values of parameter N and the smoothness threshold were determined by experiments on a large number of plant images. These parameters are related to image resolution. In this work, the plant images were scaled to a given size; therefore, the fixed parameter enhanced high performance in background removal. These values can also be determined by conducting preliminary tests under different environmental conditions. Owing to the complexity of field images, background noise images (e.g. weeds) may present features similar to those of the leaf surface. These unexpected objects are usually small and can be eliminated by morphological erosion-dilation operations. Similarly, the spots on the leaf surface produced small wholes after the background removal process, which could also be removed. The isolated small areas (e.g. less than 500 pixels) should be removed as background noise. The plant image in Fig. 1a after background removal is shown in Fig. 4.

Fig. 4 – Leaf image after background removal from the original image to eliminate noise.

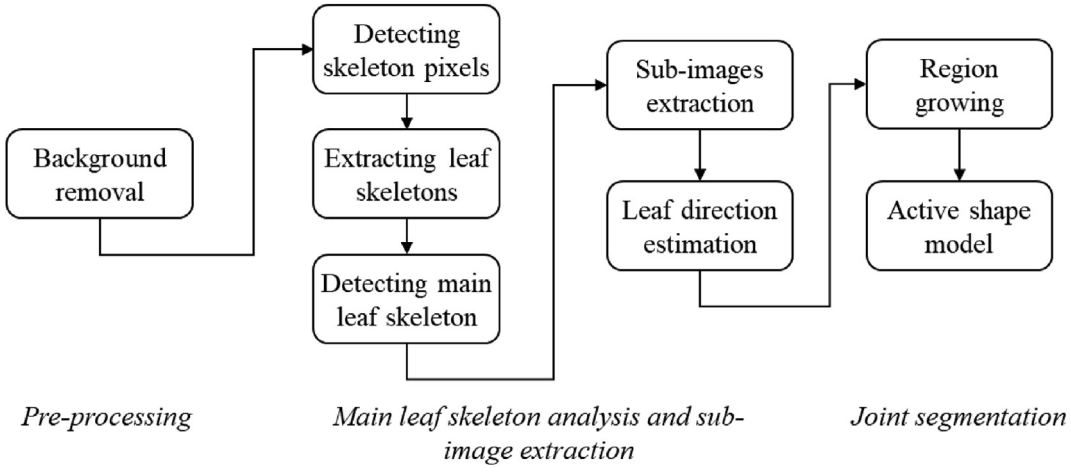

Fig. 3 – Overall procedure of proposed leaf shape-determination scheme.

Many leaf segmentation methods have been developed to eliminate background images in natural scenes, such as ExG (Meyer & Neto, 2008). The vegetation segmentation algorithm for plant images is beyond the scope of this study, and performance evaluation with other background removal methods should be conducted in extensive studies. The advantage of our method is that it extracts plant leaves rather than vegetation pixels. Small pieces of weeds or branches should be filtered out during this process. The combination of the vegetation segmentation method and our background removal method improved the stability of background removal.

3.3. *Leaf skeleton detection and sub-image extraction*

As shown in Fig. 1b, the leaf skeleton always presents brighter intensity values than the surrounding pixels on the leaf surface. Therefore, leaf skeleton detection was proposed according to colour and intensity characteristics. Initially, the image features of the leaf skeleton pixels were analysed by calculating the intensity histogram of skeleton pixels and their neighbouring pixels. To determine the intensity distribution pattern, 500 skeleton pixels were randomly selected from the plant image dataset. The intensity values of neighbouring pixels were collected from an $N*N$ ($N = 9$) window centred at the skeleton pixel (Fig. 5a). The intensity difference between the skeleton pixel and its neighbour pixels was calculated by subtraction. The average intensity difference between the skeleton pixel and neighbouring pixels is shown in Fig. 5b. Positive values indicate that the pixels showing lower intensity values and negative values are brighter than the sampled skeleton pixels. Thus, Fig. 5c shows that most of the neighbouring pixels have a lower intensity value than the skeleton pixel.

The skeleton pixels are brighter than the leaf surface, but the sampled skeleton pixels may not be the brightest pixels in the neighbouring area. The neighbouring skeleton pixels are contained in the local image. The proportion of brighter pixels in the local image of the skeleton pixel is analysed based on the intensity difference in the local image. The cumulative distribution of brighter pixels from the 500 skeleton pixel samples is shown in Fig. 5c. Only 27% of the skeleton pixels are the brightest pixels in the neighbouring image. In the local image of the skeleton pixel, the proportion of brighter pixels was 40% at most. In other words, any given pixel might be considered as a candidate skeleton pixel provided less than 40% of its neighbouring pixels are brighter than the given pixel. Based on the characteristic of the intensity distribution of skeleton pixels, a computationally inexpensive method is proposed to examine the candidate skeleton pixels from plant images.

(1) Detection of candidate skeleton pixels

The plant image is initially converted into a grey image on which the correlation filter is performed. For each pixel p in the plant image, the intensity difference of its neighbouring area is calculated, and the proportion of neighbouring pixels brighter than the given pixel p is obtained. If the proportion of brighter pixels is less than 40%, p is determined as a candidate skeleton pixel. The examined candidate skeleton pixels are shown in Fig. 6.

(2) Skeleton extraction

Leaf skeletons were effectively extracted from leaf images; however, numerous noise points were also detected in the process. Small dots can be removed by examining their size. Many of the noise points were connected, forming a large area (Fig. 6). In some cases, the noise points were connected to the

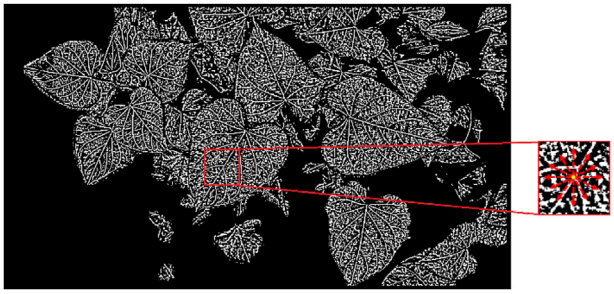

Fig. 6 – Candidate skeleton pixels.

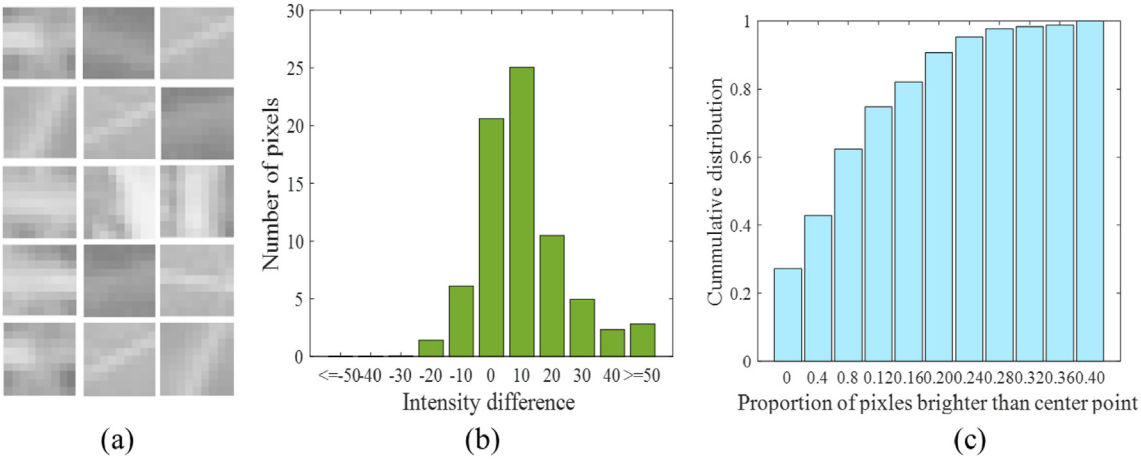

Fig. 5 – Intensity distribution of skeleton pixels, (a) sample images, (b) intensity difference, (c) proportion of brighter pixels.

skeleton. Therefore, a skeleton extraction method was proposed to extract the main leaf skeleton from the candidate pixels.

As the leaf skeleton always shows a line shape, the main leaf skeleton can be extracted by examining if the pixels lay on a line. The detailed steps are as follows:

(a) For every candidate pixel p, candidate skeleton pixels are searched for which run along straight lines originated at p in all directions (from 0° to 360°). The search strategy is illustrated in Fig. 6.
(b) If a straight line containing a candidate pixel can be found with a certain length (e.g. 48 pixels), pixel p and the pixels on the straight line are determined as skeleton pixels. Otherwise, pixel p should be removed as noise.

As shown in Fig. 7, only strong skeletons remained after this process. The length of the line containing candidate pixels was selected according to the size of the leaves in the plant image. A large leaf shows a long vein where a large threshold should be given to extract the vein. This value should be determined by preliminary tests or manual measurements of the maximal and minimal lengths of leaf veins.

(3) Retrieval of the main leaf skeleton

The leaf skeleton represents the structural information of a leaf. For leaf analysis, important information from the leaf skeleton is leaf direction. In this study, the main leaf skeleton was used to represent leaf direction. The main leaf skeleton was defined as the longest leaf vein. The main leaf skeleton was retrieved from the leaf skeletons (Fig. 8b). A straight-line fitting based on random sample consensus (RANSAC) was applied to extract the longest leaf skeleton. RANSAC is a robust method of parameter estimation that can estimate outliers from a dataset (Fischler & Bolles, 1981). In the leaf skeleton, the main leaf skeleton is considered to be the inliers, while the skeletons on both sides are outliers.

After skeleton pixel extraction as described in step (2), leaf skeletons are represented as connected components (Fig. 8a). In this process, it is possible to break the leaf skeletons into small pieces. Therefore, leaf skeletons were merged and the remaining small pieces were removed as noise. Every connected component in Fig. 6 was examined. If any two connected components have a minimal distance of less than six

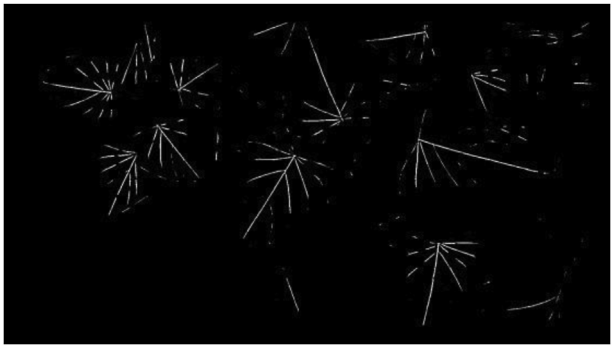

Fig. 7 – Extracted leaf skeletons.

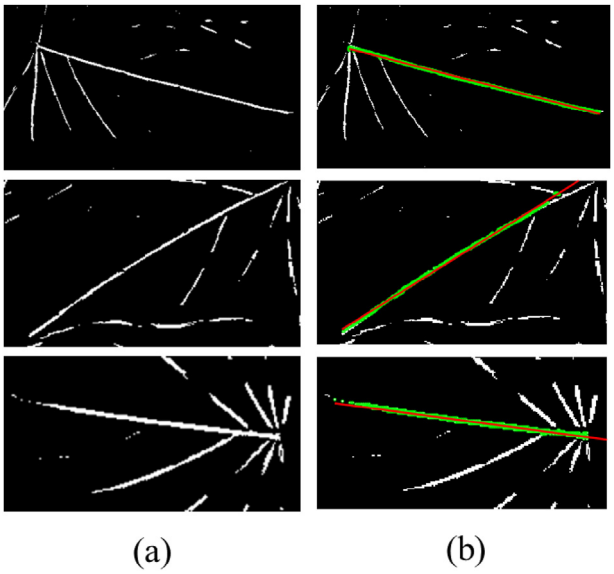

Fig. 8 – Main leaf skeleton detection. (a) leaf skeleton, (b) detected main skeleton (green pixels are the main skeleton and red lines are the fitting results). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

pixels, these two connected components are merged into a new connected component. Subsequently, RANSAC fitting was conducted to detect the main leaf skeleton in each connected component. The detailed process is described as follows:

1. A minimum number of pixels are randomly selected for a straight-line fitting.
2. The least squares method is conducted to solve the parameters of the straight-line function from those selected pixels.
3. The distances between every pixel on the skeleton and the fitted straight line are calculated. The number of pixels with a distance less than a predefined tolerance is determined (e.g. 3 pixels). These pixels can be considered as inliers.
4. If the number of inliers exceeds a predefined threshold τ, re-estimate the model parameters using all the identified inliers.
5. Steps 1 to 4 are repeated M times (e.g. $M = 500$). The fitting result with most inliers was determined as the final result.

The main skeletons detected are shown in Fig. 8b, where the slightly curved main skeletons were effectively detected in the skeletons. The red line represents the fitting results by RANSAC, and the green pixels are the original skeleton pixels (inliers).

(4) Sub-image extraction and estimation of leaf direction

Once the leaf main skeleton is extracted, the approximate position of the leaf can be acquired. The centre point of the

leaf main skeleton was assumed to be the leaf centre. The length of the sub-image was assigned according to the length of the main skeleton. Subsequently, sub-images containing a single plant leaf can be extracted from the whole plant image. As shown in Fig. 9a, one of the individual leaves shown in Fig. 1 was sampled. The length of the sub-image was given as 1.3 times the length of the main skeleton.

To represent the morphological features of the entire leaf skeleton, an adaptive threshold was applied to obtain the structure of the plant leaf (Fig. 9b). Although leaf skeletons are extracted by the previous process, the leaf skeletons are thin and cannot accurately describe the actual appearance of the leaf skeleton. For example, the leaf tip could not be identified using a thin skeleton. Therefore, the structure image of the main leaf skeleton was obtained to detect the leaf tip side. A bounding box of the partial main skeleton was calculated to sample the structure image. To reduce background noise, 3/4 of the length of the main skeleton from its centre was selected to examine the morphological features of the main skeleton (Fig. 9b), shown in Fig. 9c); the leaf tip side is thinner than the other side. Therefore, the leaf tip could be determined by comparing the size of the area of the two halves of the main skeleton.

Once the main skeleton and the leaf tip are determined, the direction of the individual leaves can be calculated from the main leaf skeleton. In this work, two end points of the partial main skeleton were adopted to calculate leaf direction (an angle ranging from $[0, 2\pi]$), as shown in Fig. 10. Because the main skeletons might be slightly curved in some leaves, especially on the tip side, calculating the leaf direction using the central part of the main skeleton should be more accurate.

3.4. *Estimation of leaf shape by joint segmentation*

After the sub-images of individual leaves are extracted from the plant image, it is possible to obtain detailed information on plant leaves based on local image analysis. Here, a fine leaf extraction strategy is presented that combines pixel-wise region growing segmentation and model-based segmentation to improve the accuracy of leaf contours definition, and to deal with occlusions.

Fig. 10 – Estimated leaf direction. The leaf direction is represented by the red arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Region growing is initially conducted to segment leaf images by examining the similarity of neighbouring pixels (Shih & Cheng, 2005). Seed points of region growing were obtained from the main leaf skeleton detected. Region growing is performed on the grey-value plant image. Figure 10a shows the segmentation results from Fig. 8a using region growing, which is a low-level image processing method that does not have any constraints to maintain object shapes.

As plant leaves generally overlap or occlude each other, region growing can extract all connected leaves showing a similar colour (Fig. 10a). Therefore, a deformable model-based segmentation method was introduced to detect and estimate the accurate leaf shape from the incomplete segmented results. The active shape model (ASM) is a widely applied deformable model for detecting elastic objects, which has been successfully applied to many fields, such as face detection and medical image analysis (Cootes et al., 1995; Hamarneh et al., 2004; Milborrow & Nicolls, 2008). The advantage of ASM is that it detects objects showing occluded portions and estimates the shape of such occluded parts. ASM also succeeded in segmenting plant leaves from a complex background (Xia et al., 2013). In ASM, the variation in leaf shape was modelled. With *a priori* knowledge of leaf shape,

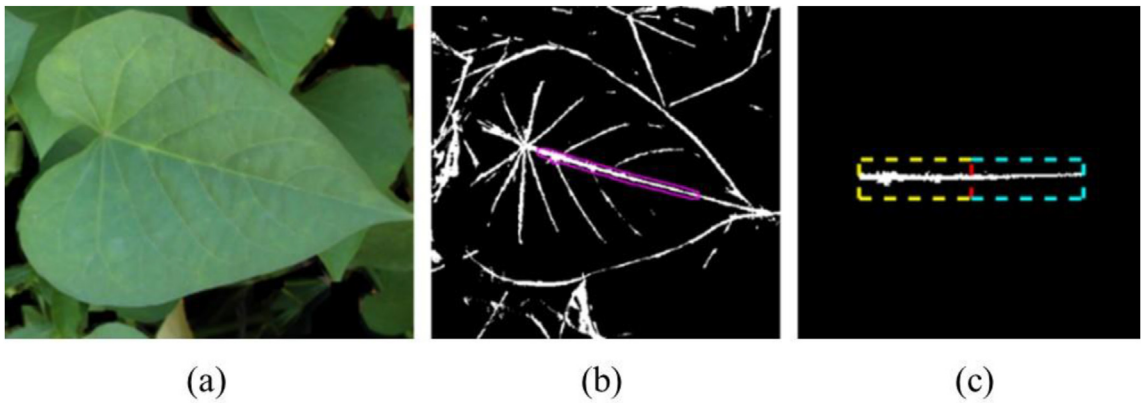

Fig. 9 – Determination of leaf tip. (a) sub-image of the leaf, (b) leaf structure image, (c) structure of main leaf skeleton.

ASM can estimate incomplete parts of a leaf image caused by occlusion or segmentation error. In ASM, a leaf shape (x_i) is represented by a group of boundary points called landmarks.

$$x_i = \left[x_{i,0}, y_{i,0}, x_{i,1}, y_{i,1}, \ \ldots\ , x_{i,n-1}, y_{i,n-1}\right]^T, i = 1, \ \ldots\ , N_{asm} \quad (2)$$

where ($x_{i,j}$, $y_{i,j}$) are the coordinates of the jth landmark on the ith shape in the training set. N_{asm} indicates the total number of leaf shapes in the training set. In this study, 36 leaves of sweet potato plants containing various leaf shapes were manually marked with 37 landmarks. After training, the pattern of variation in leaf shape was modelled to represent any leaf shape in the training set:

$$x_i = \overline{x} + Pb \quad (3)$$

where $p = [p_1, p_2, \ldots, p_t]$ is the matrix of the first t eigenvectors, and $\boldsymbol{b}$ is a weight vector. The shape variation was controlled by adjusting weight vector $\boldsymbol{b}$. In the matching process, ASM searches for the optimal boundary points at each landmark and estimates the best weight $\boldsymbol{b}$ with minimal matching errors. Finding optimal leaf boundary points is important for acquiring accurate segmentation of leaf shapes. Building accurate leaf boundary descriptors requires numerous sample data and sophisticated algorithms, which are usually computationally expensive. Region growing and ASM were combined to simplify the matching process. The leaf model matches individual leaf images on the segmentation results by region growing. The leaf model was initialised on the binary image of the segmented leaf according to the position and direction of the main leaf skeleton. The leaf model finds binary leaf boundary points and estimates the whole leaf shape from inaccurate segmentation by region growing. As shown in Fig. 11b, c the individual leaf boundary was accurately extracted by applying ASM to refine the segmentation results from region growth. In this study, the ASM was implemented based on the work by Hamarneh et al. (1998).

3.5. *Leaf detection using faster R–CNN*

To evaluate the performance of the proposed method, plant leaf detection was further implemented using a state-of-the-art object detection method, namely, Faster R–CNN (Ren et al., 2015). A Faster R–CNN consists of two parts: (1) fully convolutional region proposal networks for generating candidate regions, and (2) a downstream Faster R–CNN network. Both parts share full-image convolutional features, thus, enabling nearly cost-free region proposals. Here, all the images in Table 1 (190 images) were selected as training images. Plant leaves were manually labelled to train the Faster R–CNN model. These images were labelled manually in the PASCAL VOC 2007 dataset format. ResNet 101 (He et al., 2016) was adopted as the feature extractor of the faster R–CNN network for leaf detection. In this study, 50% of the plant images were randomly selected for training the network, and the trained faster R–CNN was compared with our proposed method.

4. Results

In this study, the detection of the main leaf skeleton the estimation of leaf direction were conducted by coarse segmentation of individual leaves. In turn, the extraction of the shape of individual leaves was performed by a joint segmentation approach. Understandably, the complexity of leaf images is significantly increased in multiple leaves and numerous leaves, because more extensively occluded leaves are present. In particular, for images with numerous leaves, an average of 21.4 leaves were shown by these images (Table 1). The leaves were relatively small, with various positions. Occlusions and overlapping of leaves are difficult to detect. The proportion of less occluded leaves is summarised for each type of leaf image to explain the complexity of the plant images. Low occlusion was defined as involving leaves having less than 1/4 of the whole leaf surface being occluded and the main leaf skeleton visible. To evaluate the performance of leaf skeleton detection, only low occlusion leaves were selected as detection targets. The detection accuracy of the main leaf skeleton is shown in Table 2. All main leaf skeletons were correctly detected in images containing a single leaf with either a non-green or a green background. The detection accuracy of the main skeleton was 98.42% for multiple-leaf

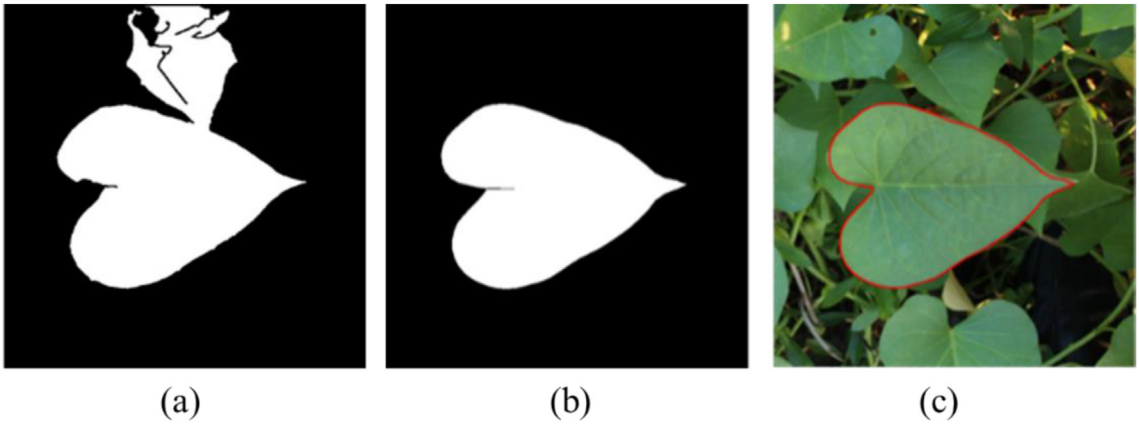

Fig. 11 – Individual leaf segmentation by region growing and ASM, (a) segmented leaf by region growing, (b) accurate segmentation by ASM and (c) ASM segmentation visualized on original image.

Table 2 – Detection accuracy of main leaf skeleton.

Image type	No. of target leaves	Correct detection	Missed leaves	Inaccurate detection	False alarms
Single leaf with non-green background	25	25 (100%)	0	0	0
Single leaf with green background	25	25 (100%)	0	0	0
Multiple leaves	317	312 (98.42%)	5 (1.58%)	9 (2.84%)	31 (9.78%)
Numerous leaves	1609	1526 (94.84%)	83 (5.16%)	114 (7.09%)	131 (8.14%)

images. In contrast, 94.84% of the target leaves were correctly detected from the images of numerous leaves. Inaccurate detection was reported to be 7.09% only for numerous leaf images. Here, inaccurate detection means that the leaf branch was incorrectly identified as the main leaf skeleton. And the detected leaf boundaries, severely occluded leaves, blurred leaves and weed leaves were considered as false positive results. False alarms were 9.78% and 8.14% for multiple leaves and numerous leaf images, respectively. A small proportion of leaf boundaries between overlapping leaves showed a similar appearance to that of the main skeleton. These leaf boundaries were possibly misrecognized as leaf skeletons. On the other hand, Table 2 shows that missed leaves were the main reason for the decrease in detection accuracy. Leaves missed by detection were 1.58% and 5.16% for images of multiple leaves and images for numerous leaves, respectively.

The estimation of leaf direction was subsequently evaluated by comparing the angle difference between estimated results and manual measurements of the main skeleton. Manual measurement of leaf direction was conducted by selecting two-thirds of 2/3 length of the main skeleton in its middle part and calculating the angle using its two end points. The straightest part of the main skeleton was selected for estimating leaf direction. The results of the evaluation of leaf direction estimation are summarised in Table 3. In this work, the correct estimation was defined as the estimated results with an angle difference of less than 10°. Otherwise, the result should be considered as an inaccurate estimation. The mean angle difference between the estimated leaf direction and human measurements was also calculated for the correct estimations and is shown in Table 3. Accurate estimation rates for all single-leaf images were 100% but decreased to 95.19% for multiple-leaf images and to 89.06% for numerous-leaf images. Directions of each leaf were all correctly measured for both single leaves with non-green and with green background. The mean estimation differences were 1.46° (±1.06°) and 1.39° (±0.92°) for single leaves with non-green background and single leaves with green background, respectively. Although the complexity was largely increased in multiple-leaf images and numerous-leaf images, the angle difference of the correctly detected leaves did not increase. The mean estimation differences were 1.31° ± 1.12° for multiple-leaf images and 1.42° ± 1.29° for numerous-leaf images. Therefore, the maximal estimation error of the leaf angle was less than 3°, which was accurate and stable. The results showed that inaccurate estimation was caused by incorrect detection of leaf tips, which produced a 180° error in leaf direction estimation. Inaccurate estimation rates for multiple-leaf and numerous-leaf images were 4.81% and 10.43%, respectively (see Table 4).

The accuracy of the proposed main leaf-skeleton scheme was subsequently compared with that of our previously reported method (Zhang et al., 2016). Three complex plant images from Zhang et al. (2016) were used to examine the accuracy of the two methods. Figure 12a shows 20 leaves on the soil background and Fig. 12b shows 21 sweet potato leaves with occlusions that overlapped with neighbouring leaves. As shown in Fig. 12c, 61 leaves of cabbage mustard were captured on the ground. The leaves in Fig. 12a,c were small and in various positions. Severe occlusions appear in Fig. 12b. Plant images in the first row in Fig. 12 were analysed by Zhang et al. (2016), while the results obtained by the new method proposed here are shown in the second row in Fig. 12. The numerical results obtained by Zhang et al. (2016) and those obtained by the proposed method are summarised in Table 2. The overall accuracy of detection and direction estimation using the proposed method was superior to that of Zhang et al. (2016). Especially for the image with numerous leaves in Fig. 12c, the detection accuracy (98.36%) and direction estimation accuracy (83.61%) by the proposed method were significantly superior, compared with the previous work in which the detection accuracy and direction estimation accuracy were only 65.57% and 54.10%, respectively. On average, the accuracy of leaf detection and direction estimation by the method were almost 20% higher than those obtained by the previous method. The method proved it can be more accurate and stable for the analysis of the leaves of diverse plant species.

The results of the comparison between the proposed method and faster R–CNN are summarised in Table 5. Both methods showed high accuracy in detecting leaves, with overall detection rates exceeding 80%. They both achieved

Table 3 – Estimation accuracy of leaf direction.

Image type	No. of detected leaves	Correct estimation	Mean angle difference (±STD)	Inaccurate estimation
Single leaf with non-green background	25	25 (100%)	1.46 ± 1.06	0
Single leaf with green background	25	25 (100%)	1.39 ± 0.92	0
Multiple leaves	312	297 (95.19%)	1.31 ± 1.12	15 (4.81%)
Numerous leaves	1526	1359 (89.06%)	1.42 ± 1.29	167 (10.43%)

Table 4 – Leaf detection accuracy of the proposed method and of the previous method by Zhang et al. (2016).

Figure	No. of target leaves	Methods	Correct leaf detection	Accuracy of leaf detection	Correct direction estimation	Accuracy of direction estimation
Fig. 12 (a)	26	Zhang 2016	20	76.92%	14	53.85%
		Proposed	24	92.31%	23	88.46%
Fig. 12 (b)	41	Zhang 2016	27	65.85%	18	43.90%
		Proposed	28	68.29%	26	63.41%
Fig. 12 (c)	61	Zhang 2016	40	65.57	33	54.10%
		Proposed	60	98.36%	51	83.61%

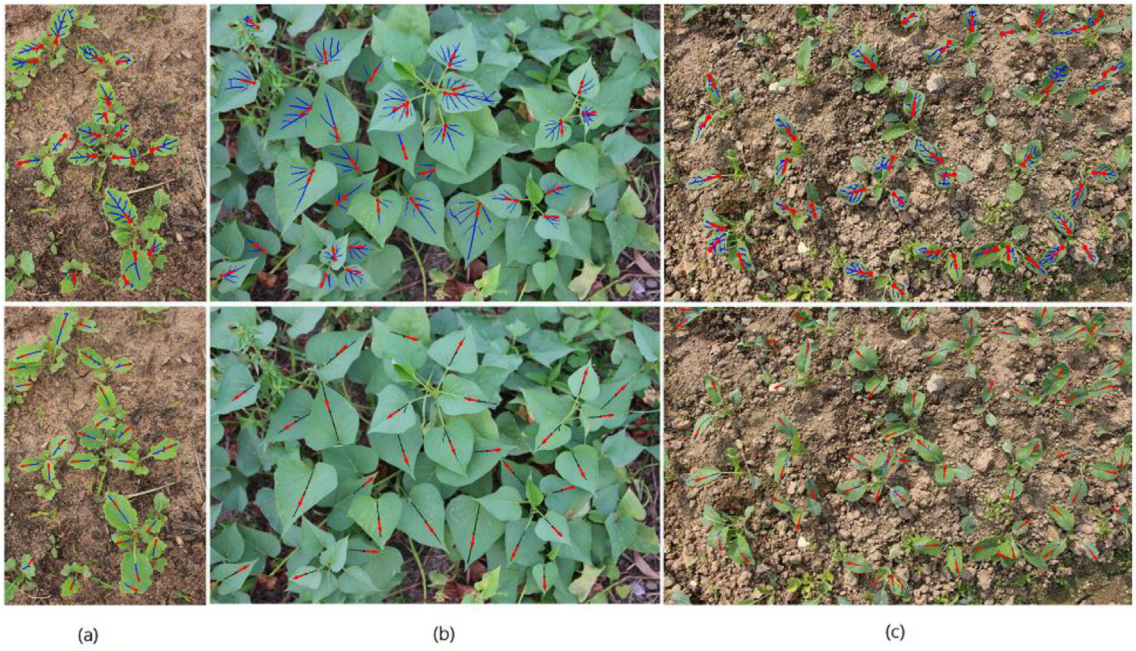

Fig. 12 – Comparison between the proposed method and the previous method reported by Zhang et al. (2016).

Table 5 – Comparison with faster R–CNN (50% training + 50% test).

Image type	No. of target leaves	Proposed		Faster R–CNN	
		Detected	False alarms	Detected	False alarms
Single leaf with non-green background	13	13 (100%)	0	13 (100%)	6 (46.15%)
Single leaf with green background	13	13 (100%)	0	13 (100%)	17 (130.77%)
Multiple leaves	170	149 (87.65%)	9 (5.29%)	162 (95.29%)	5 (2.94%)
Numerous leaves	921	747 (81.10%)	71 (7.71%)	799 (86.75%)	5 (0.54%)
In Total	1091	896 (82.13%)	80 (7.33%)	961 (88.08%)	33 (3.02%)

100% accuracy in single-leaf detection. In contrast, the detection rates for multiple leaves using the proposed method and Faster R–CNN were 87.65% and 95.29%, respectively. In multiple-leaf detection, our proposed method produced 5.29% false alarms (two leaves), while the false alarm rate was 2.94% for Faster R–CNN. As for numerous-leaf detection, the detection rates were 81.10% and 86.75%, while false alarms were 7.71% and 0.54% for the proposed method and Faster R–CNN, respectively. The overall detection rate of the proposed method was 82.13%, while Faster R–CNN showed a slightly higher accuracy, estimated at 88.08%. False alarm rate was 7.33% for our proposed method, and 3.02% for Faster R–CNN for all types of leaves. The extremely high false alarm rates of Faster R–CNN were found in the single leaf detection. In single leaf images, target leaves showed relatively large size. But small weed leaves, blurred leaves, irregular leaves or severely occluded leaves were occurred in the same image. These images were not considered as detection targets for both the proposed method and Faster R–CNN. And these unexpected leaves were not included for training Faster R–CNN. However, these unexpected leaves were still detected due the description ability of CNN features. The detection accuracy of our proposed method was comparable to that of Faster R–CNN. Although the accuracy of the proposed method was approximately 5.95% lower than that of Faster R–CNN, it actually detected more information on individual leaves, such

as leaf tip direction and skeleton structure. This information is important for applications requiring precision analysis, such as plant growth monitoring or leaf picking by robots.

Automatic individual leaf extraction was another contribution of this study. Based on individual leaf detection, leaf shape extraction was subsequently conducted using regional growing and deformable models. Plant leaves with occlusions in complex backgrounds were identified, and leaf contours were effectively extracted by the proposed method. An example of leaf shape extraction is presented in Fig. 13. Plant leaves of various sizes, directions, and occluded leaves were accurately detected by our method and Faster R–CNN. Detected leaves by Faster R–CNN were marked by the bounding boxes while our method was able to match the accurate leaf boundaries (Fig. 13). Only the results showing a regular leaf shape and containing at least 80% of the actual leaf contour were considered as effective results. Partially matched leaves should not be included in the final results; thus, for example, leaf No. 7 in Fig. 13 is a partially detected incorrect result. Leaf No. 2 in is an over-deformed result that failed to converge to the actual leaf contour and should also be removed from the final results. The overall accuracy of leaf shape extraction is shown in Table 6. Leaf extraction was evaluated based on the results of leaf skeleton detection. The leaf contour-detection rate was calculated according to the total number of detected main skeletons. Segmentation using region growing and combination segmentation of region growing and ASM were evaluated separately. For a single leaf with a non-green background, the correctly detected leaf contour was 23, with only region growing and 25 with joint segmentation. The detection rates for region growing and joint segmentation were 92% and 100%, respectively. The extracted leaf contours were reduced to 10 and 23, while detection rates were 40% and 92% for region growing and joint segmentation, respectively. Because the complexity increased in multiple-leaf images and numerous-leaf images, the accuracy of detection decreased significantly. The number of leaves detected in multiple-leaf images was 125 with region growing and 262 with joint segmentation. The detection rates of region growing and joint segmentation for multiple-leaf images were 40.19% and 84.24%, respectively. The number of detected leaves using region growing and joint segmentation in numerous-leaf images was 333 and 1032, respectively.

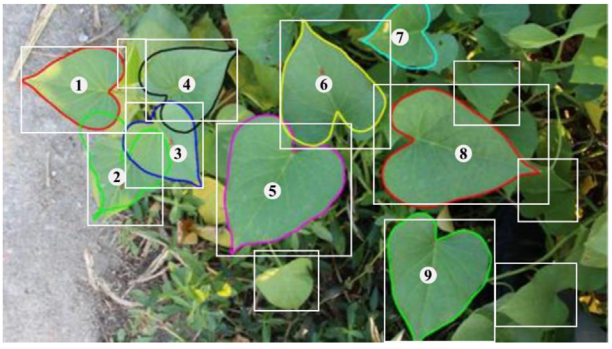

Fig. 13 – Shape determination for individual leaves (coloured shapes) and detection results by Faster R–CNN (white rectangles).

Table 6 – Accuracy of shape detection by region growing and joint segmentation.

Image type	No. of detected leaves	Region Growing	Joint segmentation
Single leaf with non-green background	25	23 (92.0%)	25 (100%)
Single leaf with green background	25	10 (40.0%)	23 (92.0%)
Multiple leaves	312	147 (47.12%)	282 (90.38%)
Numerous leaves	1526	341 (22.35%)	1159 (75.95%)

Conversely, the detection rates for numerous-leaf images decreased to 22.78% by region growth and to 70.59% by joint segmentation. Leaf shape extraction is a challenging task for analysing plant leaf images under field conditions. The accuracy of leaf extraction was excellent for single-leaf and multiple-leaf images. Our experimental results demonstrated that leaf shape extraction rates were greater than 80%, which is acceptable for many agricultural applications. Furthermore, it is understandable that the extraction rate was relatively low in numerous-leaf images, compared with the other types of plant images. Nevertheless, the extraction results clearly showed the advantages of the proposed method in such a complex scenario.

5. Discussion

Precise location of individual leaves in natural scenes is the key step for counting leaf populations, accurate individual leaf analysis, and other precision measurement tasks for developing intelligent agricultural systems. To implement automatic plant leaf analysis, a coarse-to-fine scheme was proposed to accurately detect leaf shape under field conditions, including the determination of the shape of occluded leaves. Leaf image analysis of sub-images of a single leaf may effectively reduce the complexity of leaf analysis. Therefore, coarse segmentation of individual leaves was developed by detecting the leaf main skeleton and estimating leaf direction. Sub-images containing individual leaves were extracted from the whole plant image according to the position of the main leaf skeleton. Fine-scale leaf shape determination was implemented on the sub-images of individual leaves by combining pixel-wise segmentation and deformable model-based object detection methods.

Accurate leaf shape detection was tested on four types of plant images in an agricultural field. The sample image sets contained various levels of complex plant images. The proposed coarse-to-fine leaf extraction scheme showed promising performance in accurately determining the leaf shapes from plant images under different conditions. The main leaf skeleton is considered as evidence of the presence of plant leaves. For all types of plant images, the overall accuracy of the leaf skeleton exceeded 90%. Although occlusions and background noise were severe in numerous-leaf images, the accuracy of detection of individual leaves was 94.84%, and false alarms occurred at a rate of 5.16% (Table 2). Detection accuracy was much better for all the other types of image

because the leaves were in much less complicated conditions. The main factors affecting the detection rate were missed detection and inaccurate detection. As the leaves were at various scales in this study, the main skeleton of small leaves could be removed as branches in the skeleton extraction process. Some of the main leaf skeletons were occluded and showed discontinuous lines while detecting candidate skeleton pixels. Consequently, it was difficult to detect these types of leaves using the proposed method. Inaccurate detection was the result of incorrectly recognising branches connected to the main leaf skeleton. Many inaccurate main skeleton detection results may provide an approximate position of the leaves and extract sub-images of individual leaves, but they should not be considered as correct leaf detection. Nevertheless, inaccurate detection may cause incorrect results while estimating leaf direction (Table 2).

The estimation of leaf direction showed human-level accuracy. The average measurement difference of correct detections in all plant images was 1.40°. The estimation of leaf direction was robust to noise because the RANSAC-based fitting was applied to determine the main skeleton. The results were seldom confused with noise or branches. In particular, when compared with our previous work, the accuracy in estimating leaf direction was significantly improved. In comparison, the proposed method was successfully applied on leaves of several different plant species. The results showed that the proposed leaf detection scheme is suitable for many other plant species because the main leaf skeleton exhibits similar image features. Under field conditions, leaf main skeletons show various visual features, and the proposed leaf-tip detection method should be improved to deal with variations in the main skeleton. Leaf tip detection contributed 7.62% of the estimation error. Therefore, it is possible to improve the accuracy of the estimation of leaf direction by developing a sophisticated leaf-tip detection method in the future.

To evaluate detection accuracy, the proposed method was compared with a deep learning approach, namely, Faster R–CNN. The performance of our proposed method was comparable to that of Faster R–CNN. The proposed leaf detection method is based on the presence of the main leaf vein. The leaves without clear leaf veins or blurred leaf images were the main factor reducing the detection accuracy of the proposed method (Fig. 13). The advantages of the proposed detection scheme are the following: (1) the direction of the leaf tip can be estimated during detection, (2) additional training process is not required, and (3) lower computational cost compared to convolutional networks. The convolutional neural network has demonstrated robust and reliable performance in various computer vision tasks. Recently, many leaf detection/segmentation methods have been developed based on convolution networks. Most CNNs are direction insensitive, which can locate objects in arbitrary directions but cannot measure the rotational angle of the leaves. Additional analyses should be conducted to determine the object direction. In this study, leaf direction could be directly estimated by identifying and analysing the main leaf veins. Further analysis of the leaf direction was not necessary. The proposed method can be applied to plant image analysis without training. In fact, a similar idea underlies both the proposed method and Faster R–CNN for detecting individual leaves. The proposed method predicts individual leaves from a feature of the local image, namely, the main leaf vein. The precise leaf shape was extracted using ASM. In Faster R–CNN, numerous-leaf candidates were generated according to the local CNN features (e.g. anchors) learned from the training sets. The candidates were further examined and verified using a classifier (e.g. SVM or SoftMax) (Ren et al., 2015). In some cases, thousands of candidates were proposed in an image, and those with high confidence were selected as the detection results. This process ensures that the Faster R–CNN achieves a robust detection performance. In this study, leaf image verification is not included in the proposed detection scheme. The hypothesis based on leaf veins was restricted to reduce false alarms; for example, the hypothesis could not be generated based on broken veins. Therefore, leaves showing broken veins have a high probability of being ignored. In the future, combining CNN features for leaf detection and verification of individual images should be considered to improve the accuracy and robustness of leaf detection. Furthermore, hypotheses based on leaf veins, leaf tips, and leaf boundaries should also be included in future work (Liang et al., 2015).

The accurate determination of leaf shape under field conditions is an additional contribution of this study. Regional growing was applied to segment individual leaves at the pixel level. The deformable leaf model built by the ASM was used to determine the precise shape of individual leaves. Regional growing is a low-level image segmentation method that segments images according to the similarity among pixels. Over-segmentation and under-segmentation frequently occur due to variations in illumination conditions, overlapping, or occlusions among leaves under field conditions. Therefore, the deformable leaf model was applied to accurately determine the leaf shape from the inaccurately segmented leaf images. Leaf shape variations were modelled as *a priori* knowledge by the ASM. With the deformable model, the whole leaf shape can be determined from incomplete segmentation by regional growth. The accuracy of leaf shape extraction was excellent in single-leaf and multiple-leaf images. Leaves in numerous-leaf images showed various irregular positions that might potentially reduce the robustness of leaf shape determination. Because irregular shape variations were not included in the training set, they were difficult to describe by the model. In cooperation with a robust estimation algorithm, such as expectation-maximization (EM), RANSAC can effectively improve the matching accuracy of ASM (Rogers & Graham, 2002; Santiago et al., 2015). In addition, integrating state-of-the-art feature descriptors, such as convolutional features, to find optimised leaf boundary features can improve the accuracy and stability of leaf shape determination (Jourabloo & Liu, 2016).

6. Conclusions

A leaf detection and an accurate shape determination scheme are reported herein. A leaf skeleton-extraction method was implemented and the main leaf skeleton was accurately detected by RANSAC fitting. The main leaf skeleton serves as evidence of the presence of an individual leaf. Sub-images of

single leaves were extracted for the complexity of plant images. Leaf direction was accurately calculated using the main skeleton. Accurate segmentation of individual leaves was performed by combining region growth and ASM. The proposed method could not only measure the individual shapes of numerous aggregated leaves in various poses but additionally, it can determine the entire shape of occluded leaves. The accuracy and stability of the proposed method were demonstrated by experiments on plant images under various conditions. Our results showed that the accuracy of leaf detection and leaf shape determination achieved by the proposed method were outstanding, especially for single-leaf and multiple-leaf images. The performance of the proposed method is comparable to that of state-of-the-art object detection methods (e.g. faster R–CNN). The proposed method can provide reliable leaf measurements and has great potential to implement automation systems for agricultural practises.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Key Area Research and Development Program of Guangdong Province under grant number 2019B020214002, by the CAS Key Technology Talent Program, and by the Key Research and Development Program of Yantai (2018YT06000808). This work was also supported by the China Scholarship Council (CSC) Grant and BK21PLUS, Creative Human Resource Development Program for IT Convergence.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biosystemseng.2021.03.017.

REFERENCES

Barré, P., Stöver, B. C., Müller, K. F., & Steinhage, V. (2017). LeafNet: A computer vision system for automatic plant species identification. *Ecological Informatics, 40*, 50–56.

Cerutti, G., Tougne, L., Vacavant, A., & Coquin, D. (2011). A parametric active polygon for leaf segmentation and shape estimation. In *International symposium on visual computing* (pp. 202–213). Berlin, Heidelberg: Springer.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image Understanding, 61*(1), 38–59.

De Vylder, J., Ochoa, D., Philips, W., Chaerle, L., & Van Der Straeten, D. (2011). Leaf segmentation and tracking using probabilistic parametric active contours. In *International conference on computer vision/computer graphics collaboration techniques and applications* (pp. 75–85). Berlin, Heidelberg: Springer.

Dobrescu, A., Valerio Giuffrida, M., & Tsaftaris, S. A. (2017). Leveraging multiple datasets for deep leaf counting. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 2072–2079).

Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosystems Engineering, 151*, 72–80.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395.

Giuffrida, M. V., Doerner, P., & Tsaftaris, S. A. (2018). Pheno-deep counter: A unified and versatile deep learning architecture for leaf counting. *The Plant Journal, 96*(4), 880–890.

Hamarneh, G., Abu-Gharbieh, R., & Gustavsson, T. (1998). Review e active shape models e part I: Modeling shape and gray level variation. In *Proceedings of Swedish symposium on image analysis*, Article 125e128 (Uppsala: Sweden).

Hamarneh, G., Abugharbieh, R., & McInerney, T. (2004). Medial profiles for modeling deformation and statistical analysis of shape and their use in medical image segmentation. *Image and Vision Computing, 22*(6), 461–470.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Horaisová, K., & Kukal, J. (2016). Leaf classification from binary image via artificial intelligence. *Biosystems Engineering, 142*, 83–100.

Jourabloo, A., & Liu, X. (2016). Large-pose face alignment via CNN-based dense 3D model fitting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4188–4196).

Kumar, J. P., & Dominic, S. (2020). Rosette plant segmentation with leaf count using orthogonal transform and deep convolutional neural network. *Machine Vision and Applications, 31*(1), 6.

Liang, X., Liu, S., Shen, X., Yang, J., Liu, L., & Dong, J. (2015). Deep human parsing with active template regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(12), 2402–2414.

Manh, A.-G., Rabatel, G., Assemat, L., & Aldon, M.-J. (2001). Weed leaf image segmentation by deformable templates. *Journal of Agricultural Engineering Research, 80*(2), 139–146.

Mata-Montero, E., & Carranza-Rojas, J. (2016). Automated plant species identification: Challenges and opportunities. In *IFIP world information Technology forum* (pp. 26–36). Cham: Springer.

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture, 63*(2), 282–293.

Milborrow, S., & Nicolls, F. (2008). Locating facial features with an extended active shape model. In *Proceedings of 10th European conference on computer vision* (pp. 504–514). Marseille, France: Springer.

Olsen, A., Han, S., Calvert, B., Ridd, P., & Kenny, O. (2015). In situ leaf classification using histograms of oriented gradients. In *Digital image computing: Techniques and applications (DICTA), 2015 international conference on* (pp. 1–8). IEEE.

Ota, T., Bontsema, J., Hayashi, S., Kubota, K., Van Henten, E. J., Van Os, E. A., & Ajiki, K. (2007). Development of a cucumber leaf picking device for greenhouse production. *Biosystems Engineering, 98*(4), 381–390.

Persson, M., & Åstrand, B. (2008). Classification of crops and weeds extracted by active shape models. *Biosystems Engineering, 100*(4), 484–497.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).

Rogers, M., & Graham, J. (2002). Robust active shape model search. In *European conference on computer vision* (pp. 517–530). Berlin, Heidelberg: Springer.

Santiago, C., Nascimento, J. C., & Marques, J. S. (2015). 2D segmentation using a robust active shape model with the EM algorithm. *IEEE Transactions on Image Processing, 24*(8), 2592–2601.

Shih, F. Y., & Cheng, S. (2005). Automatic seeded region growing for color image segmentation. *Image and Vision Computing, 23*(10), 877–886.

Slaughter, D. C., Giles, D. K., & Downey, D. (2008). Autonomous robotic weed control systems: A review. *Computers and Electronics in Agriculture, 61*(1), 63–78.

Teng, C. H., Kuo, Y. T., & Chen, Y. S. (2011). Leaf segmentation, classification, and three-dimensional recovery from a few images with close viewpoints. *Optical Engineering, 50*, 937–946.

Viaud, G., Loudet, O., & Cournède, P. H. (2017). Leaf segmentation and tracking in Arabidopsis thaliana combined to an organ-scale plant model for genotypic differentiation. *Frontiers of Plant Science, 7*, 2057.

Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering, 25*(2), 507–543.

Wang, J., He, J., Han, Y., Ouyang, C., & Li, D. (2013). An adaptive thresholding algorithm of field leaf image. *Computers and Electronics in Agriculture, 96*, 23–39.

Wang, X. F., Huang, D. S., Du, J. X., Xu, H., & Heutte, L. (2008). Classification of plant leaf images with complicated background. *Applied Mathematics and Computation, 205*(2), 916–926.

Xia, C., Lee, J. M., Li, Y., Song, Y. H., Chung, B. K., & Chon, T. S. (2013). Plant leaf detection using modified active shape models. *Biosystems Engineering, 116*(1), 23–35.

Xia, C., Wang, L., Chung, B. K., & Lee, J. M. (2015). In situ 3D segmentation of individual plant leaves using a RGB-D camera for agricultural automation. *Sensors, 15*(8), 20463–20479.

Zhang, L., Weckler, P., Wang, N., Xiao, D., & Chai, X. (2016). Individual leaf identification from horticultural crop images based on the leaf skeleton. *Computers and Electronics in Agriculture, 127*, 184–196.

Zhao, C., Chan, S. S., Cham, W. K., & Chu, L. M. (2015). Plant identification using leaf shapes—a pattern counting approach. *Pattern Recognition, 48*(10), 3203–3215.