Unsupervised segmentation of greenhouse plant images based on modified Latent Dirichlet Allocation

Yi Wang and Lihong Xu

College of Electronics and Information Engineering, Tongji University, Shanghai, China

ABSTRACT

Agricultural greenhouse plant images with complicated scenes are difficult to precisely manually label. The appearance of leaf disease spots and mosses increases the difficulty in plant segmentation. Considering these problems, this paper proposed a statistical image segmentation algorithm MSBS-LDA (Mean-shift Bandwidths Searching Latent Dirichlet Allocation), which can perform unsupervised segmentation of greenhouse plants. The main idea of the algorithm is to take advantage of the language model LDA (Latent Dirichlet Allocation) to deal with image segmentation based on the design of spatial documents. The maximum points of probability density function in image space are mapped as documents and Mean-shift is utilized to fulfill the word-document assignment. The proportion of the first major word in word frequency statistics determines the coordinate space bandwidth, and the spatial LDA segmentation procedure iteratively searches for optimal color space bandwidth in the light of the LUV distances between classes. In view of the fruits in plant segmentation result and the ever-changing illumination condition in greenhouses, an improved leaf segmentation method based on watershed is proposed to further segment the leaves. Experiment results show that the proposed methods can segment greenhouse plants and leaves in an unsupervised way and obtain a high segmentation accuracy together with an effective extraction of the fruit part.

Subjects Agricultural Science, Plant Science, Computational Science

Keywords Latent Dirichlet Allocation, Word-document assignment, Mean-shift, Optimal bandwidth search, Plant segmentation

Submitted 9 February 2018
Accepted 29 May 2018
Published 28 June 2018

Corresponding author
Lihong Xu, xulhk@163.com

Academic editor
Luiz Martinelli

Additional Information and Declarations can be found on page 27

DOI **10.7717/peerj.5036**

INTRODUCTION

Plant phenotype analysis based on image processing has been a popular application field of agricultural computer vision in recent years. An automatic, high-throughput, accurate and rapid imaging technique and processing method for plant phenotypic analysis will not only monitor the growth of plants, but also lay a visual foundation for the optimization of an intelligent greenhouse environment control system. It is a benefit to gene breeding, environment regulation, cultivation management, and the estimation of yield and quality. In the literature, the convolutional neural network (CNN) has been applied to agriculture, however, a large amount of reliable training data is needed which becomes an urgent problem. For the analysis of greenhouse plants, it is an important process to get enough

high-quality labelled images. In this regard, a well-segmented result of the plant and leaves can help in labelling the image quickly and accurately.

There are various imaging methods in the literature, among which a large amount of segmentation methods are developed from different aspects. Currently, there are several common methods to collect information of plants: (1) typical 2D color camera, (2) RGB-D camera, including stereo camera, ToF camera, structured light Kinect and laser scanner, (3) spectrometer, and (4) airborne laser radar. It is known that 2D imaging technologies are simple, low-cost and non-destructive, which can be suitable for plant phenotype analysis such as studying the growth of plant leaves, tracing their position and orientation, classifying new leaves from old ones, examining growth regulation and evaluating environmental stress if an accurate result of leaf segmentation is obtained (*Scharr et al., 2016*). The 2D imaging technologies lay a visual foundation for the analysis and warning of leaf disease as well.

For plant segmentation, the use of color in combination with depth images or multi-view images for supervised or unsupervised segmentation is a popular practice (*Alenya, Dellen & Torras, 2011*; *Song et al., 2007*; *Teng, Kuo & Chen, 2011*). Additionally, researchers have proposed some methods of segmenting plants from their background, in order to handle leaves with lesions (*Zou et al., 2015*; *Valliammal & Geethalakshmi, 2012*) and those without lesions (*Zhang, Zhang & Guo-Hong, 2017*; *Wang, Wang & Cui, 2011*). The background of the plants is artificial or relatively brief. One method for the detection of leaf disease spots is to directly extract the lesion part from the leaf of interest (*Ma et al., 2017*), while the other method is to separate a whole leaf from the complex background and extract the lesion part subsequently (*Fang, Lu & Lisi, 2014*). The latter can provide a better view to observe the site of lesions and analyze the severity of the disease. For single leaf segmentation, methods can be roughly categorized into four classes: shape model constraints (*Manh et al., 2001*; *Mezzo et al., 2007*; *Cerutti et al., 2011*; *Sogaard, 2005*; *Persson & Astrand, 2008*), boundary information detection (*Valliammal & Geethalakshmi, 2011*; *Noble & Brown, 2008*; *Tang et al., 2009*; *Lee & Slaughter, 2004*), depth information integration (*Guo & Xu, 2017*; *Alenya et al., 2013*; *Xia et al., 2015*; *Sanz-Cortiella et al., 2011*) and machine learning methods (*Zheng, Zhang & Wang, 2009*; *Meunkaewjinda et al., 2008*; *Hernández-Rabadán, Guerrero & Ramos-Quintana, 2012*; *Pape & Klukas, 2015*). Recently, the end-to-end recurrent neural network (RCNN) architecture (*Ren & Zemel, 2016*; *Romera-Paredes & Torr, 2016*), using an attention mechanism to model a human-like counting process, has provided a new way to handle occlusion by proceeding sequentially. Some of the methods above are mainly utilized in the classification of plants by extracting one complete leaf from a plant, while other methods segment all leaves in the image. Compared with the former, the latter methods have advantages in applications like genome-environment-phenotype synergistic analysis which cannot only rely on the information of one single leaf.

In recent years, more and more statistical methods have been applied in the image processing field. The Latent Dirichlet Allocation (LDA) model (*Blei, Ng & Jordan, 2003*) introduces a parameter θ which obeys Dirichlet distribution on the basis of Probabilistic Latent Semantic Analysis (pLSA) to establish the probability distribution of latent topic variable z. *Li & Perona (2005)* first applied the LDA model to image segmentation and many improved models came into view later. For example, a Spatial Latent Dirichlet

Allocation (SLDA) topic model (*Wang & Grimson, 2008*) encodes the spatial structure of visual words as a random hidden variable in a better way with LDA's generative procedure. A spatially coherent latent topic model (Spatial-LTM) (*Cao & Li, 2011*) provides a unified representation for a spatially coherent bag of word topic models, which can simultaneously segment and classify objects in the case of occlusion and multiple instances. *Russell et al. (2006)* partition a set of segmented objects into visual object classes using LDA and the visual object classes are further used to assess the accuracy of a segmentation. *Reddy, Singhal & Krishna (2014)* propose an algorithm that jointly infers the semantic class and motion labels of an object, integrating the semantic, geometric and optical flow based constraints into a Dense-CRF model.

In this paper, we first analyze the difficulties of applying a natural language processing model to image segmentation, and improve the spatial structure encoding of LDA through the word-document assignment strategy. Considering the problems of leaf disease spots and mosses, a spatial LDA image segmentation algorithm based on Mean-shift document bandwidths searching (MSBS-LDA) is proposed and applied to plant segmentation. In view of the problems with tomato fruits and the complicated illumination environment, an improved leaf segmentation method based on watershed is proposed to further segment the leaves. Experimental results show that the proposed methods can achieve a high accuracy of image segmentation.

LATENT DIRICHLET ALLOCATION (LDA)

The Latent Dirichlet Allocation (LDA) was first proposed by *Blei, Ng & Jordan (2003)*. It is a topic generation model which contains a three-level structure of word, topic and document (*Griffiths & Steyvers, 2004*; *David, 2010*). Both the topics of a document and the words of a topic obey polynomial distribution. It uses the Bag of Words (BoW) model, where each document is regarded as a word frequency vector, such that textual information is transformed into easy-to-model digital information. Therefore, LDA can be used to identify latent topic information in a large document collection or corpus. In recent years, LDA has been widely used in the field of machine vision as an unsupervised machine learning technology, such as target discovery, scene classification, behavior detection and visual surveillance (*Wang & Grimson, 2008*). However, spatial structures among visual words, which are vital in computer vision issues, are ignored in the language model since BoW forms a huge gap between low-level visual words and high-level semantics.

LDA gives the topic of each document in a document collection or corpus as a probability distribution. It is an unsupervised learning algorithm, which does not need manual annotation of the training set. Only a corpus and the number of topics are demanded. Note that a document can contain more than one topic, and each word in the document is generated by one of the topics. As depicted in the LDA Bayesian network structure (Fig. 1), the generative procedure of a document in LDA can be described as follows:

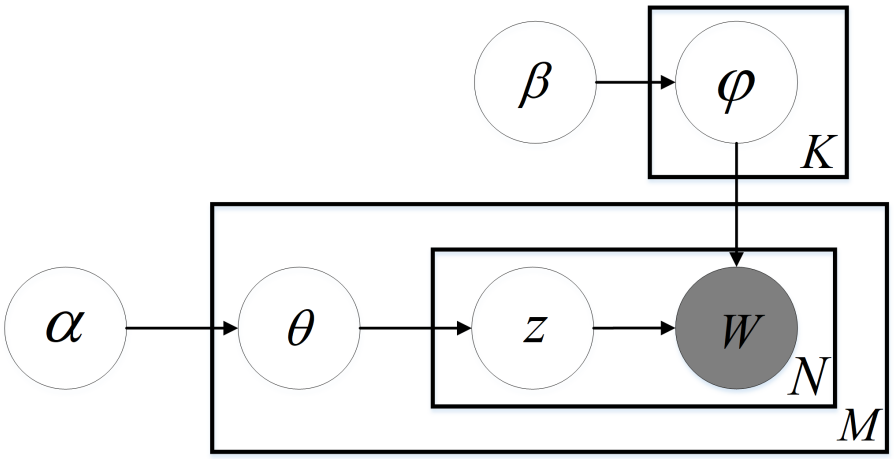

Figure 1 **LDA Bayesian network structure.**

Full-size DOI: 10.7717/peerj.5036/fig-1

(1) For a document i, its topic distribution θ_i is sampled from Dirichlet priori $Dir(\alpha)$;

(2) For a word j in document i, its topic label $z_{i,j}$ is sampled from topic multinomial distribution θ_i;

(3) For a topic $z_{i,j}$, its word distribution $\phi_{z_{i,j}}$ is sampled from Dirichlet priori $Dir(\beta)$;

(4) The value $w_{i,j}$ of word j in document i is sampled from word multinomial distribution $\phi_{z_{i,j}}$;

Thus, the joint distribution of all visible variables and latent variables in the model is:

$$p(w_i, z_i, \theta_i, \Phi|\alpha, \beta) = \prod_{j=1}^{N} p(\theta_i|\alpha)\, p\left(z_{i,j}|\theta_i\right) p(\Phi|\beta)\, p\left(\omega_{i,j}\left|\varphi_{z_{i,j}}\right.\right). \tag{1}$$

The maximum likelihood estimation for a document's word distribution can be obtained by integrating θ_i, Φ and summing z_i, as shown by:

$$p(w_i|\alpha, \beta) = \int_{\theta_i} \int_{\Phi} \sum_{z_i} p(w_i, z_i, \theta_i, \Phi|\alpha, \beta). \tag{2}$$

According to the maximum likelihood estimation of $p(w_i|\alpha, \beta)$, the parameters in the model can finally be estimated by Gibbs sampling.

SPATIAL LDA

When applied to image segmentation, the modified LDA needs to meet the following two requirements: (1) it should have a good description of local features and (2) it should be equipped with a good spatial summarization ability. The former concentrates on the construction of visual words while the latter focuses on the design of visual documents, making up for the defect of the BoW model. The two processes of local feature description and global structure summarization need to complement each other, leading to the improvement of spatial LDA algorithm.

Visual words

When we compute visual words, it should be noted that the detailed information of image feature is not the more the better. Excessively high dimension makes it difficult to achieve convergence when learning a dictionary in an unsupervised way, thus hindering LDA to extract latent topics. In this regard, moderate detailed features can be considered as the key to the ability of spatial LDA to describe local information.

Visual words are local descriptors that make up the whole content of an image. They mostly focus on local details instead of global structure or relationships among objects. In order to obtain the local descriptors, images are convolved with a filter bank which consists of three Gaussians, four Laplacian of Gaussians (LoG), and four first order derivatives of Gaussians. According to *Winn, Criminisi & Minka (2005)*, the filter bank has shown to achieve good performance for feature description and object categorization. The three Gaussian kernels (with $\sigma = 1, 2, 4$) are applied to each CIE L,a,b channel, thus producing nine filter responses. The four LoGs (with $\sigma = 1, 2, 4, 8$) are applied to the L channel only, thus producing four filter responses. The four derivatives of Gaussians are divided into the two x- and y-aligned sets, each with two different values of σ ($\sigma = 2, 4$). Moreover, these four derivatives of Gaussians are applied to the L channel only, thus producing four final filter responses. Totally, each pixel in the image has associated a 17-dimensional feature vector.

Afterwards, the image is divided into local patches on a grid and we densely sample a local descriptor for each patch. For the purpose of clustering all the local descriptors, K-means is employed to obtain N cluster centers $(c_1, c_2, \ldots, c_N)$ which form the visual dictionary $V = \{c_1, c_2, \ldots, c_N\}$. The local descriptors $G_1, G_2, \ldots, G_M$ are quantified into visual words according to the dictionary, with the minimum of Euclidean distance between each descriptor and each cluster center, as shown in Eq. (3).

$$w_j = argmin_{c_i} \left\| G_j - c_i \right\|^2, i \in [1, N], j \in [1, M]. \tag{3}$$

Thus the original image is converted into data consisting of visual words. In the next step, the visual words (image patches) will be assigned to visual documents (image regions) and further clustered into different classes (semantic objects).

Visual documents

After establishing the visual BoW, we need to further design the visual documents. The traditional LDA assumes that the corpus is a BoW model, which neglects objects and their spatial relationship. When we determine which visual document a pixel is assigned to, it actually contains some design of spatial information about the image. Therefore, we call the modified LDA with visual word-document assignment, spatial LDA. On the other hand, the design of visual documents, such as structure and size, generally imply specific visual assumptions on the image. Whether these assumptions are strong or not, it is directly related to the quality of image segmentation. For example, if we take the whole image as a document, it is assumed that if two patches belong to the same object, they often appear in the same image (Assumption 1). This assumption is reasonable but not rigorous.

In these regards, the design of visual documents should comply with the following two principles. (1) First, visual documents are required to depict spatial information

reasonably and focus on certain objects in the image space. A document expressing many different objects leads to the loss of important edge information, since image patches in a document tend to have the same topic label, while a document with few words does not have enough words to describe a certain topic. (2) Second, it is supposed that visual documents have a certain tolerance to feature differences in an object, in order to deal with a situation whose features vary widely within an object, like disease spots on leaves. The co-occurrence information of LDA, to some extent, has been able to classify different components, belonging to one object but with feature differences, into a topic. But it requires a large number of complete observation samples, which tends to make the algorithm time-consuming and accounting for storage, thus reducing the efficiency of spatial LDA. To alleviate this problem, we further enhance this co-occurrence information in the word-document assignment by blurring different parts of an object during the design of visual words, which can eliminate unnecessary weak edges and preserve critical strong edges.

Before discussing the unsupervised learning method Mean-shift Bandwidths Searching Latent Dirichlet Allocation (MSBS-LDA), we present two kinds of word-document assignment strategies that have some limitations on the above requirements.

Rectangular documents

If the document of a patch only includes other patches falling within its neighborhood, instead of the whole image, it corresponds to a better assumption: if two types of image patches are from the same object class, they are not only often in the same image but also close in space (Assumption 2). Hence, considering the tolerance to feature differences, two kinds of grid-based spatial information encoding methods are proposed.

Firstly, a rectangular region, which is taken as the document, is defined based on the grid with fixed size of $R_1 \times R_2$. The local patches in each region are quantized into visual words according to the dictionary, and subsequently, assigned to their documents. In the case that some patches are very close in space but assigned to different documents, an encoding scheme with overlapping rectangular regions is considered here. On the contrary, there is no overlapping region for another encoding scheme. That is the main difference between the proposed two encoding methods in this section. The detailed steps of image segmentation are as follows:

(1) The image is divided into local patches on a grid. Then local descriptors are calculated for each patch, and subsequently quantified into visual words $\{\mathbf{w}\}$ according to the vocabulary $\mathrm{V} = \{c_1, c_2, \ldots, c_N\}$;

(2) The image is divided into rectangular regions of $R_1 \times R_2$ size. Every two adjacent rectangular regions do not overlap (or overlap $(R_1 \times R_2)/2$), generating a number of visual documents $\{\mathbf{d}\}$;

(3) According to the visual word frequency histogram H^d, $\mathrm{d} = 1, \ldots, \mathrm{D}$, a corpus C is established for LDA;

(4) Suppose that the number of segments is K, a LDA model with K topics is trained based on the corpus C to obtain the probability that pixels of each region belong to K topics, namely $\varphi_i^d = \mathrm{P}(z|w_i, d)$. Note that, for the words of overlapping documents, the

probability φ_i is the average of φ_i^d obtained from all the corresponding regions, as shown in Eq. (4):

$$\varphi_i = \exp\left(\text{mean}\left(\sum_{d,i\in d} \log\varphi_i^d\right)\right) \tag{4}$$

(5) Each pixel is classified to a topic with the largest probability value of φ_i, thus all pixels of the image are partitioned into K classes.

With the two different encoding scheme of rectangular documents described in this section, two versions of spatial LDA segmentation algorithms are proposed, namely Non-overlapping Rectangular Documents index LDA (NR-LDA for short) and Overlapping Rectangular Documents index LDA (OR-LDA for short).

Super-pixel documents

Based on the space proximity assumption (i.e., Assumption 2), we attempt to design visual documents regarding the characteristics of the image itself. Since super-pixels combine pixels into meaningful atomic regions, they can be used to replace the rigid structure of the grid-based rectangular regions. Generally, the over-segmented super-pixel regions can retain most of the valid information for further image segmentation, and do not destroy the edge information of objects. Thus we further consider a super-pixel based spatial information encoding method in our research.

First, the basic information of the image is abstracted by super-pixels, thereby clustering a pixel-level image to a district-level map. These irregular pixel blocks, which are composed of adjacent pixels with similar texture, color, brightness and so on, are regarded as a corpus C. Accordingly, the corpus C contains a series of regions, each of which includes visual words corresponding to the pixels, forming a tree structure of spatial LDA. The subsequent image segmentation steps are described in steps (4) and (5) of section 'Rectangular documents' without the detail of the overlapping case. Here we use SLIC (Simple Linear Iterative Clustering) to implement the super-pixel encoding. Compared with other super-pixel segmentation methods, SLIC is superior in operating speed, super-pixel compactness and contour retention (*Achanta et al., 2012*).

For simplicity, the LDA segmentation algorithms based on super-pixel documents is named as Simple Linear Iterative Clustering Documents index LDA (SLIC-LDA for short).

Limitation analysis

By using rectangular regions and super-pixel regions, the word-document assignments usually have the following main limitations:

(1) In case of the rectangular documents, some parameters should be fixed such that they take the spatial structure of the image as a fixed and explicit variable. Although the documents of overlapping regions have some tolerance to the feature differences of an object, the rigid structure based on grid is separated from the characteristics of the image. Every document contains the same number of visual words and the size of all the documents is uniquely determined by the rectangular area. This kind of word-document assignment cannot describe the spatial information of objects in the image naturally.

(2) In the case of the super-pixel documents, they usually measure a cluster using the traditional distance, which is, however, unreasonable in some cases. Also, they replace the rigid structure of grid-based rectangular regions with some meaningful atomic regions, and accordingly, some unnecessary weak edge information of the object in the image may be introduced in addition to most valid edge information that has been retained. Besides, the over-segmentation result actually weakens the documents' tolerance to feature differences, which ought to make spatial LDA reduce the co-occurrence information of the same object in the word-document assignment.

Moreover, the design of rectangular documents and super-pixel documents can only classify visual words located within a certain distance into one document. If the object occupies more pixels and the distance between pixels is farther, it reflects the inaccuracy of Assumption 2. In view of the problems above, we propose a spatial LDA segmentation algorithm based on Mean-shift documents in the next section and compare traditional LDA to spatial LDAs with different designs of documents then.

MSBS-LDA

For the SLDA algorithm (*Wang & Grimson, 2008*), each document is represented by a point in the image, assuming that its region covers the whole image. If an image patch is close to a document, it has a high probability to be assigned to that document. To describe the word-document mechanism, several parameters are introduced as follows:

(1) d_j: a hidden variable indicates which document word j is assigned to;
(2) c_i^d: a hyper-parameter for document i known as a priori;
(3) g_i^d: the index of the image where document i is placed;
(4) $\left(x_i^d, y_i^d\right)$: the location of document j;
(5) $c_j = \left(g_j, x_j, y_j\right)$: storing location $\left(x_j, y_j\right)$ and image index g_j of word j;

In the generation procedure of SLDA, for a word j, a random variable d_j is sampled from prior $\mathrm{p}\left(d_j|\eta\right)$ indicating which document word j is assigned to and a uniform prior is used. The image index and location of word j is sampled from distribution $p\left(c_j|c_{d_j}^d, \sigma\right)$, and a Gaussian kernel is chosen:

$$p\left(\left(g_j, x_j, y_j\right)\middle|\left(g_{d_j}^d, x_{d_j}^d, y_{d_j}^d\right), \sigma\right) \propto \delta_{g_{d_j}^d}\left(g_j\right) \exp\left\{-\frac{\left(x_{d_j}^d - x_j\right)^2 + \left(y_{d_j}^d - y_j\right)^2}{\sigma^2}\right\} \tag{5}$$

where $p\left(c_j|c_{d_j}^d, \sigma\right) = 0$ if the word and the document are not in the same image. In the procedure of parameter estimation, z_j and d_j are sampled through a Gibbs sampling procedure integrating out ϕ_k and θ_i. The conditional distribution of z_j given d_j is the same as in LDA, which is given by:

$$p\left(z_j = k|d_j = i, d_{-j}, z_{-j}, w, \alpha, \beta\right) \propto \frac{n_{-j,w_j}^{(k)} + \beta_{w_j}}{\sum_{w=1}^{W}\left(n_{-j,w}^{(k)} + \beta_w\right)} \cdot \frac{n_{-j,k}^{(i)} + \alpha_k}{\sum_{k'=1}^{K}\left(n_{-j,k'}^{(i)} + \alpha_{k'}\right)} \tag{6}$$

The method attempts to borrow the language model to image segmentation, for which a uniform prior is applied to determine which document a word is assigned to and a

Gaussian kernel is adopted to describe the location of the word. Inspired by this idea, we consider the local maximum points of probability density function (P.D.F) of an image as documents, and the density estimation should be estimated with a nonparametric method due to the fact that the distribution of image data has no fixed pattern.

As a feature space analysis method, Mean-shift (*Comaniciu & Meer, 2002*; *Comaniciu, Ramesh & Meer, 2000*) attempts to find the local maximum points of P.D.F in a joint space, and applies nonparametric kernel density estimation (KDE) with smooth effect to density estimation, which provides a new way for word-document assignment of LDA. Thus we propose a modified LDA segmentation algorithm, namely Mean-shift Bandwidths Searching Latent Dirichlet Allocation (MSBS-LDA), for which Mean-shift is adopted to determine which document a word is assigned to. Moreover, the documents are represented by the modes of P.D.F and instead of applying Gibbs sampling to parameter estimation, it applies kernel density estimation to encode image information.

The pixels of each image include two types of information in coordinate space (x^s, $[px, py]$) and color space (x^r, $[l, u, v]$) such that the five-dimensional joint feature space $[px, py, l, u, v]$ is constituted. For a point x, Mean-shift iteratively searches its mode y in the joint space and assigns the color value of the mode to itself, that is $x^r = y^r$. If we set the footprint of x_i to $\{y_{i,0}, y_{i,1}, y_{i,2}, \ldots, y_{i,k}, \ldots, y_{i,c}\}$ then $y_{i,0} = x_i$ at the beginning and it converges to the mode $y_{i,c}$. The procedure of mode detection is as follow:

(1) Screen points close to $y^s_{i,k}$ in the coordinate space to the next step. Note that h_s is the kernel function bandwidth in coordinate space.

(2) Use the points survived to calculate the center of gravity and move towards it according to Eq. (7).

$$y^s_{i,k+1} = \frac{\sum_{n=1}^{N} x^s_n g\left(\left\|\frac{x^r_n - y^r_{i,k}}{h_r}\right\|^2\right)}{\sum_{n=1}^{N} g\left(\left\|\frac{x^r_n - y^r_{i,k}}{h_r}\right\|^2\right)} \tag{7}$$

where h_r is the kernel function bandwidth in color space.

(3) Decide if the stopping condition is met or if the number of iterations exceed the maximum limit. If so, stop the search and turn to next step, otherwise, return to step (1) and start from y_{k+1}. The stopping condition for search is determined by:

$$\begin{cases} y^s_{i,k+1} = y^s_{i,k} \\ \left\| y^r_{i,k+1} - y^r_{i,k} \right\| \le thr \end{cases} \tag{8}$$

(4) Assign the color $y^r_{i,c}$ of the mode $y_{i,c}$ to the starting point $x_i/y_{i,0}$, namely $x^r_i \leftarrow y^r_{i,c}$.

For density estimation at point x, the non-parametric density estimation is computed as:

$$\hat{f}(x) = \frac{1}{Nh^d}\sum_{n=1}^{N} K\left(\frac{x_n - x}{h}\right) = \frac{1}{Nh^d}\sum_{n=1}^{N} c_{k,d}\, k\left(\left\|\frac{x_n - x}{h}\right\|^2\right) \tag{9}$$

where $\mathrm{K}(x)$ is the kernel function, for which the radially symmetric function is used here such that $\mathrm{K}(x) = c_{k,d}\, k\left(\|x\|^2\right)$. Note that, $c_{k,d}$ is a normalization constant, making $\int_{R^d} K(x)\, dx = 1$, and $\mathrm{k}(x)$ is the profile for $\mathrm{K}(x)$. The Epanechnikov Kernel is adopted and

its profile is given by:

$$k(x)=\begin{cases}1-x & 0\le x\le 1\\ 0 & \text{otherwise}\end{cases}$$
$$K_E(x)=\begin{cases}\frac{1}{2}(d+2)\left(1-\|x\|^2\right) & \|x\|\le 1\\ 0 & \text{otherwise}\end{cases} \tag{10}$$

In addition, the density gradient estimation for point x is computed as:

$$\hat{\nabla} f_{h,K}(x)=\frac{2c_{k,d}}{Nh^{(d+2)}}\sum_{n=1}^{N}(x_n-x)\left[-k'\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)\right] \tag{11}$$

We know where to move and then the step length is determined by:

$$\begin{aligned}\hat{\nabla} f_{h,K}(x) &= \tfrac{2c_{k,d}}{Nh^{(d+2)}}\textstyle\sum_{n=1}^{N}(x_n-x)\left[-k'\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)\right]\\ &= \tfrac{2c_{k,d}}{Nh^{(d+2)}}\textstyle\sum_{n=1}^{N}(x_n-x)g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)\\ &= \tfrac{2c_{k,d}}{Nh^{(d+2)}}\left[\textstyle\sum_{n=1}^{N}x_n g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)-x\sum_{n=1}^{N}g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)\right]\\ &= \underbrace{\tfrac{2c_{k,d}}{h^2c_{g,d}}}_{C}\left[\underbrace{\tfrac{c_{g,d}}{Nh^d}\textstyle\sum_{n=1}^{N}g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)}_{\hat{f}_{h,G}(x)}\right]\left[\underbrace{\frac{\sum_{n=1}^{N}x_n g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)}{\sum_{n=1}^{N}g\left(\left\|\frac{(x_n-x)}{h}\right\|^2\right)}-x}_{m_{h,G}(x)}\right]\end{aligned} \tag{12}$$

where $\mathrm{m}_{h,G}(x)$ is the mean shift vector. It is a directional vector, which corresponds to the gradient direction. For a point x moving to x' in the gradient direction, the new coordinate is:

$$x'=x+m_{h,G}=\frac{\sum_{n=1}^{N}x_n g\left(\left\|\frac{x_n-x}{h}\right\|^2\right)}{\sum_{n=1}^{N}g\left(\left\|\frac{x_n-x}{h}\right\|^2\right)}. \tag{13}$$

As we can see from the above steps, the document assignment of word i is hugely affected by the bandwidths h_s and h_r of multivariate kernel density estimation. When we fix h_s and increase h_r monotonically within a certain range, the number of clusters decreases. A large h_r leads to inadequate documents (regions) while a small h_r results in excessive documents (regions). Similarly, the number of clusters decreases when h_r is fixed and h_s increases. In this regard, we propose to use the word frequency statistics and structural features of the image to find the required parameters (h_s, h_r). According to the meaning of the coordinate scale parameter h_s, we count the word frequency w_{fren} of the quantified visual words in accordance with the LDA dictionary. Then h_s is calculated on the basis of the proportion of the first major word by Eq. (14) where C is a constant in $[1.41, 1.67]$. For the color scale parameter h_r, we iteratively call the LDA segmentation procedure based on Mean-shift documents to obtain an optimal result. The mean value of LUV distance between the K classes is calculated by Eq. (15) and the last three $\Delta\bar{E}$ serve as a cut-off condition for the

iteration of LDA.

$$h_s = 10^C \times \frac{w_{fren}^{(1)}}{N} \tag{14}$$

$$\Delta\bar{E} = \frac{2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sqrt{\left(\bar{L}_i - \bar{L}_j\right)^2 + \left(\bar{U}_i - \bar{U}_j\right)^2 + \left(\bar{V}_i - \bar{V}_j\right)^2}}{K(K-1)}. \tag{15}$$

Under the framework of LDA, the number of topics K is determined manually. After setting h_s, we initialize h_r and decrease it with a step of one. In the beginning, Mean-shift obtains few clusters where the structure information is not obvious and changes a lot. Hence, LDA based on these documents leads to mis-segmentation, and the resulting $\Delta\bar{E}$ between the K classes is unstable. During the implementation process, h_r gradually decreases and the number of clusters increases, and under-segments change to over-segments as a result. Thereafter, the mis-segmentation situation of LDA is significantly improved, and $\Delta\bar{E}$ tends to stabilize after a big jump. If h_r is quite small, Mean-shift turns up excessive segments and the structural information is destroyed. Therefore, we aim to find a set of over-segmented Mean-shift documents, which should maintain certain structural description. The qualitative estimate of the value of h_r should fall in a middle range.

To obtain the final segmentation result, we need to consider the stopping criterion for the iteration of LDA. Note that, when under-segments change to over-segments, the Mean-shift documents have a great impact on LDA segmentation result. When h_r falls in a suitable middle range, the LDA segmentation result tends to be stable, and if h_r continues to decrease, the segmentation result is adversely effected. Therefore, we need to find the segmentation result immediately when the stable state is reached. We take it as the condition of convergence that the difference of three consecutive $\Delta\bar{E}$ is less than one. But sometimes there may occur special cases where two consecutive $\Delta\bar{E}$ are close and then spread out or the optimal range of the LDA segmentation result is limited. At this time, we take the segmentation corresponding to the first large jump of $\Delta\bar{E}$ as the optimal result to output. In addition, when h_r decreases within 10, the number of Mean-shift clusters increases rapidly, which leads to an "overleaping" on the number of spatial documents, thus affecting the judgement of convergence. Therefore, when h_r is less than 10, we change the step length to 0.5 to reduce the adverse effect of sudden changes in the document number on the segmentation results.

Based on the calculation above, the topic probability distribution for each pixel can be obtained by LDA given the word-document assignment. The generation process is as follows:

(1) For a document i, multinomial parameter θ_i for K topics is sampled from Dirichlet priori, $\theta_i \sim Dir(\alpha)$;

(2) For a topic k, multinomial parameter ϕ_k is sampled from Dirichlet priori, $\phi_k \sim Dir(\beta)$;

(3) For a word j, its document d_j is determined by the search of kernel function bandwidth $h = (h_s, h_r)$ of Mean-shift;

(4) For a word j, its topic label z_j is sampled from discrete distribution of document d_j, $z_j \sim Discrete\left(\theta_{d_j}\right)$;

(5) The value w_j of word j is sampled from discrete distribution, $w_j \sim Discrete\left(\phi_{z_j}\right)$.

Based on the analysis above, MSBS-LDA makes an improvement to Assumption 2 and proposes Assumption 3: If two image patches are from the same object class, they often appear in one image, and they belong to the pre-classified document in image space. That is to say, word i is likely to be assigned to document j if satisfying the following conditions: (1) they are in the same image, (2) word i belongs to the same document as its mode, and (3) the mode is found by Mean-shift combined with word frequency statistics and LDA iterative search. The assumption it adds is that a word tends to have the same topic label as other words in the pre-classified document of image space.

To overcome the limitations of two word-document assignments proposed in Sect. 'Visual documents', MSBS-LDA makes some breakthroughs in the two major difficulties of document design: image data modelling and document tolerance.

(1) Spatial information of the image is modelled by nonparametric estimation and the filtering effect of Mean-shift enhances the tolerance of the documents to feature differences, leading to a good global summary ability of the algorithm.

(2) Consider that the bandwidth parameters need to be manually adjusted for spatial documents, which is a non-trivial issue in case of real-world problems without domain knowledge. We propose to use word frequency statistics to determine the coordinate space bandwidth parameter h_s and iteratively search the optimal color space bandwidth parameter h_r in combination with the LDA segmentation algorithm of Mean-shift documents. The optimization of word-document assignment can reduce the time and resource consumption and greatly improve the accuracy of LDA segmentation result.

With the help of the Gaussian filter bank, we increase the diversity of words to ensure the local description ability of the LDA algorithm along with a rapid convergence of the unsupervised learning dictionary. The word-document assignment based on Mean-shift improves the image data modelling of documents and the tolerance to different features of the same object, thereby improving the global summary ability of the spatial LDA algorithm. Based on LDA's word frequency statistics and structural features of segmentation result, Mean-shift is guided to automatically find the optimal bandwidth parameters. The combination of the three processes improves the ability of LDA to solve image segmentation problems. Fig. 2 shows the flowchart of the MSBS-LDA algorithm.

PLANT SEGMENTATION AND LEAF SEGMENTATION

The segmentation of greenhouse plant images is divided into two processes: (1) plant extraction from background and (2) single leaf segmentation. The superiority of MSBS-LDA to describe local details in a complex scene and its tolerance to differences in features of the same object make it suitable for the problem of plant segmentation under the greenhouse scene, especially in the case of complex lighting, leaf lesions and mosses. Therefore, we first adopt MSBS-LDA to handle plant segmentation, extracting plants from the complex scene.

The illumination in greenhouses is ever-changing and the images captured during the fruits bearing period sometimes include tomato fruits of various growth stages, which have a negative impact on leaf segmentation. In these regards, we first compute color

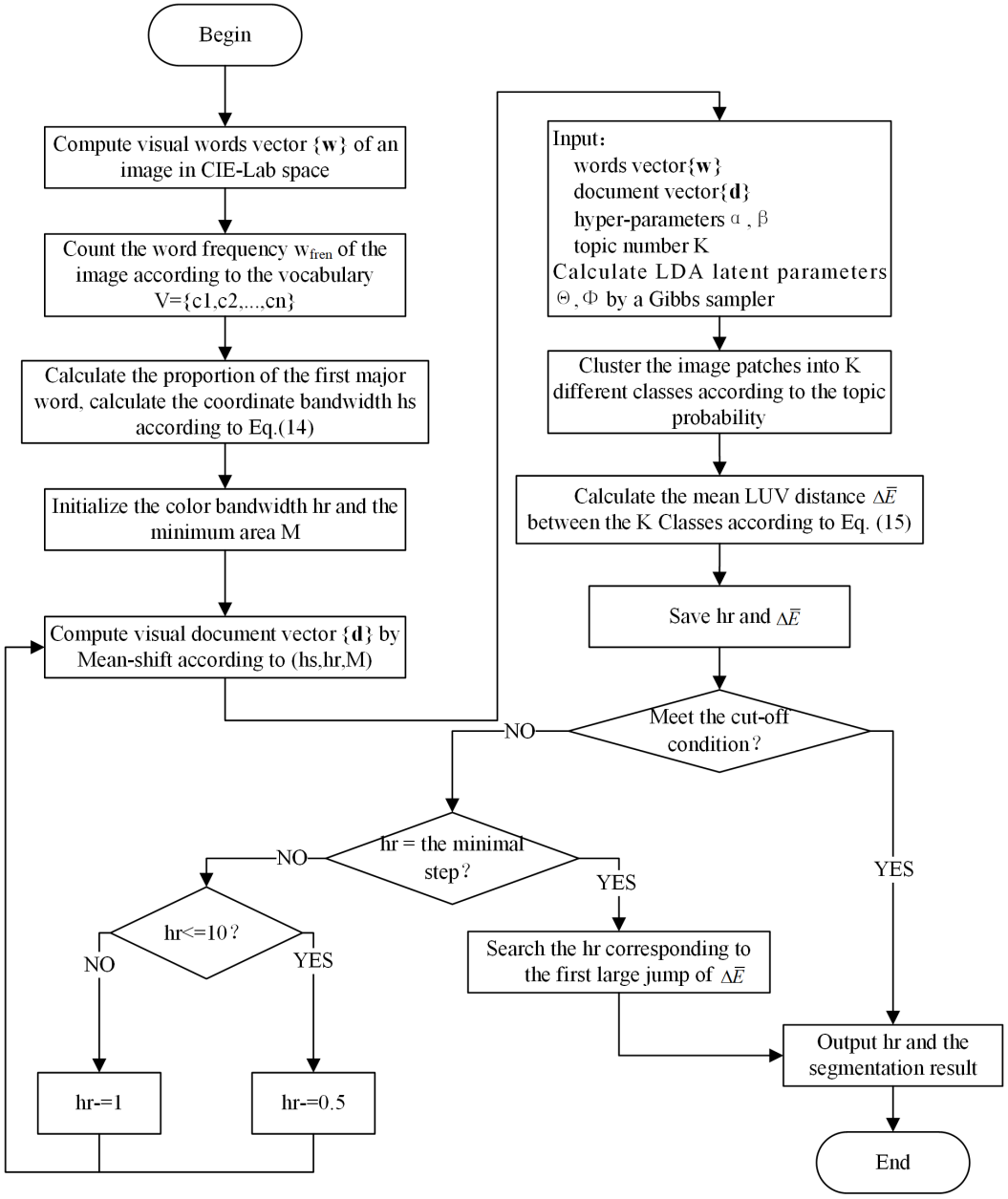

Figure 2 **The flowchart of MSBS-LDA segmentation algorithm.**
Full-size DOI: 10.7717/peerj.5036/fig-2

difference (R-B) on plant foreground with fruit and then perform threshold segmentation to remove the fruit part. Thereafter, the lighting environment is investigated based on the gray distribution of the leaf foreground histogram, as shown in Fig. 3. The grayscale distribution rate α' is calculated as:

$$\alpha' = \frac{C_{hist}(M_s - R_s, M_s + R_s)}{C_{hist}(1, 255)}, R_s = 255 \times \beta'/2 \quad (16)$$

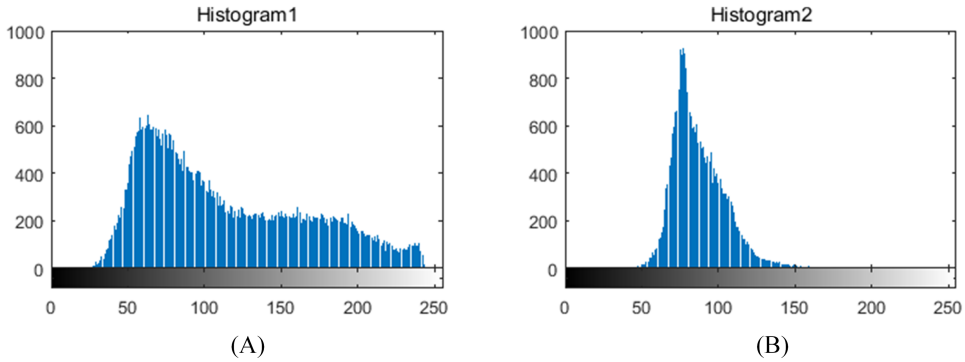

Figure 3 Grayscale histograms of foreground leaf images under different illumination conditions in greenhouses. (A) histogram of uneven illumination image, (B) histogram of even illumination image.
Full-size DOI: 10.7717/peerj.5036/fig-3

where C_{hist} is a grayscale statistical function, M_s is the maximum gray value of the accumulated grayscale, R_s is a grayscale radius, and β' represents a grayscale distribution range. We set β' and the threshold of α' as 0.2 and 0.8, respectively. Equation (16) denotes that if more than 80% of the pixels are distributed within 20% of the grayscale range, the lighting environment of the foreground leaf image is considered to be ideal, otherwise, illumination correction of homomorphic filtering is required to weaken its effect.

Then, leaf segmentation is carried out on the extracted leaf part after illumination response in three steps: strong edge detection, leaf centroid location, and marker-controlled watershed segmentation. Removal of strong edges from foreground leaf mask is a preprocessing step for leaf center detection, which can significantly improve the efficiency and accuracy of local maximum filtering. In the literature, some scholars have provided new ideas for closed contour effects (*Valliammal & Geethalakshmi, 2011*; *Sampath & Shan, 2007*; *Dollár & Zitnick, 2015*; *Ming, Li & He, 2012*). To decrease algorithm complexity and meet the demand of rapidity, the Structured Edge (SE) detector (*Dollár & Zitnick, 2015*) is applied to obtain leaf-leaf boundaries. Then we subtract the detected edges from the foreground leaf mask and compute a Euclidean distance map on it. Afterwards, the local maximum of the distance map is searched to locate each leaf's centroid based on the dilation operation. Further segmentation by watershed algorithm is conducted as the post-process step. The schematic diagram of the segmentation process is shown in Fig. 4 and the overall flowchart of the proposed plant segmentation and leaf segmentation method is shown in Fig. 5.

EXPERIMENTS AND ANALYSIS

In our research, we first test MSBS-LDA on the images from the Microsoft Research Cambridge Object Recognition Image Database v2 (MSRC-v2) (*Microsoft, 2000*) to evaluate its performance on the general dataset. For the segmentation of plant and leaf, all the tomato plant images were taken under real greenhouse conditions from three Venlo greenhouses of the Chongming Base of National Facility Agricultural Engineering Technology Research Center (abbreviated as CM), the Sunqiao Modern Agricultural Development Zone in

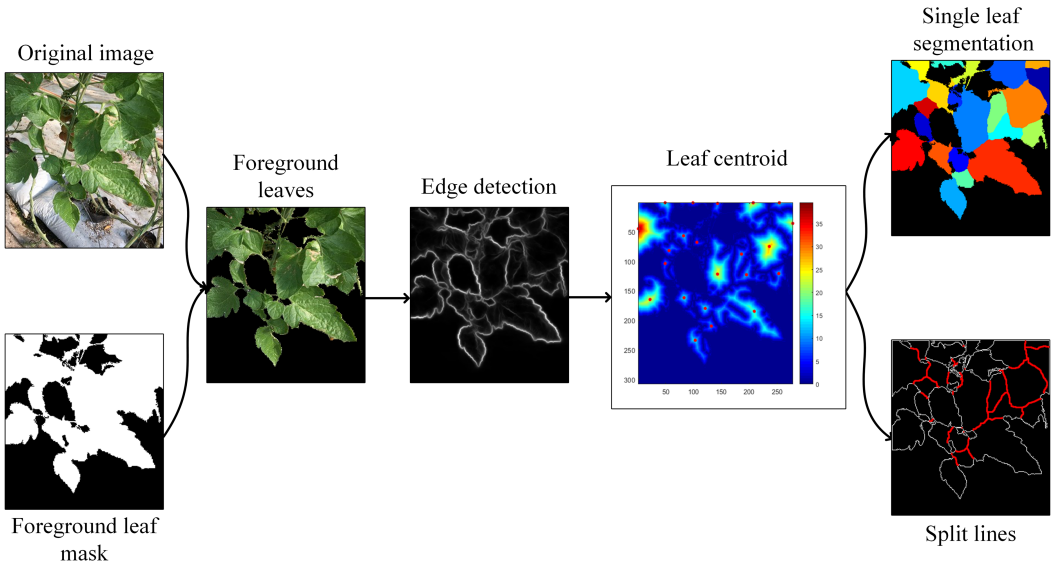

Figure 4 **The segmentation process.**

Full-size DOI: 10.7717/peerj.5036/fig-4

Shanghai (abbreviated as SQ), and the Jiading Experimental Greenhouse in Tongji University (abbreviated as JD). The main characteristic of the CM images is that they have uniform illumination and the plant is in vegetal stage with smooth leaves; while for the SQ images, the plant is in the stage of blossoming and bearing fruits, and the complicated light environment causes reflection as well as shadows on the leaves. The leaves with light or serious disease spots of the JD images increase the diversity of the plant images. It deserves pointing out that all the ground-truths and training set for comparison experiments were labelled manually by the first author. In addition to the tomato plant images, we took 13 images (among which, nine images contain obvious mosses) of Arabidopsis from the Computer Vision Problems in Plant Phenotyping (CVPPP) dataset A1 subset (*Scharr et al., 2014*; *Minervini, Abdelsamea & Tsaftaris, 2014*) to test the ability of MSBS-LDA to handle mosses. All experiments were conducted on a PC HP-g4-1059TX machine (Shanghai, China) with 2.10 GHz CPU and 6GB RAM.

Evaluation metrics

In order to quantitatively evaluate the accuracy of the spatial LDA segmentation algorithms including NR-LDA, OR-LDA, SLIC-LDA and MSBS-LDA, we define the following three metrics:

(1) SA (Segmentation Accuracy) measures the area of overlap between ground-truth and algorithm result:

$$SA = \frac{1}{K}\sum_{i=1}^{K}\frac{\left|P_i^{gt} \cap P_i^{ar}\right|}{\left|P_i^{gt}\right|} \tag{17}$$

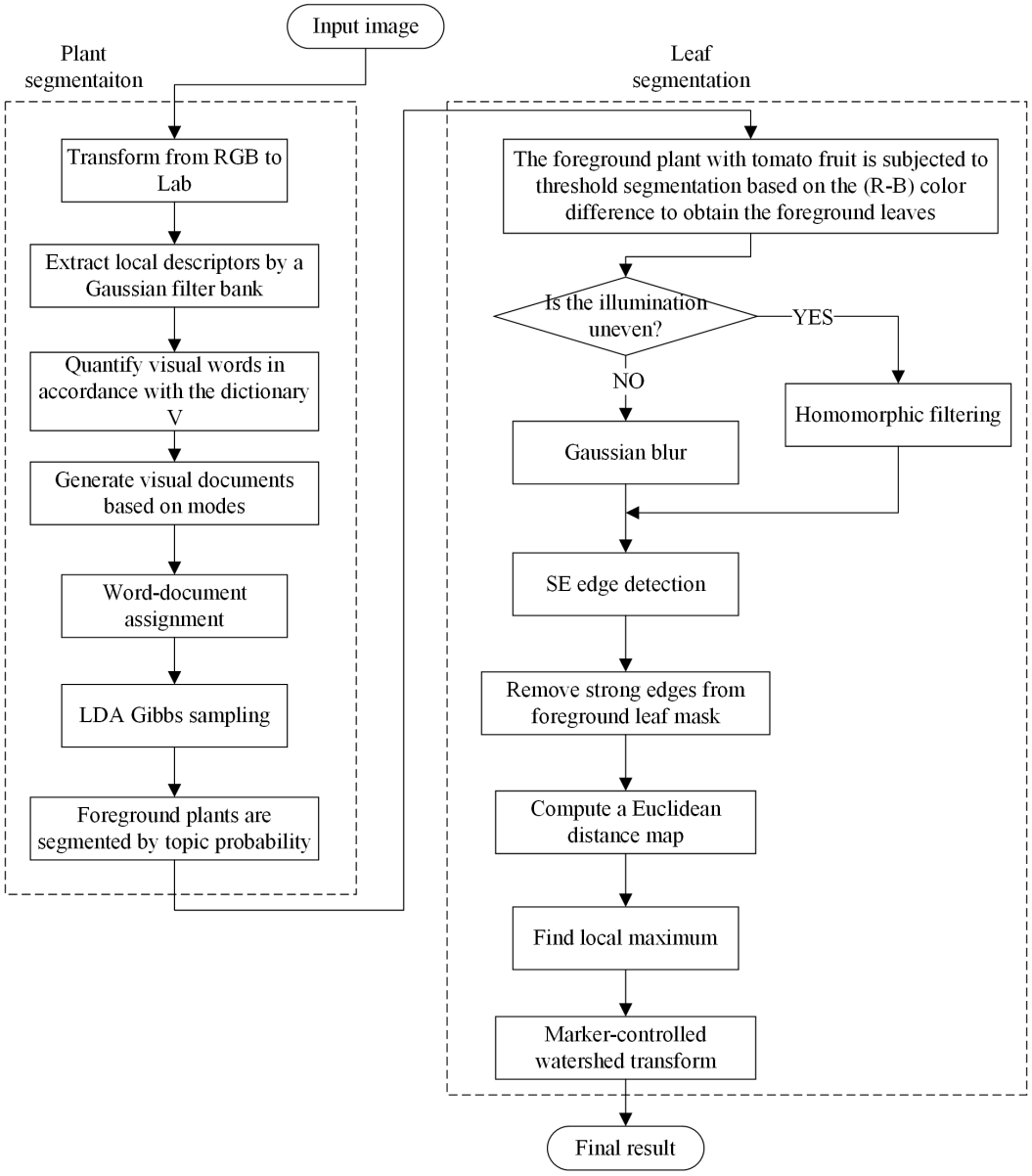

Figure 5 **The flowchart of the proposed segmentation method.**
Full-size DOI: 10.7717/peerj.5036/fig-5

(2) OR (Over-segmentation Rate) measures the area of algorithm result not in ground-truth:

$$OR = \frac{1}{K}\sum_{i=1}^{K}\frac{\left|P_i^{ar}\right| - \left|P_i^{gt} \cap P_i^{ar}\right|}{\left|P_i^{gt}\right|}\varepsilon\left(\left|P_i^{ar}\right| - \left|P_i^{gt} \cap P_i^{ar}\right|\right) \tag{18}$$

(3) UR (Under-segmentation Rate) measures the area of ground-truth not in algorithm result:

$$UR = \frac{1}{K}\sum_{i=1}^{K}\frac{\left|P_i^{gt}\right| - \left|P_i^{gt} \cap P_i^{ar}\right|}{\left|P_i^{gt}\right|}\varepsilon\left(\left|P_i^{gt}\right| - \left|P_i^{gt} \cap P_i^{ar}\right|\right) \tag{19}$$

where P_i^{gt} and P_i^{ar} denote the ground-truth and the algorithm result of the ith class, and $\varepsilon(\cdot)$ is the step function. All of these metrics are counted in pixels, with larger values of SA, smaller values of OR and UR representing higher agreement between ground-truth and algorithmic results.

The metrics above are suitable for the evaluation of general multi-class classifiers, but for plant and leaf segmentation, more targeted quantitative evaluation metrics are required. In this regard, some evaluation criteria proposed in *Scharr et al. (2016)* are adopted here, which are, respectively, defined as follows:

(1) FBD (Foreground-Background Dice) is Dice score of foreground plant:

$$Dice\,(\%) = \frac{2\left|P^{gt} \cap P^{ar}\right|}{\left|P^{gt}\right| + \left|P^{ar}\right|} \tag{20}$$

which measures the degree of overlap between ground-truth P^{gt} and segmentation result P^{ar}.

(2) SBD (Symmetric Best Dice) is symmetrical mean Dice of all leaves, as computed by:

$$SBD\left(L^{ar}, L^{gt}\right) = \min\left\{BD\left(L^{ar}, L^{gt}\right), BD\left(L^{gt}, L^{ar}\right)\right\} \tag{21}$$

where BD (Best Dice) is defined as:

$$BD\left(L^{a}, L^{b}\right) = \frac{1}{M}\sum_{i=1}^{M}\max_{1\le j\le N}\frac{2\left|L_i^a \cap L_j^b\right|}{\left|L_i^a\right| + \left|L_j^b\right|} \tag{22}$$

(3) Dic (Difference in Count) is difference in leaf number between ground-truth and algorithm result:

$$Dic = \#L^{ar} - \#L^{gt} \tag{23}$$

(4) |Dic| is absolute value of Dic,

(5) MHD (modified Hausdorff distance) measures accuracy of shapes and boundaries by comparing two sets of points, A and B, on the edge of a leaf:

$$MHD(A,B) = \max\{D(A,B), D(B,A)\} \tag{24}$$

with

$$D(A,B) = \frac{1}{|A|}\sum_{p\in A}\min_{q\in B}\|p-q\|. \tag{25}$$

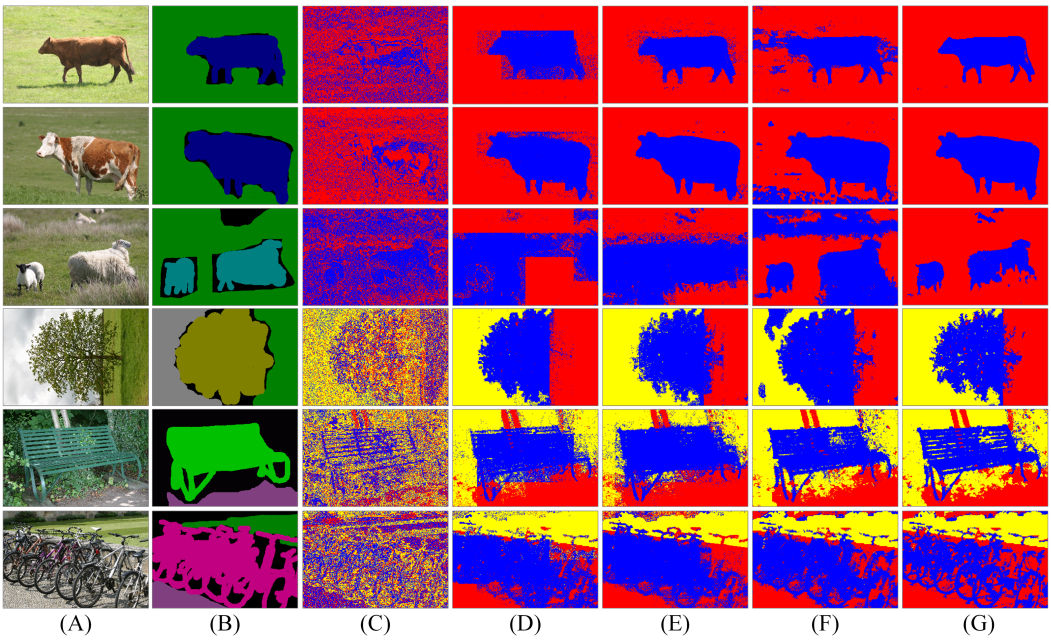

Figure 6 Segmentation results of spatial LDA algorithms with different word-document assignments. Given a collection of images as shown in the column (A), the goal is to segment images into different objects. The column (B) is weakly labelled ground-truth. The columns (C)–(G) are the results of LDA, NR-LDA, OR-LDA, SLIC-LDA and MSBS-LDA, respectively. The original images and their ground-truth credit: the Microsoft Research Cambridge Object Recognition Image Database v2 (MSRC-v2).
Full-size DOI: 10.7717/peerj.5036/fig-6

Spatial LDA evaluation

The experiment is carried out on the MSRC-v2 dataset to test the spatial LDA segmentation algorithms including NR-LDA, OR-LDA, SLIC-LDA and MSBS-LDA.

First, we investigate the influence of different word-document assignment strategies on the performance of spatial LDA. The segmentation results are shown in Figs. 6 and 7. We can figure out that the traditional LDA takes local descriptors as words and the whole image as one document, then it clusters the visual words that often appear in the same image into one object class, which exists two problems: (1) the segmentation result is noisy because spatial information is not considered and (2) since the whole image is treated as one document, if an object is dominant in the image, then other non-dominant objects could be labelled as the dominant one. NR-LDA improves the effect of noise, but documents of non-overlapping regions make rectangle edges noticeable. This kind of borderline situation is further improved by OR-LDA. Compared with spatial structure, SLIC-LDA more meticulously depicts the homogeneity of the atomic regions and the differences between them, so that weakens the tolerance of the documents to the different features in the same object. As a result, some unnecessary edges and noise in the segmentation results are generated. With the description of color and location information of the documents, MSBS-LDA tends to get better segmentation results in both details and global structure. The smooth effect of kernel density estimation and filtering effect of Mean-shift greatly reduce the noise of LDA caused by the BoW model. The results are no longer limited by

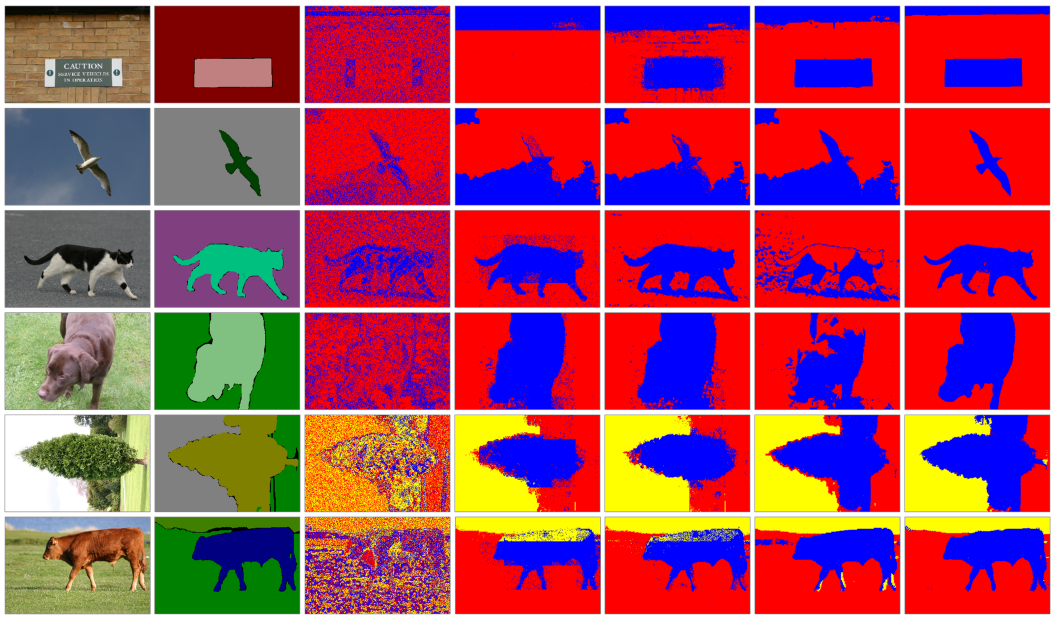

Figure 7 **Segmentation results of spatial LDA algorithms on images with high quality ground-truth.** The column (A) are original images. The column (B) is high quality ground-truth. The columns (C)–(G) are the results of LDA, NR-LDA, OR-LDA, SLIC-LDA and MSBS-LDA, respectively. The original images and their ground-truth credit: the Microsoft Research Cambridge Object Recognition Image Database v2 (MSRC-v2).

Full-size DOI: 10.7717/peerj.5036/fig-7

Table 1 **Segmentation evaluation of spatial LDA algorithms for images in Fig. 7.**

	SA	OR	UR
LDA	0.7483	0.7072	0.2517
NR-LDA	0.8582	0.7885	0.1419
OR-LDA	0.9509	0.7998	0.0491
SLIC-LDA	0.8976	0.7342	0.1024
MSBS-LDA	**0.9899**	**0.0787**	**0.0101**

the size of the rectangular documents, also, the shapes and boundaries are more natural. The evaluation results are shown in Table 1, where MSBS-LDA achieves the highest segmentation accuracy (0.9899), the lowest under-segmentation rate (0.0787) and the lowest over-segmentation rate (0.0101).

Afterwards, we compare MSBS-LDA with some unsupervised learning segmentation methods. The segmentation results are shown in Fig. 8. It can be seen that MSBS-LDA shows superior performance in general. Compared with 2D-OTSU and ICM, MSBS-LDA can significantly improve the noise situation, and it also has better tolerance to feature differences of the same object. Meanwhile, compared with Normalized-cut and Co-segmentation, the description of the details of MSBS-LDA has not been weakened by the tolerance.

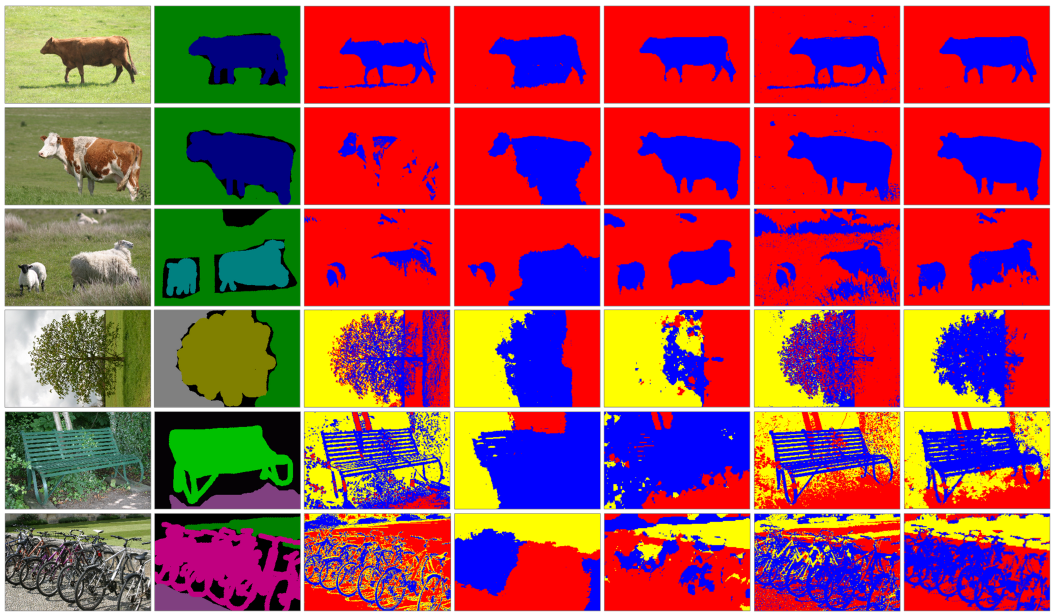

Figure 8 Segmentation results of five unsupervised algorithms. The column (A) are original images. The column (B) is weakly labelled ground-truth. The columns (C)–(G) are the results of 2D-OTSU, Normalized-cut, Co-segmentation, ICM and MSBS-LDA, respectively. The original images and their ground-truth credit: the Microsoft Research Cambridge Object Recognition Image Database v2 (MSRC-v2).

Full-size DOI: 10.7717/peerj.5036/fig-8

Plant segmentation evaluation

For three kinds of tomato plant images with distinct characteristics, we adopt MSBS-LDA to fulfill the plant segmentation task. Meanwhile, we test some other segmentation algorithms and features in comparison: (1) BP with 3D color and texture features proposed in *Minervini, Abdelsamea & Tsaftaris (2014)*, (2) OTSU with a* channel color of SLIC super-pixels, (3) co-segmentation, (4) dense-CRF based on texton-boost (*Krähenbühl & Koltun, 2011*; *Yang & Li, 2014*). The results of the five foreground segmentation methods on three tomato plant image sets are illustrated in Fig. 9, and the quantitative evaluation is shown in Tables 2 and 3. For Arabidopsis images from CVPPP A1 subset, the results of MSBS-LDA are illustrated in Fig. 10, and the quantitative evaluation is shown in Table 4.

According to the experimental results on tomato plant images, for the CM images, MSBS-LDA obtains the segmentation result comparable to BP and SLIC-OTSU, the stems and other details can be well separated from the background. The performance of MSBS-LDA is not affected by the complex background of the SQ image, and there is almost no noise in the result. Moreover, MSBS-LDA shows superior segmentation performance on the JD image which is most difficult for plant extraction. From Fig. 9 it is clear that not only the plant and the background can be effectively separated, but also the leaf disease spots can be accurately included in the foreground plant. When compared with the other four segmentation methods, MSBS-LDA achieves better segmentation accuracy, leaf shape and edge precision in general. In fact, through experiments we also find that gmentation

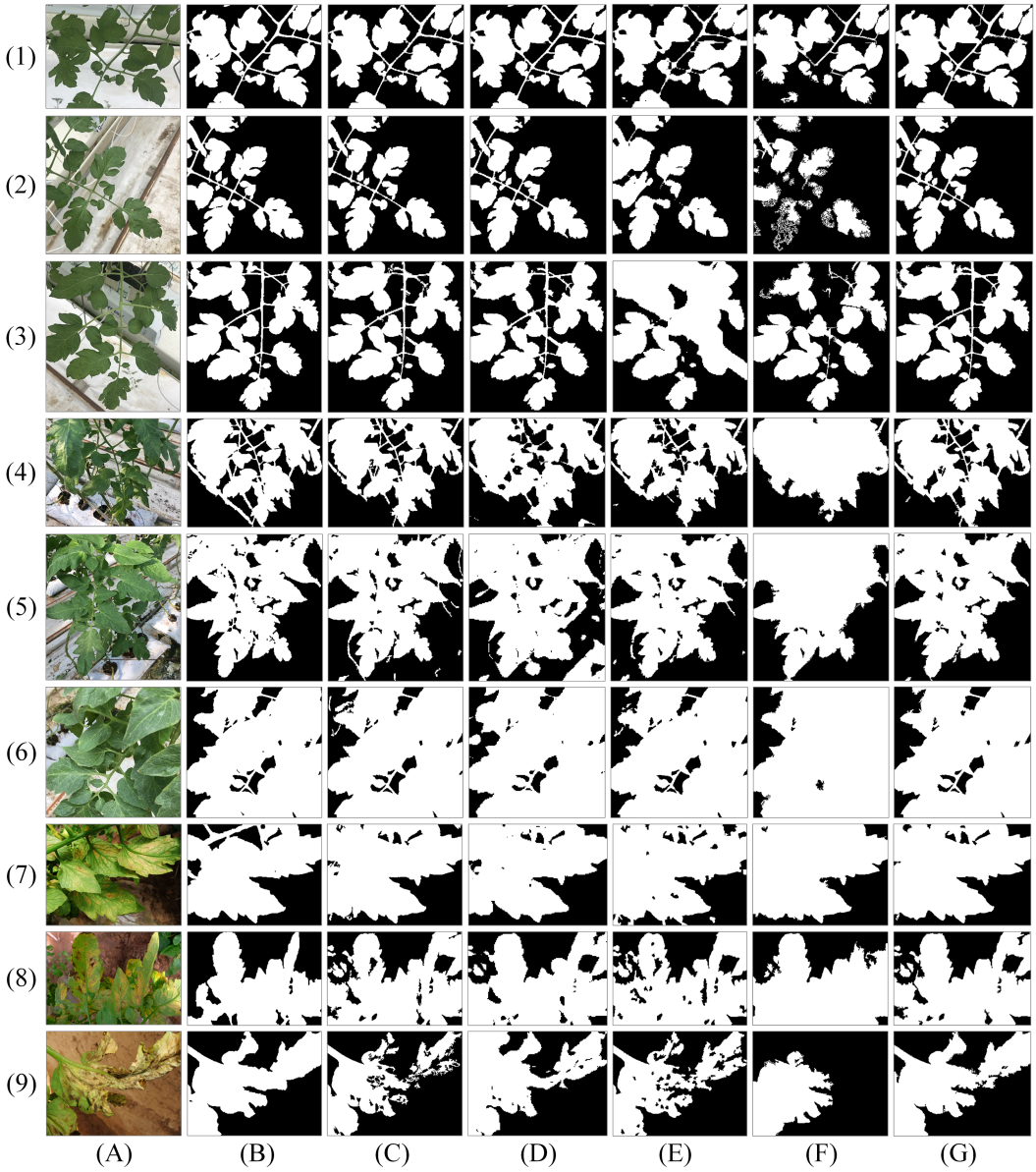

Figure 9 Results of five plant segmentation methods on three tomato plant image sets. The column (A) are original images. The column (B) are ground-truth images labelled by the first author. The columns (C)–(G) are the results of BP, SLIC-OTSU, Co-segmentation, Dense-CRF, and MSBS-LDA, respectively. Original images in the rows (1)–(3) are from Chongming Greenhouse, those in the rows (4)–(6) are from Sunqiao Greenhouse and in the rows (7)–(9) are from Jiading Experimental Greenhouse.

Full-size DOI: 10.7717/peerj.5036/fig-9

performance of some methods, such as BP and co-segmentation which rely on the training set or the same type of objects in a group, could be greatly affected if the characteristic between images are quite different or the design of the training set is not appropriate. In addition, from Table 5 we can assume that the time consumption of MSBS-LDA is competitive and practicable for most greenhouse applications. From the experimental results on CVPPP A1 subset, we can see that MSBS-LDA is able to separate mosses and

Table 2 FBD of five methods on three tomato plant image sets.

FBD (for plant, %)	BP	SLIC-OTSU	Co-segmentation	Dense-CRF	MSBS-LDA
CM1	97.90	**98.50**	93.00	92.76	98.08
CM2	**97.83**	97.65	89.44	80.65	97.54
CM3	97.31	97.12	82.04	90.47	**98.02**
SQ1	96.32	96.03	92.44	85.78	**96.42**
SQ2	97.07	96.93	88.99	92.63	**97.44**
SQ3	98.80	99.13	97.15	96.18	**99.18**
JD1	89.17	86.65	89.07	88.85	**90.78**
JD2	94.00	92.02	92.89	92.62	**94.31**
JD3	84.93	90.82	84.31	72.19	**96.54**

Table 3 MHD of five methods on three tomato plant image sets.

MHD (for plant)	BP	SLIC-OTSU	Co-segmentation	Dense-CRF	MSBS-LDA
CM1	0.71	**0.48**	5.92	7.04	1.48
CM2	0.46	0.72	1.89	3.14	**0.44**
CM3	**1.51**	2.51	6.90	6.41	2.03
SQ1	4.88	6.88	9.90	8.22	**3.51**
SQ2	6.43	4.17	14.79	5.05	**2.10**
SQ3	3.83	2.23	4.66	2.13	**0.45**
JD1	8.80	9.93	10.36	**8.12**	8.75
JD2	3.92	5.17	3.71	4.36	**3.02**
JD3	7.67	4.69	6.86	32.55	**2.35**

plants effectively by adjusting the topic number. The appearance of mosses does not affect the performance of the algorithm much.

Leaf segmentation evaluation

For the plant segmentation results containing fruit part, the separation results of leaves and fruits are shown in Fig. 11. The leaf segmentation results on three tomato plant image sets and the CVPPP A1 subset are shown in Figs. 12 and 13, respectively. The evaluation scores for all the testing images are listed in Table 6. For more complicated cases, like the JD images with lesions and the SQ images with both complex lighting environment and seriously overlapped leaves, our method reaches segmentation accuracy ranging from 50% to 60%. For the CM images with more uniform illumination and smooth leaves, the accuracy becomes more than 65%. The leaf segmentation scores of CVPPP A1 images are listed in Table 7. For Arabidopsis images with mosses and weak boundaries between overlapping leaves, the segmentation accuracy is up to 70%.

In this study, we find that if centroids of leaves are not detected correctly owing to serious overlapping, complicated lighting environment or side-effect of lesions, the performance of watershed algorithm will be limited. Besides, low contrast, weak boundaries and complex posture can lead to low segmentation accuracy. The method can also be fine-tuned to achieve a higher accuracy on one particular image, such as utilizing a more accurate

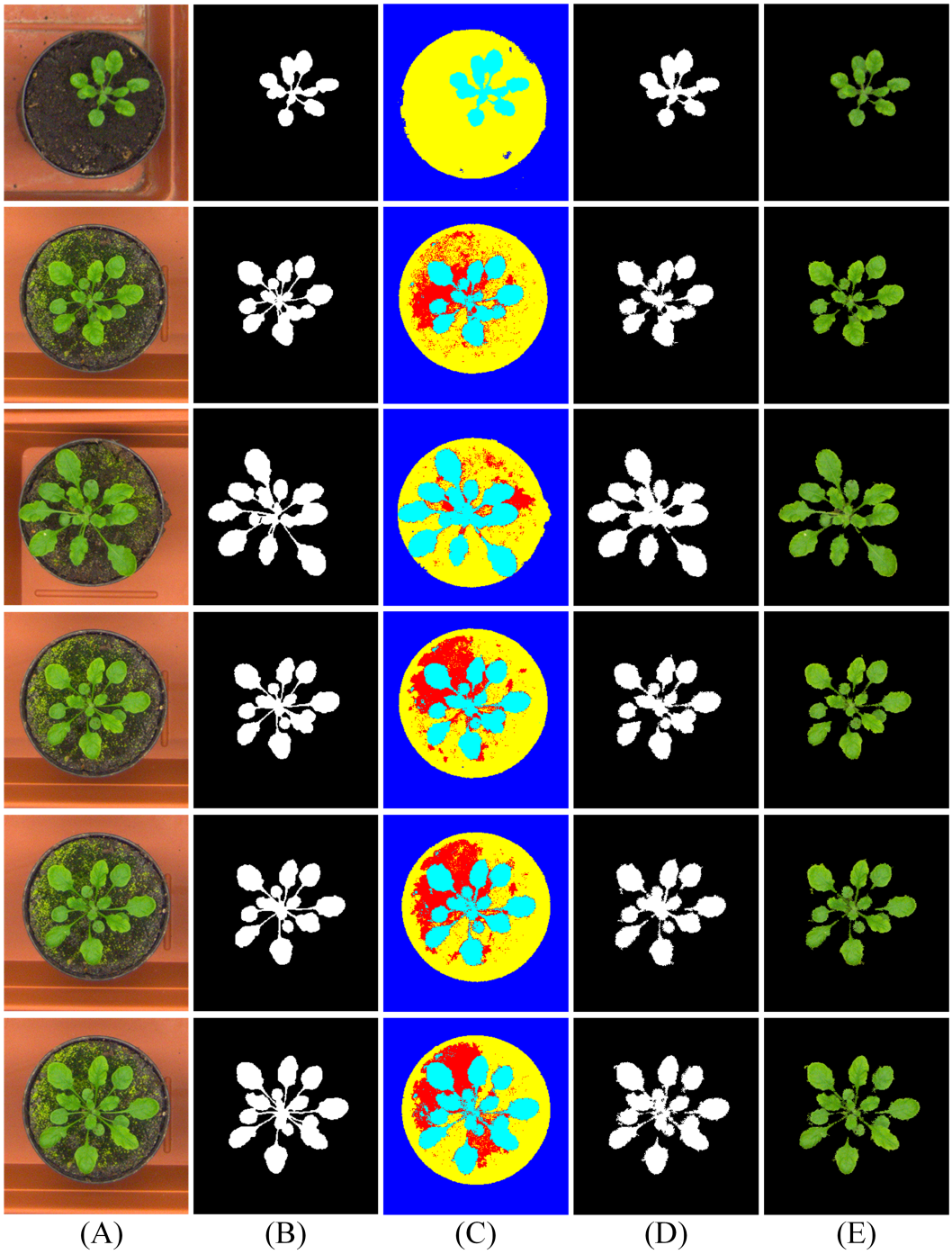

Figure 10 Results of MSBS-LDA on CVPPP A1 subset. The column (A) are original images. The column (B) is high quality ground-truth. The column (C) are the results of MSBS-LDA, the column (D) are the foreground plant masks and the column (E) are the extracted foreground plants. The original images and their ground-truth credit: the CVPPP dataset (A1).

Full-size DOI: 10.7717/peerj.5036/fig-10

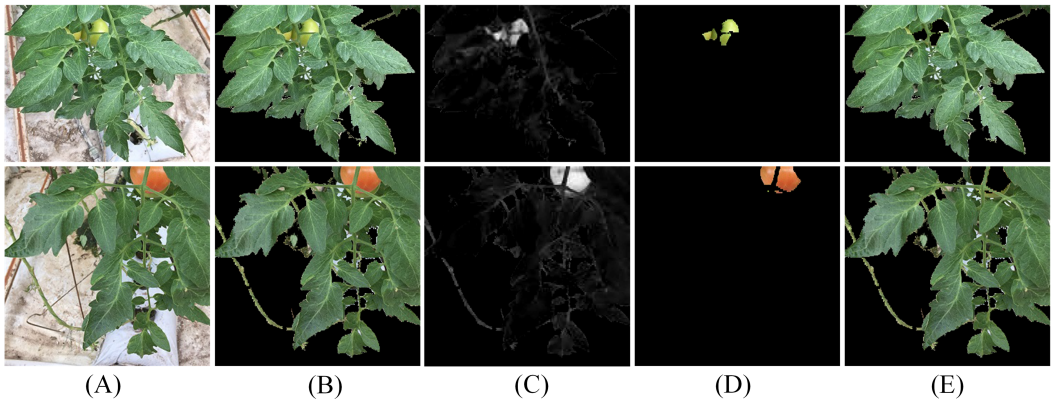

Figure 11 Results of leaf and fruit separation. The column (A) are original images containing fruit, the column (B) are the plant segmentation results, the column (C) are the R-B color differences, the column (D) are the extracted fruit part, and the column (E) are the extracted leaf part.
Full-size DOI: 10.7717/peerj.5036/fig-11

Table 4 Scores of MSBS-LDA on CVPPP A1 subset.

Samples	FBD (%)	MHD
13	95.99	0.51
59	95.69	0.62
91	95.70	0.98
96	95.55	0.70
115	95.24	0.70
153	94.54	0.88

Table 5 Time complexity of five methods for each image.

Methods	BP	SLIC-OTSU	Co-segmentation	Dense-CRF	MSBS-LDA
Time consumption (s)	22.51	5.51	213.14	166.90	141.64

edge detector, a more appropriate illumination normalization module or a better local maximum filter.

DISCUSSION AND CONCLUSION

In this study, we propose a modified statistical model of LDA, namely MSBS-LDA, to segment greenhouse tomato plants in an unsupervised way, and leaf segmentation is carried out subsequently. Through our experiments in different cases, some conclusions are drawn as follows:

(1) After analyzing the difficulties of using natural language processing model in image segmentation, it is proposed to improve the LDA algorithm in the spatial structure encoding of images through the word-document assignment. The diversity of visual vocabulary is guaranteed by constructing visual words, which can improve the ability of LDA to describe

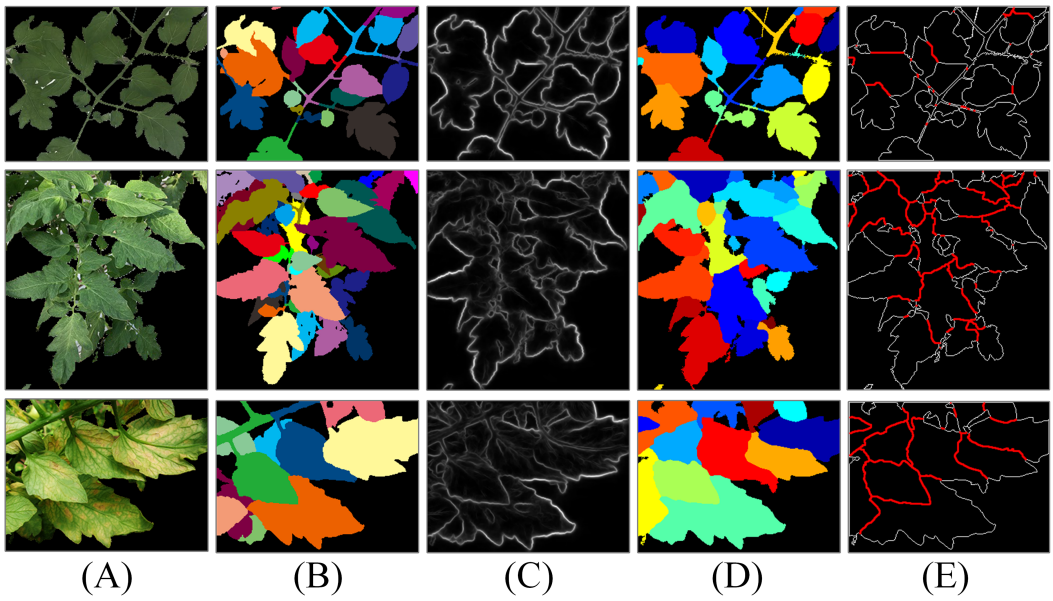

Figure 12 Results of leaf segmentation on three tomato plant image sets. The column (A) are original images, the column (B) are ground-truth images labelled by the first author, the column (C) are the SE detector results, the column (D) are the leaf segmentation results and the column (E) are the split lines.
Full-size DOI: 10.7717/peerj.5036/fig-12

image details. Moreover, the single document mapping is changed by designing visual documents, ameliorating the spatial summarization ability of LDA. The comparison experiments show that the spatial LDA algorithms with word-document assignments outperform the traditional one in image segmentation.

(2) In regards to leaf lesion spots and complicated backgrounds of greenhouses, a spatial LDA segmentation algorithm based on bandwidths searching of Mean-shift documents (MSBS-LDA) is proposed and applied to plant segmentation. The non-parametric estimation is adopted to space modelling, and the modes of Mean-shift are mapped as documents. The comparison experiments show that the proposed MSBS-LDA algorithm can simultaneously give good expression to image details and the global structure information, so that it can accurately include the lesion part in the plant and separate the plant from mosses of similar color through the adjustment of the topic number.

(3) In regards to tomato fruits and ever-changing illumination in greenhouses, an improved watershed segmentation method is proposed and applied to leaf segmentation. The fruit separation, illumination normalization, strong edge detection, leaf centroid location, and marker-controlled watershed segmentation are carried out sequentially to complete leaf segmentation. The experiments show that the method can guarantee a certain segmentation accuracy for three greenhouse tomato plant image sets of different characteristics.

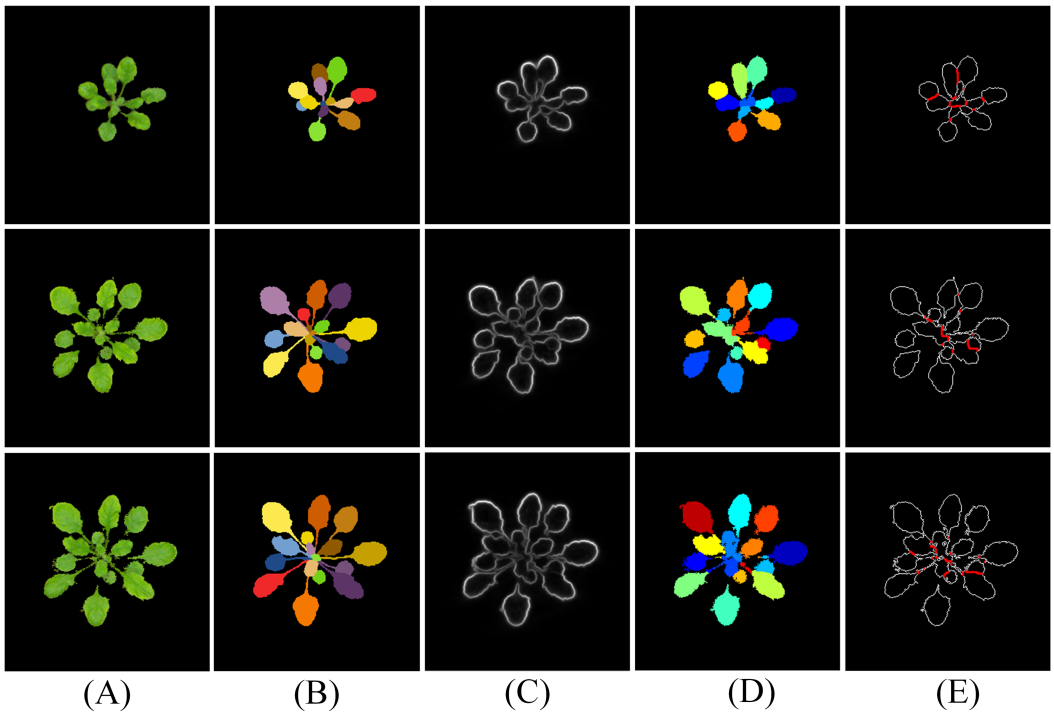

Figure 13 Results of leaf segmentation on CVPPP A1 subset. The column (A) are original images, the column (B) is high quality ground-truth, the column (C) are the SE detector results, the column (D) are the leaf segmentation results and the column (E) are the split lines. The original images and their ground-truth credit: the CVPPP dataset (A1).

Full-size DOI: 10.7717/peerj.5036/fig-13

Table 6 Leaf segmentation evaluation on three tomato plant image sets.

Image sets	FBD (for leaves, %)	SBD (%)	|DiC |	DiC
CM	96.77 (1.83)	65.37 (6.25)	1.80 (1.32)	−1.00 (2.05)
SQ	96.75 (1.71)	55.86 (4.00)	2.60 (2.17)	−2.60 (2.17)
JD	93.73 (2.79)	51.88 (7.16)	1.25 (0.89)	1.25 (0.89)
All	95.89 (2.46)	58.12 (8.03)	1.93 (1.63)	−0.93 (2.37)

Notes.
Average values are shown for metrics described in 'Evaluation metrics' and in parenthesis standard deviation.

Table 7 Leaf segmentation evaluation on CVPPP A1 subset.

	FBD (%)	SBD (%)	|DiC|	DiC
Mean	94.42 (1.48)	73.81 (5.32)	2.62 (1.33)	−2.62 (1.33)
Median	95.24	74.71	3.00	−3.00
Max	95.99	82.40	5.00	−1.00
Min	92.03	66.12	1.00	−5.00

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was supported by the National High-Tech R&D Program of China under Grant 2013AA102305, the National Natural Science Foundation of China under Grants 61573258, and the US National Science Foundation's BEACON Center for the Study of Evolution in Action, under cooperative agreement DBI-0939454. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National High-Tech R&D Program of China: 2013AA102305.
National Natural Science Foundation of China: 61573258.
US National Science Foundation's BEACON Center for the Study of Evolution in Action: DBI-0939454.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Yi Wang conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Lihong Xu prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data are provided in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.5036#supplemental-information.

REFERENCES

Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S. 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **34(11)**:2274–2282 DOI 10.1109/TPAMI.2012.120.

Alenya G, Dellen B, Foix S, Torras C. 2013. Robotized plant probing: leaf segmentation utilizing time-of-flight data. *IEEE Robotics & Automation Magazine* **20(3)**:50–59 DOI 10.1109/MRA.2012.2230118.

Alenya G, Dellen B, Torras C. 2011. 3D modelling of leaves from color and ToF data for robotized plant measuring. In: *IEEE international conference on robotics and automation*. Piscataway: IEEE, 3408–3414 DOI 10.1109/ICRA.2011.5980092.

Blei DM, Ng AY, Jordan MI. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* **3**:993–1022.

Cao L, Li FF. 2011. Spatially coherent latent topic model for concurrent object segmentation and classification. In: *Proceedings of IEEE international conference in computer vision (ICCV)*. Piscataway: IEEE DOI 10.1109/ICCV.2007.4408965.

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*. Berlin, Heidelberg: Springer, 202–213 DOI 10.1007/978-3-642-24028-7_19.

Comaniciu D, Meer P. 2002. Mean-shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Analysis & Machine Intelligence* **24(5)**:603–619 DOI 10.1109/34.1000236.

Comaniciu D, Ramesh V, Meer P. 2000. Real-time tracking of non-rigid objects using Mean-shift. In: *Computer vision and pattern recognition, 2000. Proceedings. IEEE conference on, vol. 2*. Hilton Head Island: IEEE, 142–149 DOI 10.1109/CVPR.2000.854761.

David M. 2010. Probabilistic topic models. *IEEE Signal Processing Magazine* **27(6)**:55–65 DOI 10.1109/MSP.2009.934715.

Dollár P, Zitnick CL. 2015. Fast edge detection using structured forests. *IEEE Transactions on Pattern Analysis & Machine Intelligence* **37(8)**:1558–1570 DOI 10.1109/TPAMI.2014.2377715.

Fang Z, Lu W, Lisi F. 2014. Segmentation method for cucumber disease leaf images under complex background. *Acta Agriculturae Zhejiangensis* **26(5)**:1346–1355.

Griffiths TL, Steyvers M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America* **101(Suppl 1)**:5228–5235 DOI 10.1073/pnas.0307752101.

Guo J, Xu L. 2017. Automatic segmentation for plant leaves via multiview stereo reconstruction. *Mathematical Problems in Engineering* **2017(4)**:1–11 DOI 10.1155/2017/9845815.

Hernández-Rabadán DL, Guerrero J, Ramos-Quintana F. 2012. Method for segmenting tomato plants in uncontrolled environments. *Engineering* **04(10)**:599–606 DOI 10.4236/eng.2012.410076.

Krähenbühl P, Koltun V. 2011. Efficient inference in fully connected CRFs with gaussian edge potentials. In: *Advances in neural information processing systems*. Granada, 109–117.

Lee WS, Slaughter DC. 2004. Recognition partially occluded plant leaves of using a modified watershed algorithm. *Transactions of the Asae* **47(4)**:1269–1280 DOI 10.13031/2013.16561.

Li FF, Perona P. 2005. A bayesian hierarchical model for learning natural scene categories. In: *Proc of the IEEE conference on CVPR*. San Diego, USA, II, 524–531 DOI 10.1109/CVPR.2005.16.

Ma J, Du K, Zhang L, Zheng F, Sun Z. 2017. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. *Computers & Electronics in Agriculture* **142(142)**:110–117 DOI 10.1016/j.compag.2017.08.023.

Manh AG, Rabatel G, Assemat L, Aldon MJ. 2001. AE—Automation and emerging technologies: weed leaf image segmentation by deformable templates. *Journal of Agricultural Engineering Research* **80(2)**:139–146 DOI 10.1006/jaer.2001.0725.

Meunkaewjinda A, Kumsawat P, Attakitmongcol K, Srikaew A. 2008. Grape leaf disease detection from color imagery using hybrid intelligent system. In: *International conference on electrical engineering/electronics, computer, telecommunications and information technology*. Ecti-Con. IEEE, 513–516 DOI 10.1109/ECTICON.2008.4600483.

Mezzo BD, Rabatel G, Fiorio C, Stafford J, Werner A. 2007. Weed leaf recognition in complex natural scenes by model-guided edge pairing. In: *4th European conference on precision agriculture*. Berlin, 141–147.

Microsoft. 2000. Microsoft research cambridge object recognition image database v2. *Available at https://www.microsoft.com/en-us/research/project/image-understanding/*.

Minervini M, Abdelsamea MM, Tsaftaris SA. 2014. Image-based plant phenotyping with incremental learning and active contours. *Ecological Informatics* **23(23)**:35–48 DOI 10.1016/j.ecoinf.2013.07.004.

Ming Y, Li H, He X. 2012. Connected contours: a new contour completion model that respects the closure effect. In: *Computer vision and pattern recognition (CVPR), 2012 IEEE conference on*. Piscataway: IEEE, 829–836 DOI 10.1109/CVPR.2012.6247755.

Noble SD, Brown RB. 2008. Spectral band selection and testing of edge-subtraction leaf segmentation. *Canadian Biosystems Engineering* **50(2)**:1–8.

Pape JM, Klukas C. 2015. Utilizing machine learning approaches to improve the prediction of leaf counts and individual leaf segmentation of rosette plant images. In: *Computer vision problems in plant phenotyping workshop*. 3.1–3.12 DOI 10.5244/C.29.CVPPP.3.

Persson M, Astrand B. 2008. Classification of crops and weeds extracted by active shape models. *Biosystems Engineering* **100(4)**:484–497 DOI 10.1016/j.biosystemseng.2008.05.003.

Reddy ND, Singhal P, Krishna KM. 2014. Semantic motion segmentation using dense CRF formulation. In: *Proceedings of the 2014 Indian conference on computer vision graphics and image processing*. New York: ACM, 1–8 DOI 10.1145/2683483.2683539.

Ren M, Zemel RS. 2016. End-to-End Instance Segmentation with Recurrent Attention.

Romera-Paredes B, Torr PHS. 2016. Recurrent instance segmentation. In: *European conference on computer vision*. New York: Springer International Publishing.

Russell BC, Freeman WT, Efros AA, Sivic J, Zisserman A. 2006. Using multiple segmentations to discover objects and their extent in image collections. In: *Computer vision and pattern recognition, 2006 IEEE computer society conference on*. IEEE, 1605–1614 DOI 10.1109/CVPR.2006.326.

Sampath A, Shan J. 2007. Building boundary tracing and regularization from airborne lidar point clouds. *Photogrammetric Engineering & Remote Sensing* **73(7)**:805–812 DOI 10.14358/PERS.73.7.805.

Sanz-Cortiella R, Llorens-Calveras J, Escolà A, Arnó-Satorra J, Ribes-Dasi M, Masip-Vilalta J. 2011. Innovative LIDAR 3D dynamic measurement system to estimate fruit-tree leaf area. *Sensors* **11(6)**:5769–5791 DOI 10.3390/s110605769.

Scharr H, Minervini M, Fischbach A, Tsaftaris SA. 2014. Annotated image datasets of rosette plants. In: *European conference on computer vision*. Zürich: Suisse, 6–12.

Scharr H, Minervini M, French AP, Klukas C, Kramer DM, Liu X, Luengo I, Pape JM, Polder G, Vukadinovic D, Yin X, Tsaftaris SA. 2016. Leaf segmentation in plant phenotyping: a collation study. *Machine Vision & Applications* **27(4)**:585–606 DOI 10.1007/s00138-015-0737-3.

Sogaard HT. 2005. Weeds classification by active shape models. *Biosystems Engineering* **91(3)**:271–281 DOI 10.1016/j.biosystemseng.2005.04.011.

Song Y, Wilson R, Edmondson R, Parsons N. 2007. Surface modelling of plants from stereo images. In: *3-D digital imaging and modeling, 2007. 3DIM'07. Sixth international conference on*. Montreal, Quebec: IEEE, 312–319 DOI 10.1109/3DIM.2007.55.

Tang X, Liu M, Zhao H, Tao W. 2009. Leaf extraction from complicated background. In: *International congress on image and signal processing*. IEEE, 1–5 DOI 10.1109/CISP.2009.5304424.

Teng C, Kuo Y, Chen Y. 2011. Leaf segmentation, classification, and three-dimensional recovery from a few images with close viewpoints. *Optical Engineering* **50(3)**:103–108 DOI 10.1117/1.3549927.

Valliammal N, Geethalakshmi SN. 2011. Hybrid image segmentation algorithm for leaf recognition and characterization. In: *International conference on process automation, control and computing*. Piscataway: IEEE, 1–6 DOI 10.1109/PACC.2011.5978883.

Valliammal N, Geethalakshmi SN. 2012. Plant leaf segmentation using non linear K means clustering. *International Journal of Computer Science Issues* **9(3)**:212–218.

Wang J, Wang S, Cui Y. 2011. Research on the color image segmentation of plant disease in the greenhouse. In: *International conference on consumer electronics, communications and network*. Piscataway: IEEE, 2551–2553 DOI 10.1109/CECNET.2011.5768494.

Wang X, Grimson E. 2008. Spatial latent dirichlet allocation. In: *Conference on neural information processing systems*. Vancouver: DBLP, 1577–1584.

Winn J, Criminisi A, Minka T. 2005. Object categorization by learned universal visual dictionary. In: *Tenth IEEE international conference on computer vision*. Piscataway: IEEE Computer Society, 1800–1807 DOI 10.1109/ICCV.2005.171.

Xia C, Wang L, Chung BK, Lee JM. 2015. In situ 3D segmentation of individual plant leaves using a RGB-D camera for agricultural automation. *Sensors* **15(8)**:20463–20479 DOI 10.3390/s150820463.

Yang Y, Li XU. 2014. Remote sensing image classification using layer-by-layer feature associative conditional random field. *Journal of Computer Applications* **34(6)**:1741–1745.

Zhang QQ, Zhang YL, Guo-Hong QI. 2017. Segmentation of cucumber disease leaves based on otsu method. *Journal of Anhui Agricultural Sciences* **45**(**12**):193–195.

Zheng LY, Zhang JT, Wang QY. 2009. Mean-shift-based color segmentation of images containing green vegetation. *Computers & Electronics in Agriculture* **65**(**1**):93–98 DOI 10.1016/j.compag.2008.08.002.

Zou QX, Yang LN, Peng L, Zheng Q. 2015. Segmentation algorithm based on blade lab space and K-means clustering. *Journal of Agricultural Mechanization Research* **2015**(**9**):222–226.