

A Method for Tree Image Segmentation Combined Adaptive Mean Shifting with Image Abstraction

Ting-ting Yang*, Su-yin Zhou*, Ai-jun Xu*, and Jian-xin Yin*

Abstract

Although huge progress has been made in current image segmentation work, there are still no efficient segmentation strategies for tree image which is taken from natural environment and contains complex background. To improve those problems, we propose a method for tree image segmentation combining adaptive mean shifting with image abstraction. Our approach perform better than others because it focuses mainly on the background of image and characteristics of the tree itself. First, we abstract the original tree image using bilateral filtering and image pyramid from multiple perspectives, which can reduce the influence of the background and tree canopy gaps on clustering. Spatial location and gray scale features are obtained by step detection and the insertion rule method, respectively. Bandwidths calculated by spatial location and gray scale features are then used to determine the size of the Gaussian kernel function and in the mean shift clustering. Furthermore, the flood fill method is employed to fill the results of clustering and highlight the region of interest. To prove the effectiveness of tree image abstractions on image clustering, we compared different abstraction levels and achieved the optimal clustering results. For our algorithm, the average segmentation accuracy (SA), over-segmentation rate (OR), and under-segmentation rate (UR) of the crown are 91.21%, 3.54%, and 9.85%, respectively. The average values of the trunk are 92.78%, 8.16%, and 7.93%, respectively. Comparing the results of our method experimentally with other popular tree image segmentation methods, our segmentation method get rid of human interaction and shows higher SA. Meanwhile, this work shows a promising application prospect on visual reconstruction and factors measurement of tree.

Keywords

Adaptive Mean Shift, Image Abstraction, Image Segmentation, Mathematical Morphology, Tree Segmentation

1. Introduction

In recent years, digital image processing based on machine vision has been widely used for various purposes, such as organ segmentation [1], fruit segmentation [2], insect detection [3], unmanned traffic control systems [4], and fingerprint identification [5]. These methods can be roughly classified into six classes: (1) Threshold approach relies on the gray-scale values in an image by comparing the value of each pixel, in turn, with the size of the threshold. Examples include threshold segmentation and the Otsu method [6,7]. This segmentation method that relies solely on gray values does not work well in the absence of color. (2) Region-based image segmentation methods divide pixels with similar properties

※ This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Manuscript received December 7, 2018; first revision August 16, 2019; accepted September 26, 2019.

Corresponding Author: Ai-jun Xu (xuaj1976@163.com)

* Zhejiang Agriculture and Forestry University, Hangzhou, China; Zhejiang Provincial Key Laboratory of Forestry Intelligent Monitoring and Information Technology, Hangzhou, China; Key Laboratory of State Forestry and Grassland Administration on Forestry Sensing Technology and Intelligent Equipment, Hangzhou, China (917251944@qq.com, zsy197733@163.com, xuaj1976@163.com, 10262029@qq.com)

Ting-ting Yang and Su-yin Zhou contributed equally to this work.

into the same region [8,9]. The phenomenon of under-segmentation or over-segmentation probably occur in these methods. (3) Edge-based approaches use edge points by judging the maximum value of the first derivative of the image or the zero-crossing information of the second derivative. Edge-based techniques include the Canny method [10,11] and Roberts method [12]. These methods are more suitable for the segmentation of regular objects and less effective for irregular objects. (4) Graph theory provides a mathematical approach that maps objects in an image onto a graph or sub-graph, as in the graph cut method [13] and grab cut method [14]. These kind of human interaction methods have good segmentation effects, but poor operability. (5) Deep learning approaches use convolutional neural network to study features of the target to achieve the purpose of semantic segmentation [15-17]. This method is more intelligent to get segmentation results, but need to train massive data, and require higher computer hardware configuration. (6) Other approaches are based on specific theories, such as wavelet transform [18] or clustering methods [19]. The mean shift algorithm [20] is a typical clustering method that does not require any prior knowledge and relies entirely on the sample point in the feature space to calculate the value of the density function. Thus, we adopted the mean shift algorithm to segment the tree from image.

The mean shift, an effective algorithm for estimating a probability density gradient, has been used widely for image segmentation and other applications [21-23]. Scholars applied the mean shift algorithm to image clusters before. The image pixels were converted to the sampling points of the feature space. Clustering in the feature space amounted to the segmentation of the image space. Usually, the image's gray-scale value was selected to constitute the feature space [24]. However, in practice, just relying on the gray-scale values cannot meet the complex requirements for image segmentation completely. Therefore, many scholars continued to seek other approaches to optimize image segmentation [25-27]. Bandwidths set by artificial approaches were replaced by the adaptive ones [28-30]. Although researchers have provided some advances in simple image segmentation, there are still problems while segmenting tree images. Tree images collected originally in the natural environment have too much noise in the background, such as houses, roads, and ground. Moreover, the branches and leaves of a tree caused many small hollows and obvious texture details in the image that are very different from the images of other types of natural objects, such as a brain or lung. The texture features and noise cannot be ignored when segmenting the image. Therefore, a meaningful and reasonable image pre-processing technique is necessary for using the mean shift algorithm to segment a tree image.

2. Materials and Methods

2.1 Materials and Data

The image data are obtained from the natural environment in Hangzhou, Zhejiang Province, on the southeast coast of China (119.72E, 30.23N), in the autumn, 2018. To verify the feasibility and accuracy of our method, we gathered a variety of tree images as input data. We used a mobile phone camera to collect 31 images of different shapes and species of trees at a distance of 2–8 m. The image has a resolution of 3120×4160 pixels. The characteristics of the tree images were as follows:

- 1) Illumination conditions: photographs were taken under a variety of lighting conditions, including direct sunlight, back-lighting, and shading, among others.

- 2) Tree species: we selected trees of different species with relatively luxuriant canopies.
- 3) Background features: photographs included uncut grass, roads, houses, ground, pedestrians, and other background features.

2.2 Overview of the Tree Image Segmentation Method

Fig. 1 provides a block diagram that summarizes the proposed algorithm. Considering the noisy and gaps in the canopies, first we abstracted each tree image by applying background smoothing and reducing the canopy gaps. This approach allowed us to obtain the abstracted salient tree images. Then, the optimized bandwidth parameters of the clustering algorithm were calculated using spatial location and gray-scale features. Combined with the Gaussian kernel function, we got the clustering results. Finally, based on the filled clustering results and the region of interest (ROI), the final tree segmentation image was obtained via mathematical morphology processing.

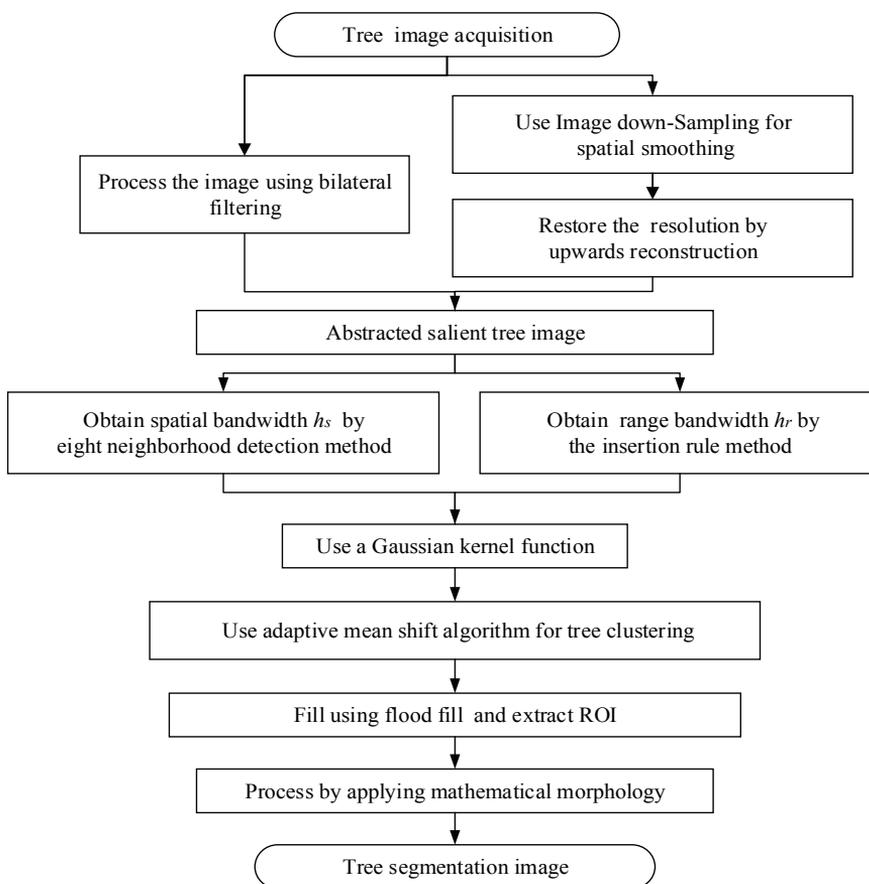


Fig. 1. Block diagram for the proposed algorithm.

A summary of the entire tree segmentation method based on image abstraction is shown below:

- 1) Image abstraction: We selected an arbitrary image from our image set. The image was processed using bilateral filtering and an image pyramid. Then we got the abstracted salient image.

- 2) Tree image clustering: For the spatial point x , we interactively searched its modal points by the joint feature space of the spatial domain and the gray domain. Then each pixel of the image was set as another new value using the flood fill method, and each was connected by a specific rule. The criteria for the realization of region connectivity were as follows:
 - (a) The space distance between two adjacent regions was less than h_s ;
 - (b) The gray distance between two adjacent regions was less than h_r ; and
 - (c) The minimum number of pixels of a region was set as M . If the number of pixels in a single region was fewer than M , the region was merged into adjacent regions.
- 3) ROI extraction of the tree image: According to the filled clustered image, the ROI of the tree image was extracted by color histogram of the image.
- 4) Mathematical morphology processing: An open operation was used repeatedly to remove the small noise found in the image and smooth the edges of the tree. Then a closed operation was used to fill the scattered hollows in the image of the branches and leaves of the target trees.

2.3 Image Abstraction

To blur the background and smooth canopy hollows of the tree image, we abstract pixel of the original image from multiple perspectives combined with the features of the tree image. This approach reduced the difficulties associated with segmentation caused by both the features of the tree itself and its background.

Bilateral filtering is a nonlinear filtering technique in which the kernel function not only considers the Euclidean distance of the pixels but also involves the difference in radiation of the pixels' range (e.g., the degree of similarity among the pixels in the convolution kernel and the central pixels, color intensity, and depth distance). As the tree is generally showed in the middle of the image in the data set, other objects located around the trees were smoothed and their edge information was well-preserved according to the filter characteristics. Therefore, bilateral filtering iterative processing was used for spatial smoothing to reduce the influence of background noise. For this paper, to describe the degree of image smoothing and denoising quantitatively, we introduced a filtering iteration time with the same spatial and range value, t ($t \in \mathbb{N}$).

Although the previous step solved the background texture interference to a certain extent, image features of the tree still need to be processed. Accordingly, we used the image pyramid method to smooth the image further. First, the smoothed image was down-sampled using the Gaussian pyramid. The first layer of the pyramid was convoluted with a Gaussian kernel, and all even rows and columns were deleted, causing the number of image pixels of the second layer to be a quarter of the original image. It can be seen from Fig. 2 that the image down-sample was filtered. We used interlaced sampling, which reduced the impact of the canopy gaps and foliage texture. Next, an upwards reconstruction method was used to restore the resolution size of the down-sampled image, filling the even rows and columns with 0 gray value. Then we used the same convolution kernel to enlarge the image to estimate the value of the "lost" pixels. By using these two steps iteratively, we achieved an abstracted salient tree image. Similarly, the image pyramid layer, l ($l \in \mathbb{N}$) (with one "layer" defined as one down-sample and one upward reconstruction), was introduced to measure the degree of image blurring.

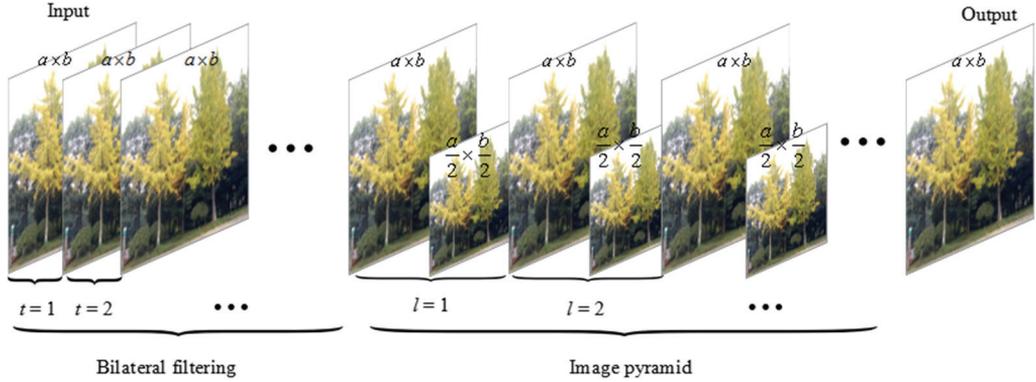


Fig. 2. Image abstraction process, where $a \times b$ is the image size.

2.4 Adaptive Mean Shift Algorithm

Mean shift is an iterative algorithm for kernel density estimation. The core of the mean shift technique is clustering the sample points in feature space. Let $x_i \in R^d$, $i = 1, 2, \dots, n$ denote the set of feature vectors in a d -dimensional feature space. The density function at a point x can be estimated by kernel function $K(x)$ and bandwidth h :

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (1)$$

The mean shift algorithm vector is given as

$$m_h(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (2)$$

where $g(x) = -K'(x)$.

In this paper, the spatial and gray feature vectors are used to estimate the corresponding feature bandwidths h_s and h_r :

$$K_{h_s, h_r}(x) = \frac{C}{h_s^2 h_r^d} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \quad (3)$$

where x_s and x_r denote the spatial part and the gray part of a feature vector, respectively. C is the normalization constant. The $K(x)$ profile kernel is named $k(x)$.

Then the adaptive mean shift translation vector based on multidimensional features is given as follows:

$$m_{h_s, h_r}(x) = \frac{\sum_{i=1}^n x_i \exp\left(-\frac{1}{2} \left\|\frac{x^s - x_i^s}{h_s}\right\|^2\right) \exp\left(-\frac{1}{2} \left\|\frac{x^r - x_i^r}{h_r}\right\|^2\right)}{\sum_{i=1}^n \exp\left(-\frac{1}{2} \left\|\frac{x^s - x_i^s}{h_s}\right\|^2\right) \exp\left(-\frac{1}{2} \left\|\frac{x^r - x_i^r}{h_r}\right\|^2\right)} \quad (4)$$

where x_i^s and x_i^r denote the spatial and gray feature vectors of x near sampling points, respectively.

An adaptive mean shift algorithm is used to determine the proper bandwidth. Spatial bandwidth h_s affects not only the mis-segmentation rate of the image but also the running speed of the algorithm by iterations. We calculate the spatial bandwidth using the method proposed by [31]. First, the initial value of spatial bandwidth H_0 is given as 8, and the increment s as 3; then $h_i = h_{i-1} + s$. When $\Sigma\sigma_j < 0.7n_i$ ($j = 1, 2, \dots, n_i$), the corresponding spatial bandwidth h_s is set as n_i . In other words, when the number of the sampling points is less than 70% of all sampling points n_i , the corresponding h_i is the required spatial bandwidth h_s . The sampling points in which gray values are similar to the smoothed points belong to the spatial bandwidth h_s . The condition under which the sampling point gray value is similar to the smoothed point gray value is $|h_j - h_0| \leq 8$, where h_j denotes the H component value of the sampling point, h_0 denotes the H component value of the smoothed point, and the symbol function σ_j is 1 when $|h_j - h_0| \leq 8$, and 0 otherwise.

The bandwidth of gray feature h_r , is an important parameter when smoothing an image. Generally, there is a global optimal fixed-range bandwidth and an adaptive-range bandwidth. In this paper, the insertion rule method [30] is used to obtain the adaptive h_r :

$$h_r = \left(\frac{4}{d+2}\right)^{1/(d+4)} n^{-1/(d+4)} \sigma_j \quad (5)$$

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (G_x - G_x)^2}$$

where d denotes the dimension of the feature space, n denotes the amount of data, σ_j denotes the standard deviation, G_x ($x = 1, 2, \dots, n$) denotes the gray value of each pixel of the image, and G_x denotes the average gray value of the image.

As the image abstraction method smoothed the original tree image and reduced the noise influence of small particles to a certain extent, the two bandwidths of the local adaptive h_s and h_r based on image abstraction achieved a compromise between over-smoothing and under-smoothing in the process of smoothing the texture.

3. Results

3.1 Algorithm Evaluation

To measure the accuracy rate of the segmentation, we evaluate our algorithm via segmentation accuracy (SA), over-segmentation rate (OR), and under-segmentation rate (UR), as shown below.

$$SA = \left(1 - \frac{|R_s - T_s|}{R_s}\right) \times 100\% \quad (6)$$

$$OR = \frac{O_s}{R_s + O_s} \quad (7)$$

$$UR = \frac{U_s}{R_s + O_s} \quad (8)$$

where R_s denotes the ground truth, and T_s denotes the area segmented by our method. Therefore $|R_s - T_s|$ denotes the number of pixels segmented mistakenly, O_s denotes the number of pixels that should not be

included in the total number of pixels of the segmentation result, but were counted nonetheless. U_s is the number of pixels that should have been included in the total number of pixels of the segmentation result, but were not counted.

3.2 Evaluation with Different Abstraction Levels

The method of tree image abstraction was validated using a filtering iteration time (t) from 0 to 15 and image pyramid layer (l) from 0 to 10. Fig. 3 (which is based on one sample image only) shows subjectively the clustering results of our algorithm versus different levels of image abstraction. From Fig. 4 (evaluation of the segmentation results of the crown and trunk of 31 sample images at different abstraction levels), it is obvious that the degree of image abstraction had a certain impact on image segmentation. Additionally, when l or t was relatively small or big, the clustering results were not satisfactory, but they were more satisfied with $t = 10$, $l = 5$ or 10. Fig. 5 shows objectively the segmentation results of our algorithm based on different tree abstraction images. Obviously, the SA was higher and more stable with t ranging from 5 to 9 and l from 0 to 10, and the SA decreased with the increase of t . Furthermore, the crown segmentation was affected more by image abstraction than the trunk.

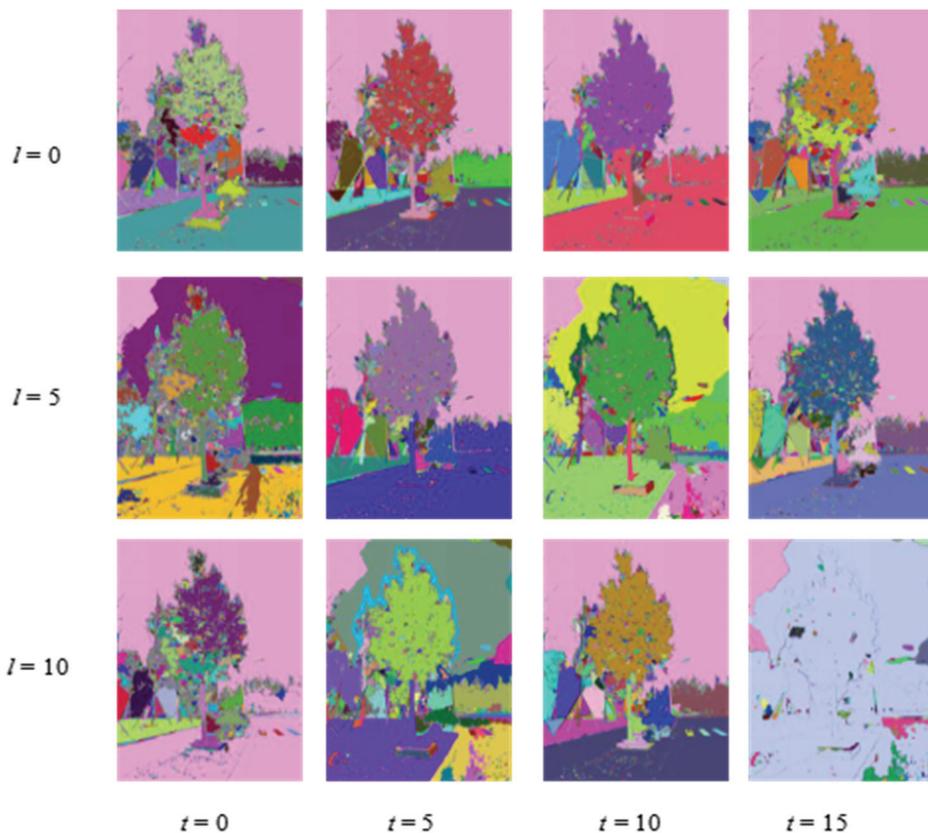


Fig. 3. Clustering results of different abstraction levels. From left to right: t is 0, 5, 10, and 15, respectively. From up to down: l is 0, 5, and 10, respectively.

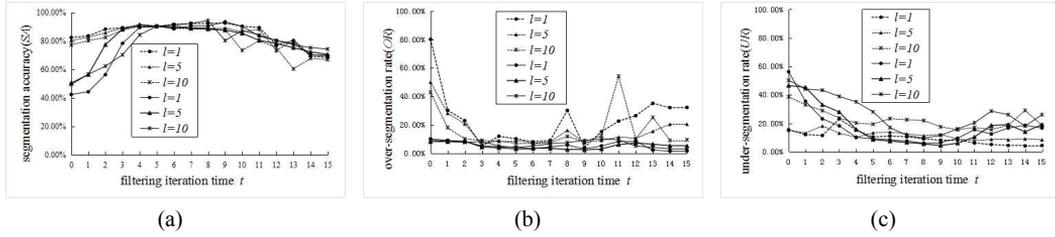


Fig. 4. An evaluation of the crown and trunk segmentation result database, where t ranges from 0 to 15 and l is 0, 5, or 10: (a) segmentation accuracy, (b) over-segmentation rate, and (c) under-segmentation rate. The solid line shows the segmentation results of the crown, and the dashed line shows the segmentation results of the trunk.

3.3 Segmentation Results based on Best Image Abstraction

In the next part of our work, we validated the proposed algorithm using tree image segmentation data. The segmentation results based on best image abstraction ($t = 10$ and $l = 5$) are shown in Fig. 5 (which is based on just one sample image). The following steps describe our procedure.

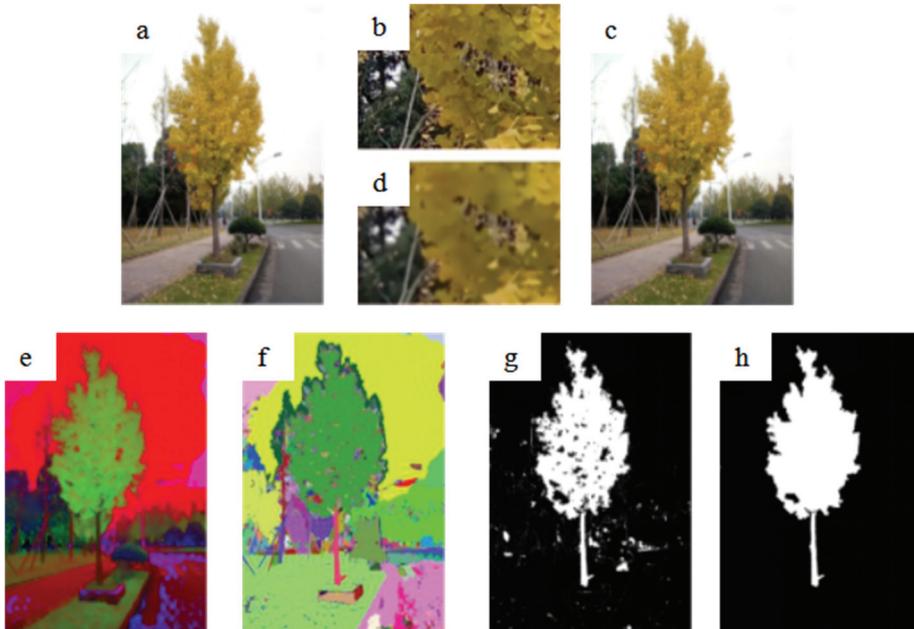


Fig. 5. Segmentation results of a tree image: (a) original tree image; (b) part of (a); (c) abstracted salient tree image; (d) part of (c); (e) clustering result; (f) filled clustering result; (g) ROI extraction result; and (h) segmentation result.

Step 1 (Searching the modal points): The spatial vector and the gray-scale vector are combined into a “space-color” field. Based on the optimal values $h_s = 29.8$ and $h_r = 19.9$, mean shift filtering is performed to obtain the tree clustering image. As the result of this filtering, m modal points are obtained, and all the pixels are clustered into the same model point to form a cluster $\{Q_p\}$, $p = 1, 2, \dots, m$, and $m \ll n$, where m is the number of clustering regions, and n is the data volume.

Step 2 (Combining similar areas and merging small areas): Step 1 is the process of image classification smoothing. In Step 2, the clustering result is filled using the flood fill function. The gray-scale value of pixels in the area is set to a new value, and the regional connection is realized by specific rules.

Step 3 (Segmenting the tree image): Based on the filled clustering image, the gray-scale thresholds of the canopy and trunk are set, and the ROI areas are exacted by color code table (crown with $R = 49$, $G = 204$, $B = 52$, and trunk with $R = 191$, $G = 23$, $B = 130$). The final result of the adaptive mean shift method (AMSM) for tree segmentation based on image abstraction is obtained through iterative application of the corrosion and expansion operations of mathematical morphology to provide further denoising.

4. Discussion

We also compared the performance of our proposed algorithm with other published methods that have been applied to tree image segmentation. The method used by [32] employed only color features for tree image segmentation in simple scenes. The method adopted by Zhao et al. [33], used fractional dimension-based color to segment tree images. Jiang [34] used color feature regions, a spatial clustering method, and color information to segment tree images divided into two parts, the trunk and crown, but just adapted to simple scenes. Ding et al. [35] utilized the grab cut algorithm to segment tree images. In addition, our own previous tree segmentation method based on the graph cut algorithm employed an interactive method to segment the image [36]. Fig. 6 shows the results of some of the segmentation methods in a visual form that can be viewed intuitively.



Fig. 6. Comparison of segmentation results from four methods. From left to right: original image, ground truth, segmentation results by color feature, segmentation results by color and fractional dimension method, segmentation results using the grab cut algorithm, segmentation results using the graph cut algorithm, and our proposed AMSM algorithm.

Table 1. Average segmentation accuracy for 31 tree images

	Automatic level	SA (%)		OR (%)		UR (%)	
		Crown	Trunk	Crown	Trunk	Crown	Trunk
Ge et al. [32]	Fully automatic	72.234	11.950	32.189	80.067	1.258	3.260
Ding et al. [35]	Semi-automatic	85.127	80.106	12.523	20.483	2.401	18.384
Yang et al. [36]	Artificial interaction	94.341	91.021	2.598	6.857	4.409	5.057
Our AMSM algorithm	Fully automatic	91.212	92.781	3.543	8.160	9.852	7.934

Table 1 provides an objective comparison of our algorithm with the average accuracy of the other methods of tree image segmentation based on best image abstraction. For our algorithm, the average values of SA, OR, and UR of the crown were 91.21%, 3.54%, and 9.85%, respectively. The average values of SA, OR, and UR of the trunk were 92.78%, 8.16%, and 7.93%, respectively. These results were better than those obtained by the color feature, fractional dimension, or grab cut methods. Compared with the graph cut algorithm, our method was more automatic, efficient, and accurate.

5. Conclusion and Future Works

In this paper, we proposed a new tree segmentation algorithm based on a combination of image abstraction and the adaptive mean shift algorithm. First, we analyzed the challenges involved in tree image segmentation, and we proposed spatial smoothing and image pyramid operations to reduce the influence of unnecessary background information and canopy gaps on clustering. The degree of image abstraction was measured by the filtering iteration time (t) and image pyramid layer (l). A comparison of segmentation results was made using different abstraction levels, with t in the range of 0–15 and l from 0–10. We found that the segmentation results were better when $l = 5$ and t ranged from 4 to 8. Then, to eliminate the need to set the parameters manually, we utilized an adaptive mean shift segmentation algorithm that used the corresponding bandwidth solution methods in both the spatial and range domains. Spatial location and gray scale features were introduced, obtained by step detection and the insertion rule method, respectively. The flood fill method was then employed to fill the results of clustering to highlight the region of interest. To prove the effectiveness of tree image abstraction on image clustering, we compared different abstraction levels and achieved the optimal clustering results.

With the experimental data set including various types of tree images, our algorithm achieved an average SA of more than 90%. In comparison with two segmentation algorithms using color and fractional dimension-based color, our proposed segmentation algorithm showed relatively high SA and finer edge information of tree. What's more, our method didn't require human marking and demonstrated a higher degree of accuracy for tree image segmentation than prior approaches. The research is also of great significance for the other target segmentation and the measurement of target size based on image information. However, when the target tree to be photographed is far away from the camera or the surrounding environment of tree is more complex, the SA of tree is reduced. To solve this problem, in the next step we will devote to use the deep learning method to study tree features such as shape or texture or edge or color. Therefore, in the future of our work, we will improve the performance of the method and engage to use this segmentation results to solve tree information measurements. For example, we are

now trying to calculate tree height and crown and diameter at breast height (DBH) based on segmentation results of tree. It may be useful for forest survey.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 31670641), Zhejiang Science and Technology Key R&D Program Funded Project (No. 2018C02013) and Zhejiang Public Welfare Project (No. LGN21C160004).

References

- [1] Y. Boykov and M. P. Jolly, "Interactive organ segmentation using graph cuts," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Heidelberg, Germany: Springer, 2000, pp. 276-286.
- [2] Q. Zeng, Y. Miao, C. Liu, and S. Wang, "Algorithm based on marker-controlled watershed transform for overlapping plant fruit segmentation," *Optical Engineering*, vol. 48 no. 2, article no. 027201, 2009.
- [3] H. T. Zhang, H. P. Mao, and D. Y. Qiu, "Feature extraction for the stored-grain insect detection system based on image recognition technology," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 25 no. 2, pp. 126-130, 2009.
- [4] C. S. Sharp, O. Shakernia, and S. S. Sastry, "A vision system for landing an unmanned aerial vehicle," in *Proceedings of IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*, Seoul, South Korea, pp. 1720-1727, 2001.
- [5] D. K. Isenor and S. G. Zaky, "Fingerprint identification using graph matching," *Pattern Recognition*, vol. 19 no. 2, pp. 113-122, 1986.
- [6] B. Wang, L. L. Chen, and J. Cheng, "New result on maximum entropy threshold image segmentation based on P system," *Optik*, vol. 163, pp. 81-85, 2018.
- [7] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.
- [8] Y. Xu, W. Yao, L. Hoegner, and U. Stilla, "Segmentation of building roofs from airborne LiDAR point clouds using robust voxel-based region growing," *Remote Sensing Letters*, vol. 8 no. 11, pp. 1062-1071, 2017.
- [9] D. Zhou and Y. Shao, "Region growing for image segmentation using an extended PCNN model," *IET Image Processing*, vol. 12 no. 5, pp. 729-737, 2018.
- [10] L. Ding and A. Goshtasby, "On the Canny edge detector," *Pattern Recognition*, vol. 34 no. 3, pp. 721-725, 2001.
- [11] Y. M. Luo and R. Duraiswami, "Canny edge detection on NVIDIA CUDA," in *Proceedings of 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, 2008.
- [12] A. K. Cherri and M. A. Karim, "Optical symbolic substitution: edge detection using Prewitt, Sobel, and Roberts operators," *Applied Optics*, vol. 28 no. 21, pp. 4644-4648, 1989.
- [13] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in *Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, 2008.
- [14] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309-314, 2004.
- [15] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 1, pp. 386-397, 2020.

- [16] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 2017.
- [17] X. Li, H. Chen, X. Qi, Q. Dou, C. W. Fu, and P. A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663-2674, 2018.
- [18] S. Loutridis, E. Douka, L. J. Hadjileontiadis, and A. Trochidis, "A two-dimensional wavelet transform for detection of cracks in plates," *Engineering Structures*, vol. 27, no. 9, pp. 1327-1338, 2005.
- [19] S. Balla-Arabe, X. Gao, D. Ginjac, V. Brost, and F. Yang, "Architecture-driven level set optimization: From clustering to subpixel image segmentation," *IEEE Transactions on Cybernetics*, vol. 46 no. 12, pp. 3181-3194, 2016.
- [20] Z. Zhuge, M. Xu, and Y. Liu, "Fabric image segmentation algorithm based on Mean Shift," *Journal of Textile Research*, vol. 28, no. 10, pp. 108-111, 2007.
- [21] L. Lin, D. Garcia-Lorenzo, C. Li, T. Jiang, and C. Barillot, "Adaptive pixon represented segmentation (APRS) for 3D MR brain images based on mean shift and Markov random fields," *Pattern Recognition Letters*, vol. 32 no. 7, pp. 1036-1043, 2011.
- [22] J. Sun, "A fast MEANSHIFT algorithm-based target tracking system," *Sensors*, vol. 12, no. 6, pp. 8218-8235, 2012.
- [23] Y. Liu, S. Z. Li, W. Wu, and R. Huang, "Dynamics of a mean-shift-like algorithm and its applications on clustering," *Information Processing Letters*, vol. 113 no. 1-2, pp. 8-16, 2013.
- [24] M. H. Jeong, B. J. You, Y. Oh, S. R. Oh, and S. H. Han, "Adaptive mean-shift tracking with novel color model," in *Proceedings of IEEE International Conference Mechatronics and Automation*, Niagara Falls, Canada, 2005, pp. 1329-1333.
- [25] H. Cho, S. J. Kang, S. I. Cho, and Y. H. Kim, "Image segmentation using linked mean-shift vectors and its implementation on GPU," *IEEE Transactions on Consumer Electronics*, vol. 60 no. 4, pp. 719-727, 2014.
- [26] K. Du, Y. Ju, Y. Jin, G. Li, Y. Li, and S. Qian, "Object tracking based on improved MeanShift and SIFT," in *Proceedings of 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, Yichang, China, 2012, pp. 2716-2719.
- [27] G. B. Li and H. F. Wu, "Weighted fragments-based mean-shift tracking using color-texture histogram," *Journal of Computer-Aided Design & Computer Graphics*, vol. 23, no. 12, pp. 2059-2066, 2011.
- [28] C. Liu, A. Zhou, Q. Zhang, and G. Zhang, "Adaptive image segmentation by using mean-shift and evolutionary optimization," *IET Image Processing*, vol. 8, no. 6, pp. 327-333, 2014.
- [29] A. Mayer and H. Greenspan, "An adaptive mean-shift framework for MRI brain segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1238-1250, 2009.
- [30] J. Zhou, J. Zhu, X. Mei, and H. Ma, "An adaptive MeanShift segmentation method of remote sensing images based on multi-dimension features," *Geomatics & Information Science of Wuhan University*, vol. 37 no. 4, pp. 419-418, 2012.
- [31] Y. Wang and Y. Sun, "Adaptive Mean Shift based image smoothing and segmentation," *Acta Automatica Sinica*, vol. 36 no. 12, pp. 1637-1644, 2012.
- [32] Y. F. Ge, H. P. Zhou, J. Q. Zheng, and H. C. Zhang, "A tree image segmentation algorithm based on relative color indices," *Journal of Nanjing Forestry University (Natural Science Edition)*, vol. 2 no. 4, pp. 19-22, 2012.
- [33] M. Zhao, J. Qiang, X. Lin, and X. Feng, "Tree image Segmentation method based on the fractional dimension," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 35, no. 2, pp. 72-75, 2004.
- [34] S. H. Jiang, "Research segmentation methods of stumpage image based on Android System," M.S. thesis, Northeast Forestry University, Harbin, China, 2015.

- [35] W. Ding, S. Zhao, S. Zhao, J. Gu, W. Qui, and B. Guo, "Measurement methods of fruit tree canopy volume based on machine vision," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 47, no. 6, pp. 1-10, 2016.
- [36] T. Yang, F. Guan, and A. Xu, "Multiple trees contour extraction method based on Graph Cut algorithm," *Journal of Nanjing Forestry University (Natural Sciences Edition)*, vol. 42, no. 6, pp. 91-98, 2018.



Ting-ting Yang <https://orcid.org/0000-0001-5441-1188>

She received B.S. degree in GIS from Anhui Science and Technology University in 2017. She is currently working as Master Degree Candidate of Zhejiang Agriculture and Forestry University in Zhejiang, China. Her main research covers deep learning and image processing.



Su-yin Zhou <https://orcid.org/0000-0001-8501-3879>

She received the Master's degree in Control Theory and Control Engineering from Shandong University of Science and Technology in 2003. She is currently working as a Lecturer in the School of Information Engineering, Zhejiang Agriculture and Forestry University, Hangzhou, China. Her current research interest includes the application of the Internet of Things and computer application technology in agriculture.



Ai-jun Xu <https://orcid.org/0000-0001-6789-6938>

He received the Ph.D. degree in Photogrammetry and Remote Sensing from Wuhan University in 2007. He is currently working as a Professor in School of Information Engineering, Zhejiang Agriculture and Forestry University, Hangzhou, China. His current research interest includes Computer Application Technology and the application of GIS in the direction of agricultural informatization, etc.



Jian-xin Yin <https://orcid.org/0000-0003-4443-5914>

She received the Master's degree in Mechatronic Engineering from Zhejiang University of Technology in 2007. She is currently working as an Associate Professor in School of Information Engineering, Zhejiang Agriculture and Forestry University, Hangzhou, China. Her current research interest includes computer vision and image processing, etc.