BIOSYSTEMS ENGINEERING 200 (2020) 215–230

Available online at www.sciencedirect.com

ScienceDirect

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

Research Paper

Automated classification of stems and leaves of potted plants based on point cloud data

Zichu Liu, Qing Zhang, Pei Wang*, Zhen Li, Huiru Wang

College of Science, Beijing Forestry University, No.35 Qinghua East Road, Haidian District, Beijing, 100083, China

ARTICLE INFO

Article history:
Received 28 March 2020
Received in revised form
7 October 2020
Accepted 12 October 2020
Published online 29 October 2020

Keywords:
Point cloud data
Automated classification
SVM
Leaf samples
Stem samples

The accurate classification of plant organs is a key step in monitoring the growing status and physiology of plants. A classification method was proposed to classify the leaves and stems automatically based on the point cloud data of the potted plants. Leaf samples and stem samples were selected automatically by using the three-dimensional (3D) convex hull algorithm and the two-dimensional (2D) projection grid density respectively, and were used to construct the leaf and stem training sets. Then, the point cloud data were classified into leaf points and stem points by using the support vector machine (SVM) algorithm. The point cloud data of three potted plants were used in the experiment. The proposed method was compared with the standard classification, the random selection method and the manual selection method. Among these methods, the proposed method is automated and time-saving. The results show that the proposed method had a good overall performance on accuracy and running time. The proposed method is efficient and effective on the leaf and stem classification of the plant point cloud data.

1. Introduction

Plant organs and their characteristics are very important to many plant studies. They can be used to monitor plant growth status and study plant physiological characteristics. Williams and Ayars (2005) measured grapevine organs to estimate plant photosynthesis. Conde (2011) estimated the effects of different nutrients on plant growth through measuring the growth statuses of different organs. Davi et al. (2005) studied carbon-water circulation through tree canopy structures. Plant organ research also plays an important role in environmental management. Huang, Zhao, Xu, Zhang, and Jiang (2019) studied *Broussonetia papyrifera* organs to study the effects of saline-alkali stress on plant morphology and growth. Fitter (1987) studied the adaptability of plants in different environments by comparing different root structures. Sampling plant organs was the most common method in practice in the above studies. However, the traditional destructive sampling method usually requires cutting the plant organs for measurement, which is time-consuming and destructive. Traditional methods not only require a large amount of time and energy but can also result in the destruction of plants and trees, which will have negative consequences on the ecological environment when conducting a large number of tests and data collections.

3D laser scanning technology can obtain the point cloud data of plants accurately and quickly, which provides a

* Corresponding author.
E-mail addresses: lzc0330@bjfu.edu.cn (Z. Liu), zhangq@bjfu.edu.cn (Q. Zhang), wangpei@bjfu.edu.cn (P. Wang), lizhen@bjfu.edu.cn (Z. Li), whr2019@bjfu.edu.cn (H. Wang).
https://doi.org/10.1016/j.biosystemseng.2020.10.006
1537-5110/

Nomenclature

L	Leaf sample point set
S	Stem sample point set
r_1	Radius for taking leaf training set
r_2	Radius for taking stem training set
X_{Leaf}	Leaf training set
X_{Stem}	Stem training set
D	The plant point cloud data set
p_i	Randomly sampled points
s_i	Sphere for projection
r	Radius for s_i
l	The number of points in sphere s_i
num_i	The number of grids occupied by the projections of all points inside the sphere s_i
m_i	The grid density of point p_i
σ	The parameter of RBF kernel
L_1	Artificially selected leaf tip point set
L_2	Artificially selected leaf root point set
L_S	Standard leaf point set
L_C	Classified leaf point set
S_S	Standard stem point set
S_C	Classified stem point set
$Card()$	Cardinal number of a set
TP	The number of correctly classified leaf points
TN	The number of correctly classified stem points
FP	The number of mistakenly classified leaf points
FN	The number of mistakenly classified stem points
N	The points number of point cloud data
$kappa$	Kappa coefficient
p_o	One of the variables for calculating kappa coefficient
p_e	One of the variables for calculating kappa coefficient

solution for non-destructive data collection and the fine-grained analysis of plant organs. The point cloud data of plants, collected with high precision and high density, record the plant geometry accurately, which is good for the analysis of plant organs. This technology has been applied in related studies on plants and agroforestry (Caccamo et al., 2018; Maltamo, Korpela, Tokola, Hyyppa, & Orka, 2010; Méndez, Rosell-Polo, Pascual, & Escolà, 2016; Morsdorf, Allgower, Hetherington, Danson, & Koetz, 2007; Su, Zhu, Huang, & Guo, 2018; Zheng, Moskal, & Kim, 2013; Zheng & Moskal, 2012), and the use of plant point cloud data to accurately classify and identify different organs of plants is a prerequisite for conducting such studies. Many scholars have proposed related classification methods. Wahabzada, Paulus, Kersting, and Anne-Katrin (2015) proposed a data-driven method for plant organ segmentation when plants are occluded. Yun, Gao, Wang, and Liu (2013) proposed a classification method for constructing a covariance matrix based on neighbourhood information. The method extracts the feature vector of each point from the scanned point cloud data and then uses the manifold learning method to reduce the point cloud. Yun et al. (2016) combined the shape, normal vector distribution and structure tensor of point cloud data features with SVM to separate various tree organs and estimate the leaf area. Paulus, Dupuis, Mahlein, and Kuhlmann (2013) used plant histological surface histograms to classify individual plant organs through a fully automated system and applied the proposed method to wheat estimation. Hétroy-Wheeler, Casella, and Boltcheva (2016) proposed a semi-automatic point cloud classification method to divide small tree seedlings into leaves, petioles, and stems. The method has strong robustness, and its false-positive rate and false-negative rate were both approximately 1%. Frasson and Krajewski (2010) proposed a method for using a preprocessed grid to represent plant organs and manually dividing the meshes into different morphological regions. Ma, Zheng, Eiter, Magney, and Moskal (2016) used a geometric feature-based automatic forest point classification (GAFPC) algorithm to divide tree point cloud data into photosynthetic and non-photosynthetic forest canopy components, and bare earth. Ferrara et al. (2018) proposed a method for using voxel density and density-based spatial clustering of applications with noise (DBSCAN) algorithm to divide cork oak trees into wood points and non-wood points. Tao et al. (2015) proposed a geometric method for leaves and wood classification based on different shapes of trunk/branch boundaries and leaf clusters. (Hang et al., 2017) used a 3D-to-2D projection and an x-axis pixel density distribution method to segment the stems and leaves. Xiang et al. (2019) proposed a skeletonisation method, which searched for the stem by using the Hough line transform, then segmented each individual leaf. Sodhi, Vijayarangan, and Wettergreen (2017, pp. 5180–5187) presented an automated method for 3D reconstructions and classified the 3D point cloud data into stem and leaves by using the SVM method. Dey, Mummert, and Sukthankar (2012) used the SVM method to classify the organs of grapevines.

In general, for rough manual sampling, the results are subjectively influenced by the operator and the sample. For careful manual sampling classification, it is necessary to make repeated sample selection attempts according to the different morphological characteristics and physiological structures of plants. To obtain better classification results, the complicated operation caused by careful manual sampling consumes considerable time and effort, which increases the research cycle and cost. Therefore, automation of the process of selecting sample points and classification can greatly reduce the time cost of related work and has great significance in practical applications.

In this paper, we propose an automated SVM classification method based on the spatial distribution characteristics and density distribution characteristics of plant point cloud data, experiment on three plant point cloud data sets, and discuss the feasibility, advantages and classification accuracy of this algorithm.

The steps of this work are as follows: (1) The point cloud data of three different potted plants are scanned and extracted. (2) Next, the standard classification result is subsequently constructed. (3) The automated selection method is proposed. (4) Two different methods were used for comparison. (5) Finally, the compared results are discussed, and conclusions are drawn.

2. Materials and methods

2.1. *Experimental data*

In the experiments, three potted plants were scanned using the HDI 120A-B scanner (3D LMI technologies company, Burnaby BC, Canada), which has the ability to acquire high-precision point cloud data with structured light, as shown in Fig. 1.

This device is composed of an imaging module and a light source module, and its overall size is small, which greatly reduces the requirements for the scanner operator, greatly improves the scanning efficiency and maintains a high scanning accuracy. The specific information of this 3D scanner is shown in Table 1.

Three different potted plants were scanned in the experiments: *Zamioculas zamiifolia*, *Pachyphytum bracteosum*, and *Dieffenbachia picta* (Fig. 2).

To obtain more visual test results and better classification results, the raw point cloud data of the three plants were denoised and rotated to make the stems of the three plants as perpendicular as possible to the XOY plane of the 3D coordinate system, as shown in Fig. 3.

The numbers of points of these three potted plants after denoising and the information of the circumscribed cuboids of the three plants are listed in Table 2.

2.2. *Method*

2.2.1. *Method description*

First, a standard classification result was provided and was regarded as a standard result to evaluate the proposed method. The potted plant point clouds were carefully and manually classified into leaf points and stem points. The result was used as the standard result.

Then, an automated selection method was proposed to classify the leaves and stems of the potted plants in this paper. In the method, the leaf samples and stem samples were acquired automatically by using different algorithms. Then the training sets of the leaves and stems were constructed and the point cloud data were classified by using the SVM algorithm. Only x, y, z coordinates of point cloud data were used as features for training and classification.

In addition, two classification methods with different sample selection methods were introduced to compare with the automated selection method. The first method is the random selection method. A certain number of points in the point cloud were randomly sampled. Then, the points were classified into leaf samples and stem samples manually. The second method is the manual selection method. Leaf samples and stem samples were selected manually in the point cloud. Obviously, these two methods obtained the leaf samples and stem samples in a partly or totally manual way. When the leaf samples and stem samples are ready, the two methods classify the point cloud with the same process of the construction of training sets and SVM classification that are described in above sections.

Finally, the results of three methods were compared with the standard result and analysed. The flow chart is shown in Fig. 4.

2.2.2. *Automated selection of leaf samples*

Considering the spatial structure of the potted plants, which has the leaf point clouds of the plants distributed on the

Table 1 – Information of HDI 120A-B 3D scanner.

3D SCANNER	LMI HDI 120
Camera	2 × 13000 pixel
Scanning Software	FlexScan3D
Scan Speed	0.3 second per scan
Field of View	124 × 120 *mm* – 192 × 175 *mm*
Average Number of Points	985000 per scan
Average Number of Polygons	19700 per scan
Point-to-Point Distance	0.162 *mm*
Accuracy	±0.02 *mm*

Fig. 1 – HDI 120 A-B scanner.

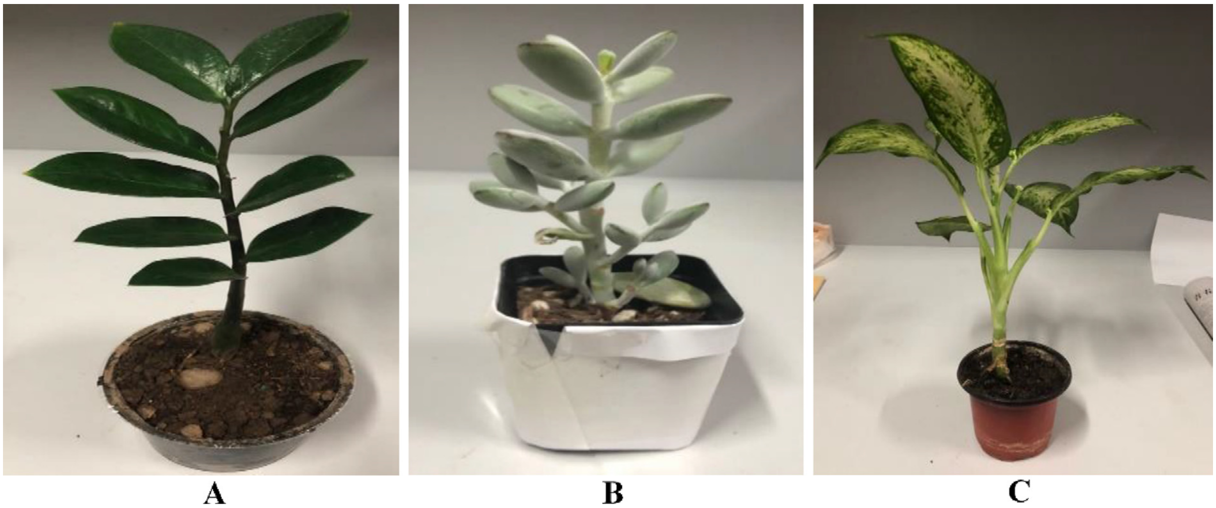

Fig. 2 – Three scanned plant pictures. A: *Zamioculas zamiifolia*. B: *Pachyphytum bracteosum*. C: *Dieffenbachia picta*.

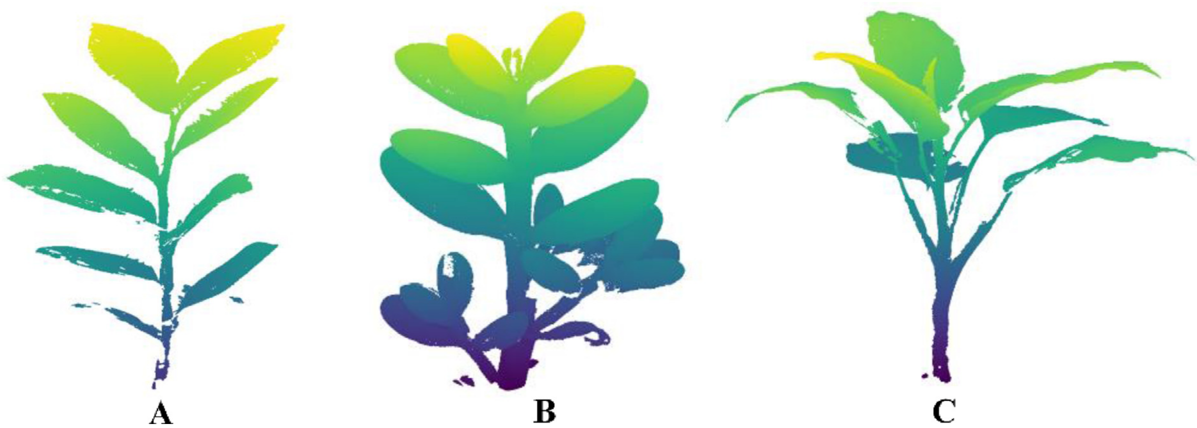

Fig. 3 – Point cloud data of the three plants. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

Table 2 – Point-related information of the three plants.

PLANT	X length/ mm	Y length/ mm	Z length/ mm	Number
Z. zamiifolia	137.8	108.8	206.4	1044220
P. bracteosum	73.4	72.7	104.5	871577
D. picta	248.8	225.0	269.1	2691531

periphery of the point cloud data, the leaf samples were selected by using the 3D convex hull algorithm.

The convex hull of a given set refers to the intersection of all convex sets containing the set in the real vector space. The 3D convex hull is the smallest convex polyhedron that contains all points in the given point cloud set (Claret & Day, 1994).

Since the stem is generally inside the plant and the points of the leaf tips are generally in the outermost part of the plant, the turning points of the obtained 3D convex hull, i.e., the apexes of the smallest convex polyhedron, are usually the points of the leaf tips; thus, these turning points were selected as samples for the plant leaves.

Taking the *Z. zamiifolia* as an example, the 3D convex hull can be constructed from the point cloud data which is shown in Fig. 5A-B. Then, the turning points of the 3D convex hull were selected as the leaf samples of the *Z. zamiifolia* (Fig. 5C).

2.2.3. *Automated selection of stem samples*

Because the stem point cloud data are mainly concentrated in the middle part of the plant point cloud data, it is difficult to directly obtain its sample distribution. Considering that the point density of the stems is greater along the stem's direction than the point density of leaves, the point density can be used to determine the stem samples. Therefore, the random sampling and 2D projection grid density method is used to automatically obtain the stem samples.

First, n points of the plant point cloud data were sampled randomly. The point p_i was selected as the centre, r was set as the radius, and the sphere s_i was created, where p_i denotes each randomly sampled point, $i = 1, \cdots, n$.

Second, orthogonal projection is used to calculate the projection grid density. The Lambert azimuthal equal-area projection and the stereographic equal-angle projection have been used to convert the 3D point cloud data to 2D raster

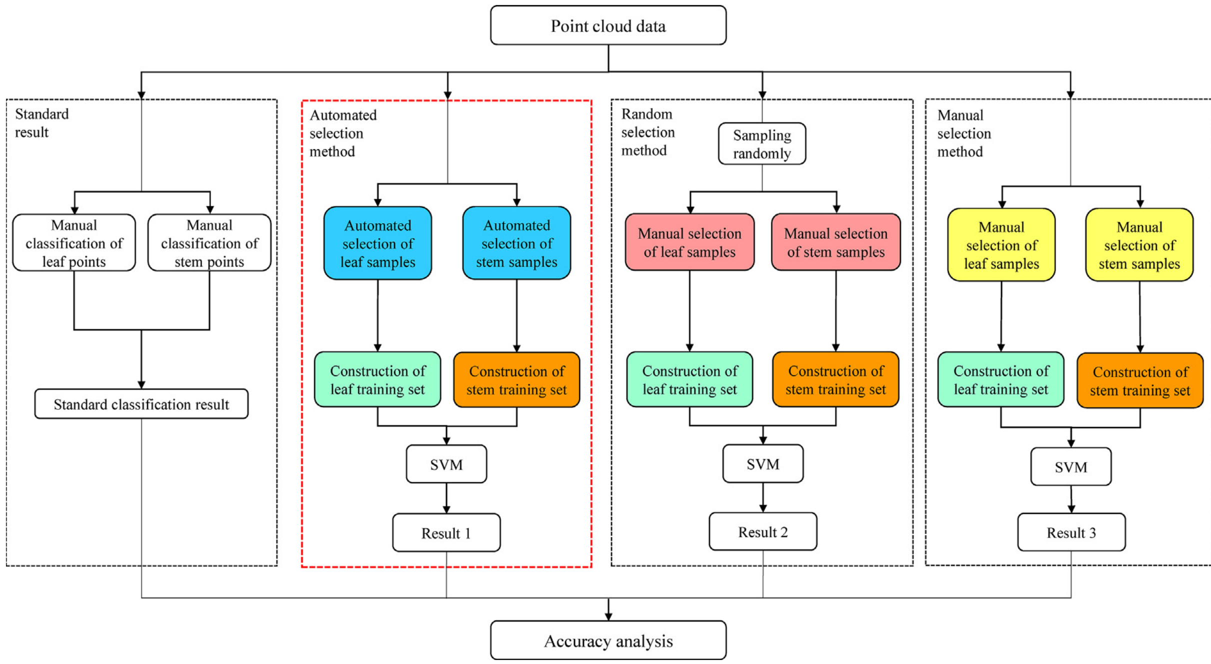

Fig. 4 – Flowchart.

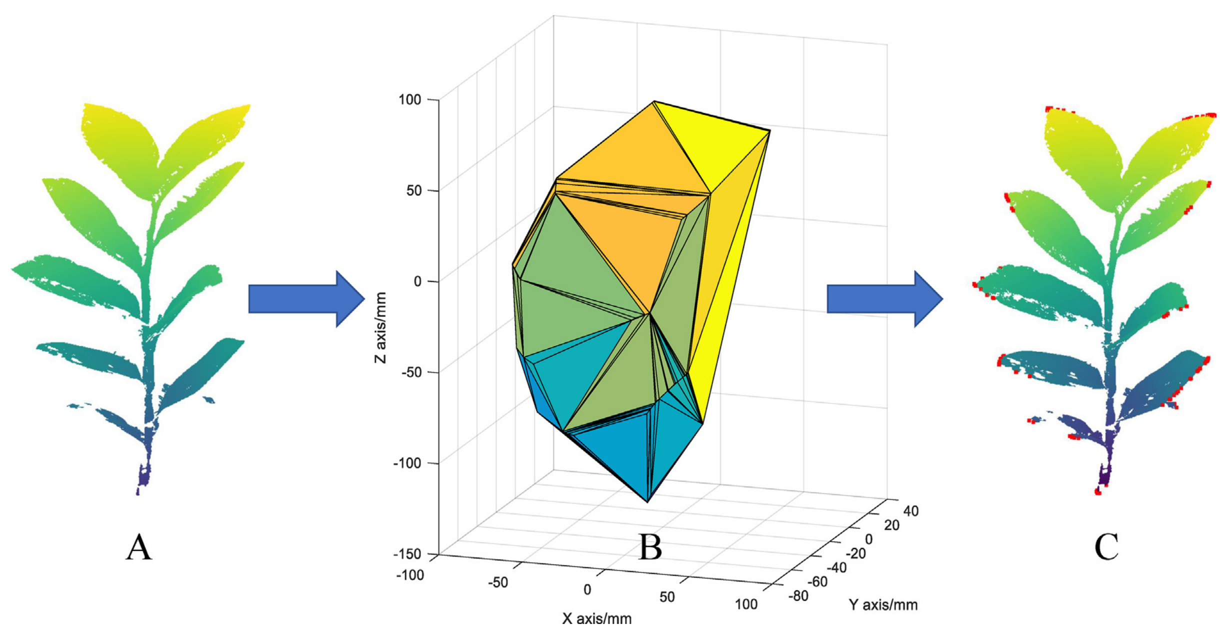

Fig. 5 – Process of selecting leaf sample points automatedly. A: *Z. zamiifolia*. B: 3D convex hull. C: Turning points.

images (Zheng et al., 2013). However, the description of the density of sample points is along the stem's direction which is almost vertical. Therefore orthogonal projection is more suitable and easier to calculate.

Therefore, all the points inside the above-mentioned sphere s_i were projected onto the XOY plane with a 0.1 *mm* grid spacing to obtain the 2D projection. The circumscribed square of the circle on the XOZ plane projected by the sphere s_i was then made, and this square was meshed. Next, the number num_i of grids occupied by the projections of all points inside the sphere s_i was counted, with the projection grid density m_i of each point p_i given as:

$$m_i = \frac{l}{num_i} \tag{1}$$

where l is the number of points in the sphere s_i. Then, some points with the higher grid densities were selected as the stem samples.

Taking the *Z. zamiifolia* as an example, a point a_1 on the stem and a point a_2 on the leaf were selected to demonstrate the above process shown in Fig. 6. The centres of the two spheres are a_1 and a_2, and the radii of the two spheres are both 0.5 *mm*.

As highlighted in Fig. 6, the two projections were obviously different in terms of the point distribution. In the stem sphere, there were 52 points that were projected into 10 grids. However, in the leaf sphere, there were 99 points that were projected into 43 grids. The projection grid density of the stem point a_1 and the leaf point a_2 were 5.2 and 2.3023, respectively, which is a significant difference that was used to discriminate the stem and the leaves in the experiment.

2.2.4. *Construction of training sets*

The leaf and stem training sets can be constructed based on the leaf samples and stem samples respectively.

The leaf training sets were constructed by using the leaf samples. Each leaf sample (x_i, y_i, z_i) was chosen as the centre of a leaf training points sphere, and r_1 was set as the radius. Then all the points inside these spheres were selected as the leaf training set:

$$X_{Leaf} = \{(x,y,z) \times |(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2 < {r_1}^2, (x,y,z) \in D, i = 1, \cdots m\} \tag{2}$$

where D denotes the plant point cloud data set and r_1 denotes the selection radius of the leaf training set.

Similarly, the stem training sets were constructed by using the stem samples. Each stem sample point (x_j, y_j, z_j) was chosen as the centre of a stem training points sphere, and r_2 was set as the radius. Then all the points inside these spheres were selected as the stem training set:

$$X_{Stem} = \left\{(x,y,z) \times \left|(x-x_j)^2 + \left(y-y_j\right)^2 + (z-z_j)^2 < {r_2}^2, (x,y,z) \in D, j = 1, \cdots n\right.\right\} \tag{3}$$

where D denotes the plant point cloud data set and r_2 denotes the selection radius of the stem training set.

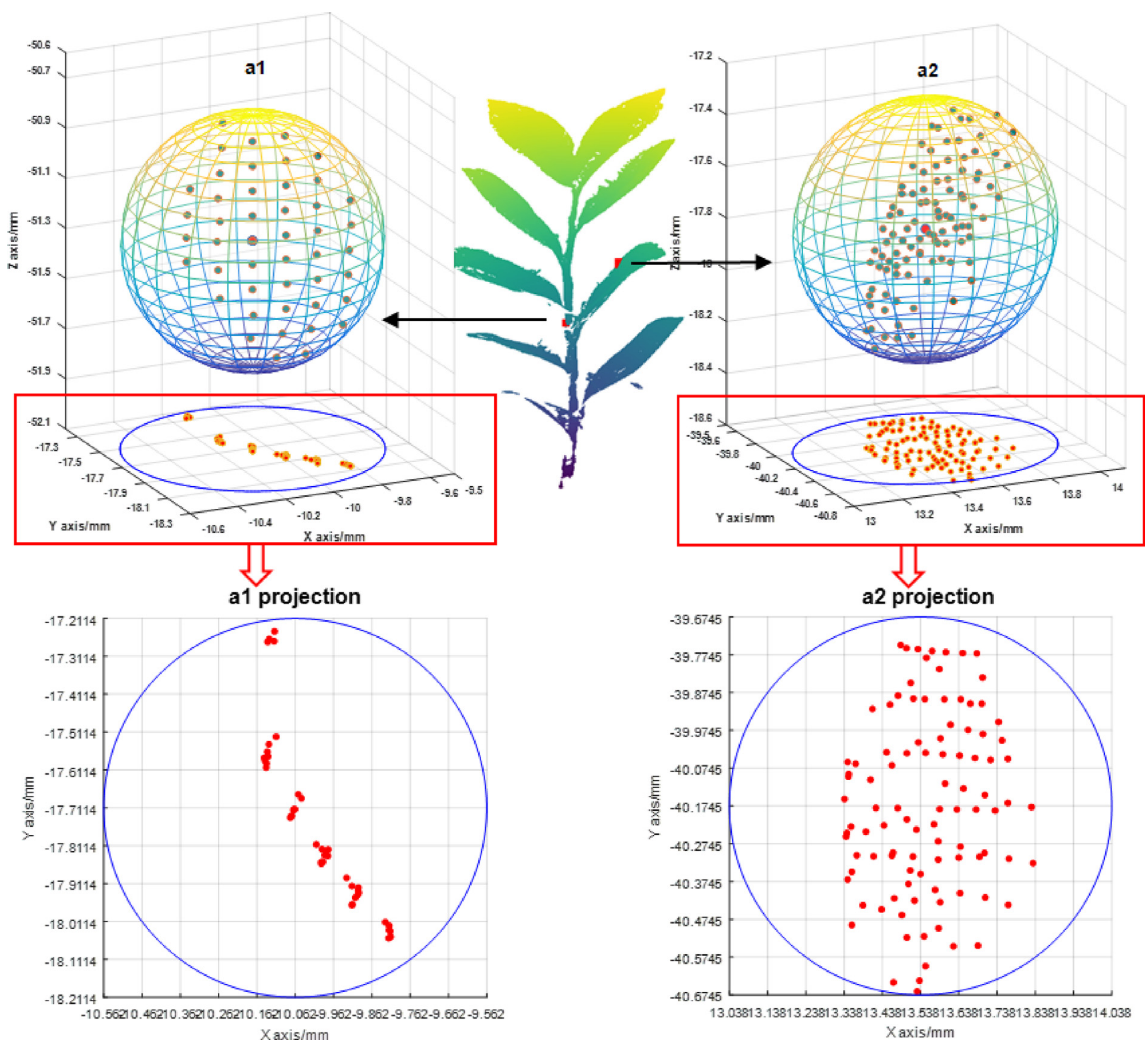

Fig. 6 – Sphere projections on the XOY plane.

2.2.5. SVM classification

Based on the leaf and stem training sets, the point clouds of the potted plants were classified into stem points and leaf points by using the SVM method. In the SVM method, a hyperplane is constructed by the training sets. For nonlinear problems, all the point cloud points were mapped from the original 3D space to another feature space with higher dimensions. Then, all the point cloud points were classified by using the hyperplane in the high-dimensional space.

The mapping from the original 3D space to the high-dimensional space was based on kernel function $K(x_i, x_j)$ (Zhang, 2000). Commonly used kernel functions are the linear kernel, polynomial kernel and radial basis function (RBF) kernel. The RBF kernel function had a better performance than the other two in our experiments. The parameter σ in the RBF kernel function has an impact on the performance of SVM (Alvarsson et al., 2014). If the parameter σ is too small and $\sigma \rightarrow 0$, although the training samples can be classified without errors theoretically, the model will show severe overfitting. If the value of σ is too large and $\sigma \rightarrow \infty$, all the samples will be regarded as one class, which means the model can't classified the samples correctly (Zhou, Liu, & Ye, 2009). In the experiment, the parameter σ was set to 0.15 after many attempts. The leaf and stem training sets X_{Leaf} and X_{Stem} were marked as class1 and class2, respectively, and were put into the SVM classifier for training and classification.

2.2.6. Accuracy analysis method

Since there may be a big difference between the number of leaf points and the number of stem points, the confusion matrix and kappa coefficient were calculated to evaluate the classified results. The confusion matrix can measure the classification accuracy of a classifier, the diagonal of the matrix is the number of points that are correctly predicted and the rest are the number of points with incorrect predictions (Fawcett, 2005). The kappa coefficient is an indicator used for consistency test, the accuracy of classification (Cohen, 1960) and many point cloud classification studies (Ferrara et al., 2018; Tao et al., 2015).

The specific steps of the accuracy analysis are as follows.

First, several evaluation indicators were set to compare the different results in detail:

(1) Standard leaf point set: L_s
(2) Classified leaf point set: L_c.
(3) Standard stem point set: S_s.
(4) Classified stem point set: S_c.
(5) Cardinal number of set A (that is, the number of elements of set A): $card(A)$.

Second, based on the definitions mentioned above and the related definitions of the confusion matrix, the following additional indicators can be set:

True Positives TP, which denotes the number of correctly classified leaf points.

True Negatives TN, which denotes the number of correctly classified stem points.

False Positives FP, which denotes the number of mistakenly classified leaf points.

False Negatives FN, which denotes the number of mistakenly classified stem points.

These indicators can be calculated by using the following equations:

$$TP = card(L_S \cap L_C) \tag{4}$$

$$TN = card(S_S \cap S_C) \tag{5}$$

$$FP = card(L_C) - TP \tag{6}$$

$$FN = card(S_C) - TN \tag{7}$$

Then, the confusion matrices of the three plant classification results obtained by using the three methods can be calculated. Based on the above-mentioned indicators and confusion matrices, the kappa coefficient can be calculated:

$$p_o = \frac{TP + TN}{N} \tag{8}$$

$$p_e = \frac{(TP + FP) \times (TP + TN) + (TN + FN) \times (FP + FN)}{N \times N} \tag{9}$$

$$kappa = \frac{p_o - p_e}{1 - p_e} \tag{10}$$

where N denotes the points number of the point cloud data.

3. Results

3.1. Selected results of leaf samples

The 3D convex hulls of the three plants were calculated based on the point cloud data, and the turning points of these convex hulls were regarded as leaf samples which are shown in Fig. 7. For comparison, the selections of leaf samples using the random selection method and the manual selection method are shown in Fig. 8 and Fig. 9.

The numbers of leaf samples of each method for each plant are listed in Table 3.

3.2. Selected results of stem samples

In the selection of stem samples, 500 points were chosen randomly in a plant point cloud. Then some points would be selected as the stem samples. Twenty points with the highest grid density were selected as stem samples of *Z. zamiifolia* and *D. picta*. Thirty points with highest grid density were selected for *P. bracteosum* because of its thicker stem.

Because of the different characteristics of the three potted plants, we tested three different r values, which is the sphere radius of stem samples described in method section. The results with r values of 5 mm, 10 mm and 15 mm are demonstrated in Fig. 10. Since the result with r value of 5 mm showed a better distribution, the r value was set as 5 mm in the experiment.

The random selected stem samples and the manual selected stem samples are shown in Fig. 11 and Fig. 12, respectively. The numbers of stem samples of each method for each plant are listed in Table 4.

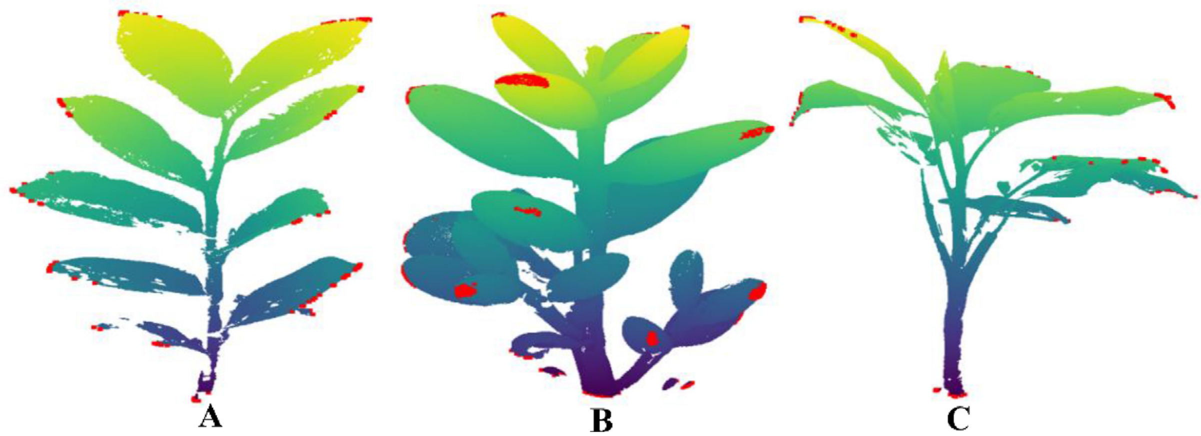

Fig. 7 – Leaf samples of automated selection method. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

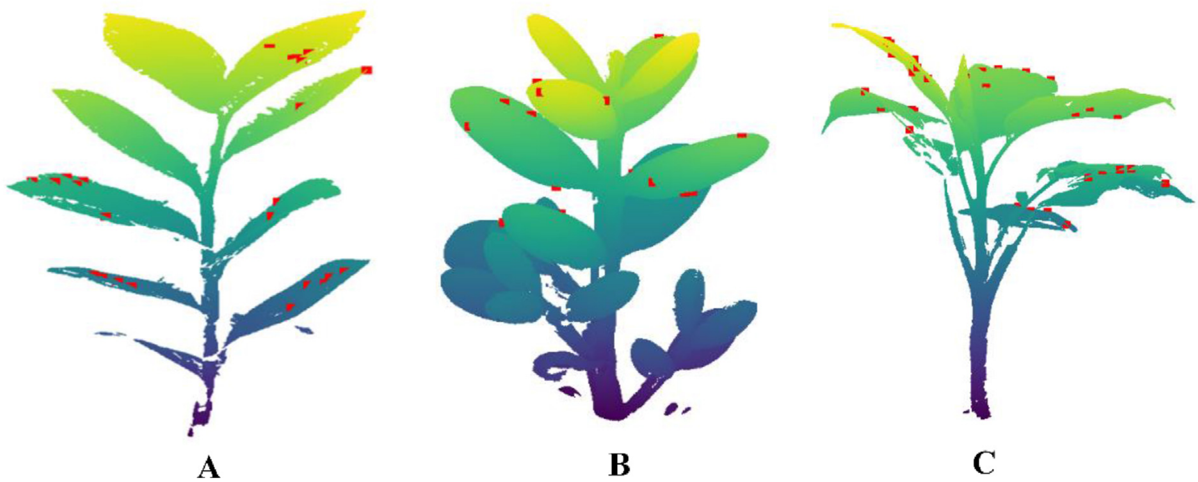

Fig. 8 – Leaf samples of random selection method. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

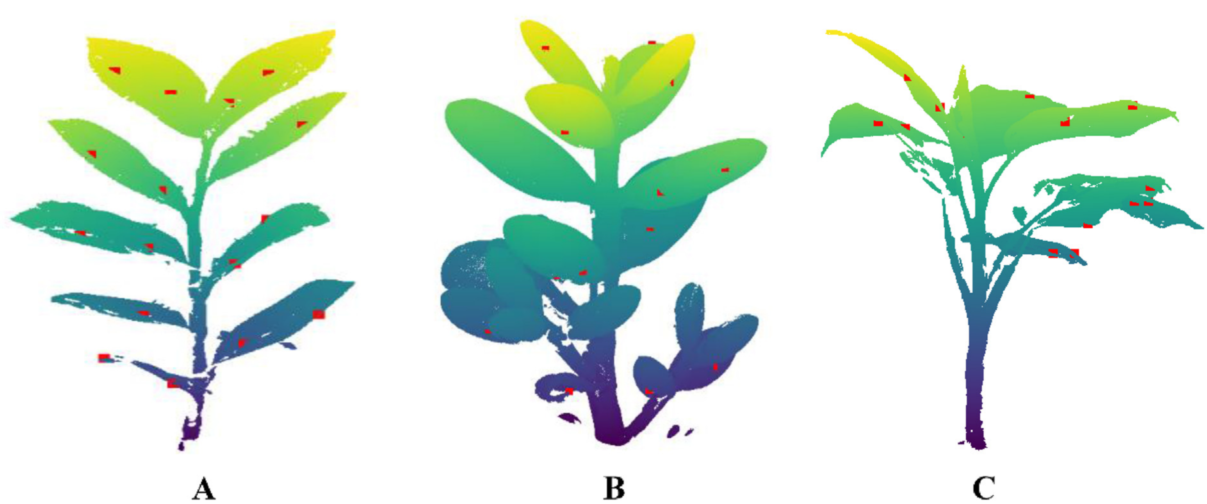

Fig. 9 – Leaf samples of manual selection method. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

Table 3 – The numbers of leaf samples and leaf training points.

Methods	Plant	Number of leaf samples	r_1	Number of leaf training points
Automated selection method	*Z. zamiifolia*	110	0.2 *mm*	544
	P. bracteosum	2373		21860
	D. picta	206		1137
Random selection method	*Z. zamiifolia*	24	0.2 *mm*	232
	P. bracteosum	21		178
	D. picta	33		330
Manual selection method	*Z. zamiifolia*	16	0.2 *mm*	140
	P. bracteosum	19		123
	D. picta	14		122

3.3. *Constructed results of training sets*

The training sets of the automated selection method, the random selection method and the manual selection method were constructed in the same way.

Centring on each leaf sample point, a sphere was constructed with a radius of 0.2 *mm*, which is the r_1 described in the method section. All the points in these spheres were defined the leaf training sets. The number of points in the leaf training set of each method is listed in Table 3.

Similarly, the stem training sets were constructed with a radius of 0.2 *mm*, which is the r_2 described in the method section. The number of points in the stem training set of each method is listed in Table 4.

3.4. *Classification results*

The standard classification results are shown in the first row of Fig. 13, in which the plants are *Z. zamiifolia*, *P. bracteosum* and *D. picta* from left to right. The next three rows of Fig. 13 demonstrate the classified results of three methods.

The detailed numbers of points of classified results are listed in Table 5.

3.5. *Accuracy analysis of results*

First, in terms of accuracy, as shown in Fig. 13, the automated selection method results of *Z. zamiifolia* and *P. bracteosum* were better, and the result of *D. picta* was not as good as the other two plants because of its morphology characteristics.

The confusion matrices (Fig. 14) show the numbers of points that are correctly classified and incorrectly classified. Each submatrix with 2 rows and 2 columns shows the classification result of one plant by using one method. And the colour scale represents the number of points, the lighter the colour is, the more the points there are.

The accuracy of the three methods can be evaluated based on the kappa coefficients (Table 6); the closer the kappa coefficient is to 1, the more accurate is the classification result. The results for the random selection method are not as good as the other two methods. For *Z. zamiifolia*, the manual selection method has the best result with a kappa coefficient of 0.9031 which is close to the kappa coefficient of 0.7825 of the proposed method. For *P. bracteosum*, the proposed method has the kappa coefficient of 0.8316, which was the best result. However, for the *D. picta*, the kappa coefficient of the manual selection method was 0.7038 which was much higher than the kappa coefficients of other two methods.

The running time of the automated selection method (Table 7) indicates its time cost. Due to different numbers of points in the point clouds (Table 5), the running time for different plants is different. However, the running time for other two methods are slow and hard to calculate because they both have the random and manual processes. So, the automated selection method is better on time cost because of its totally automated sample selection.

4. Discussion

According to the results of the three methods, the proposed method had the best performance on the running time, and a better performance on accuracy. The random selection method is not good enough on accuracy or time-cost. The manual selection method is also time-consuming, but had the best overall performance on accuracy, especially on *D. picta*. On further observation, *D. picta* has a thinner and shorter stem than the other two plants, and some leaves of *D. picta* are shaped like stems. These characteristics may affect the classification and decrease the accuracy of the proposed method and the random selection method. Of the three methods, the proposed method is automated and most efficient, with a better performance.

Yet, the proposed method still has some limitations. First, it has a requirement for integrity of the data, that is, the robustness of the algorithm still needs to be improved. For example, the thick stem of the *P. bracteosum* accounts for a larger proportion of the total volume. When the stem samples were not distributed on the entire stem very well, the classification accuracy may decrease. Second, some leaves with special shapes will interfere with the algorithm and affect the classification accuracy. For example, a slender leaf in the middle part of *D. picta* is shaped like a stem, and was finally classified as stem (Fig. 15) and the kappa coefficient dropped significantly. Third, occlusion in the scanning results in incomplete data which will hamper the selection of sample points. And many overlapped leaves may increase the point density in the projection, which can prevent the correct decision between leaf samples and stem samples. Finally, the proposed method is more suitable for plants with unbent stems and distinguishable leaves, and the classification accuracy will decrease on processing the curved stems and

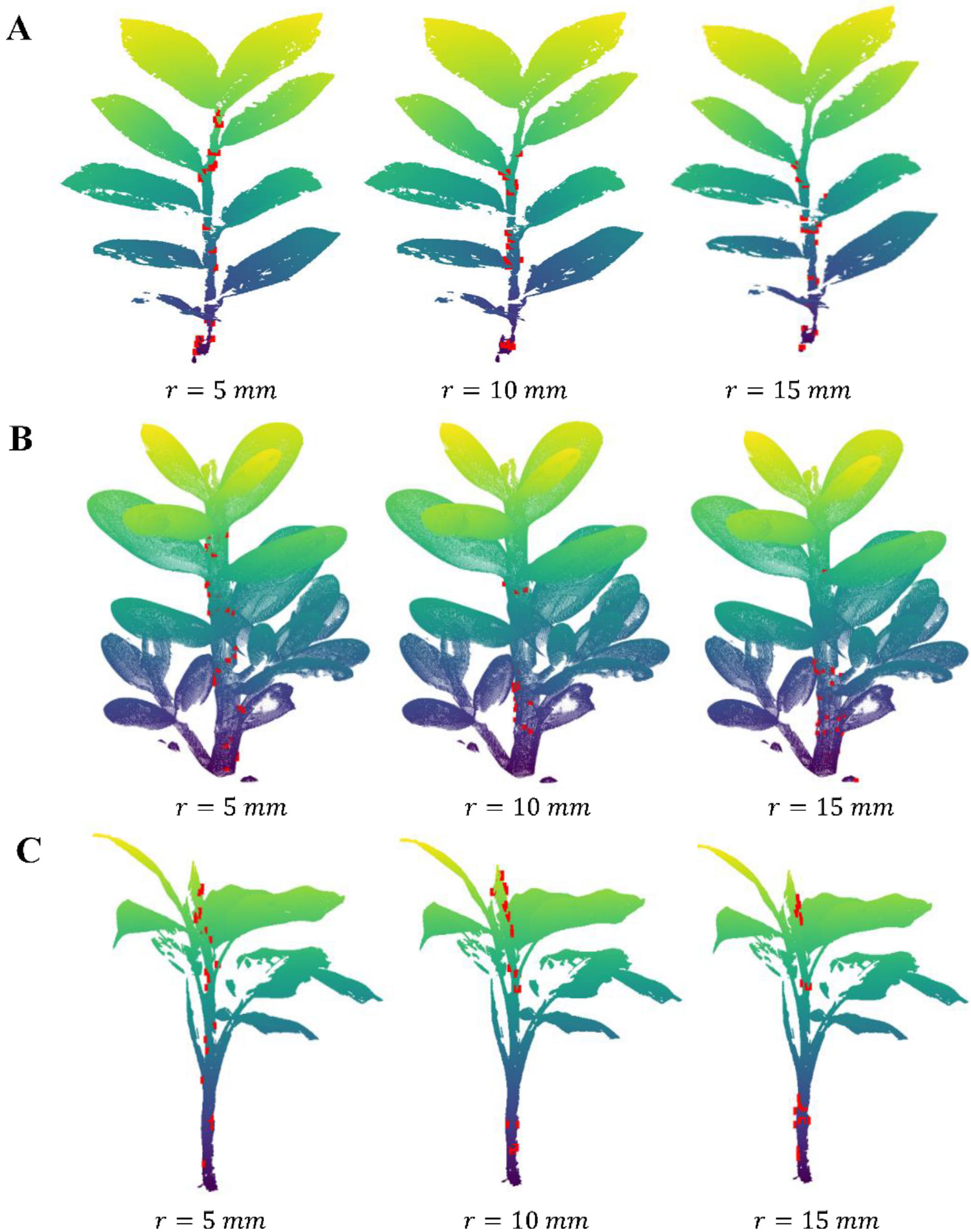

Fig. 10 – Stem samples selected by using different values of *r*. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

dense leaves. All the above problems may cause a decrease in classification accuracy.

In order to verify the interference caused by the leaves with special shapes and data deficiency, the slender leaf in the middle part of *D. picta* was deleted. The new point cloud of *D. picta* was processed again using the three methods. The new classification results and the new kappa coefficients are shown and listed in Fig. 16 and Table 8, respectively. The kappa coefficients of three methods have increased significantly. The kappa coefficient of the proposed method improved the most and is close to the manual selection method.

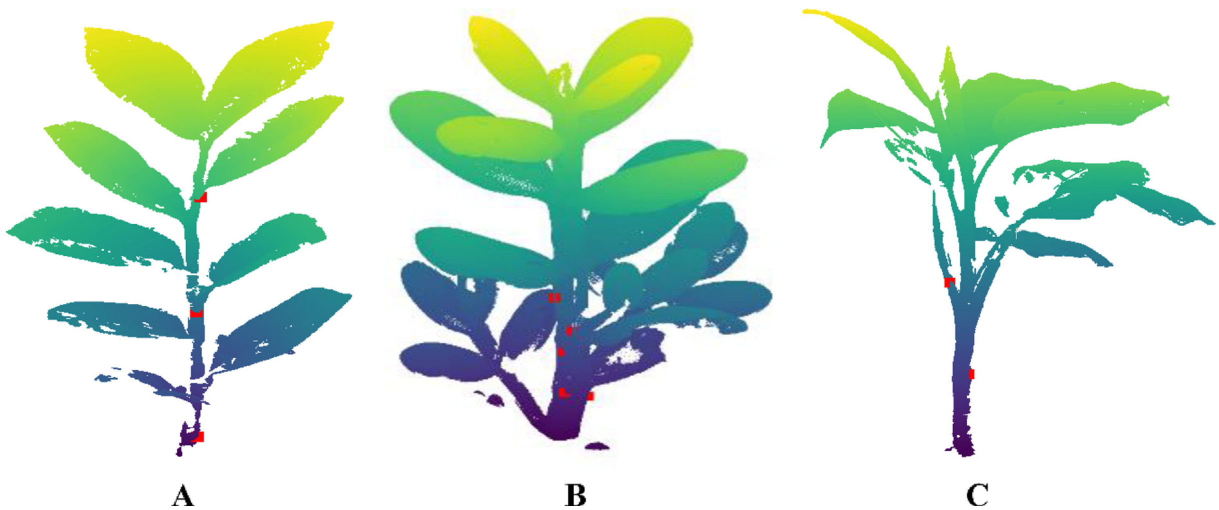

Fig. 11 – Stem samples of random selection method. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

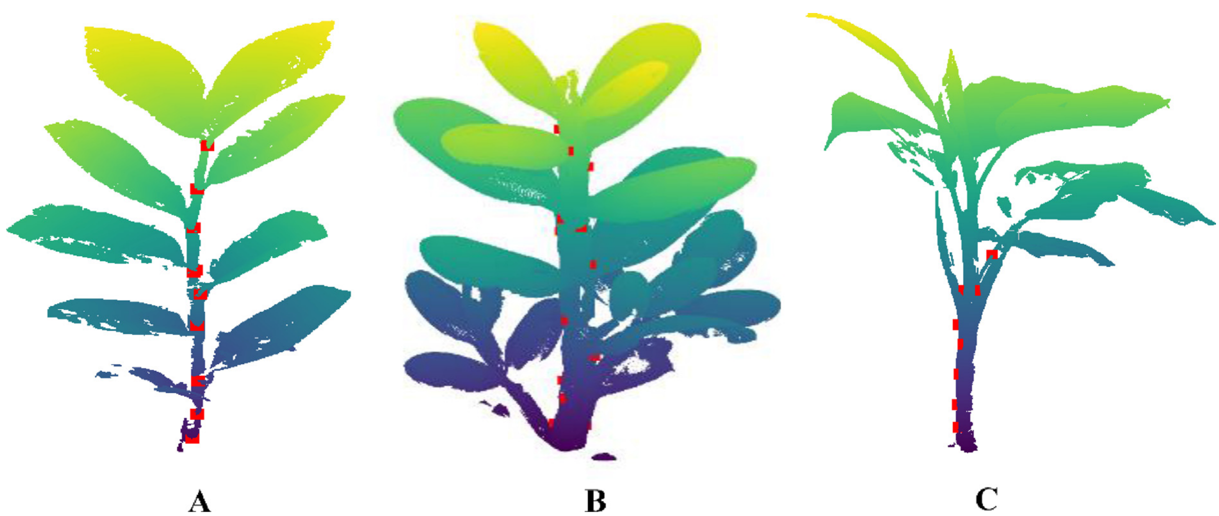

Fig. 12 – Stem samples of manual selection method. A: *Z. zamiifolia*. B: *P. bracteosum*. C: *D. picta*.

Table 4 – The numbers of stem samples and stem training points.

Methods	Plant	Number of stem samples	r_2	Number of stem training points
Automated selection method	*Z. zamiifolia*	20	0.2 *mm*	201
	P. bracteosum	30		280
	D. picta	20		143
Random selection method	*Z. zamiifolia*	3	0.2*mm*	28
	P. bracteosum	5		41
	D. picta	2		12
Manual selection method	*Z. zamiifolia*	10	0.2 *mm*	70
	P. bracteosum	23		180
	D. picta	8		36

Some similar studies have reported accuracy of classification. Paulus et al. (2013) used plant histological surface histograms to classify individual plant organs. The correct rates of classification are from 93.2% to 98.3%. Ferrara et al. (2018) classified wood and non-wood points in point cloud data and the reported kappa coefficients with different thresholds were from 0.75 to 0.88. Tao et al. (2015) separated wood points and leaf points and the reported kappa coefficients were from 0.79 to 0.89.

Compared with these above-mentioned methods, the proposed method can achieve similar performance in accuracy. The correct rates of classification (which were calculated by equation (8)) of the proposed method are 97.13%, 97.16% and 93.16% for *Z. zamiifolia*, *P. bracteosum* and *D. picta*,

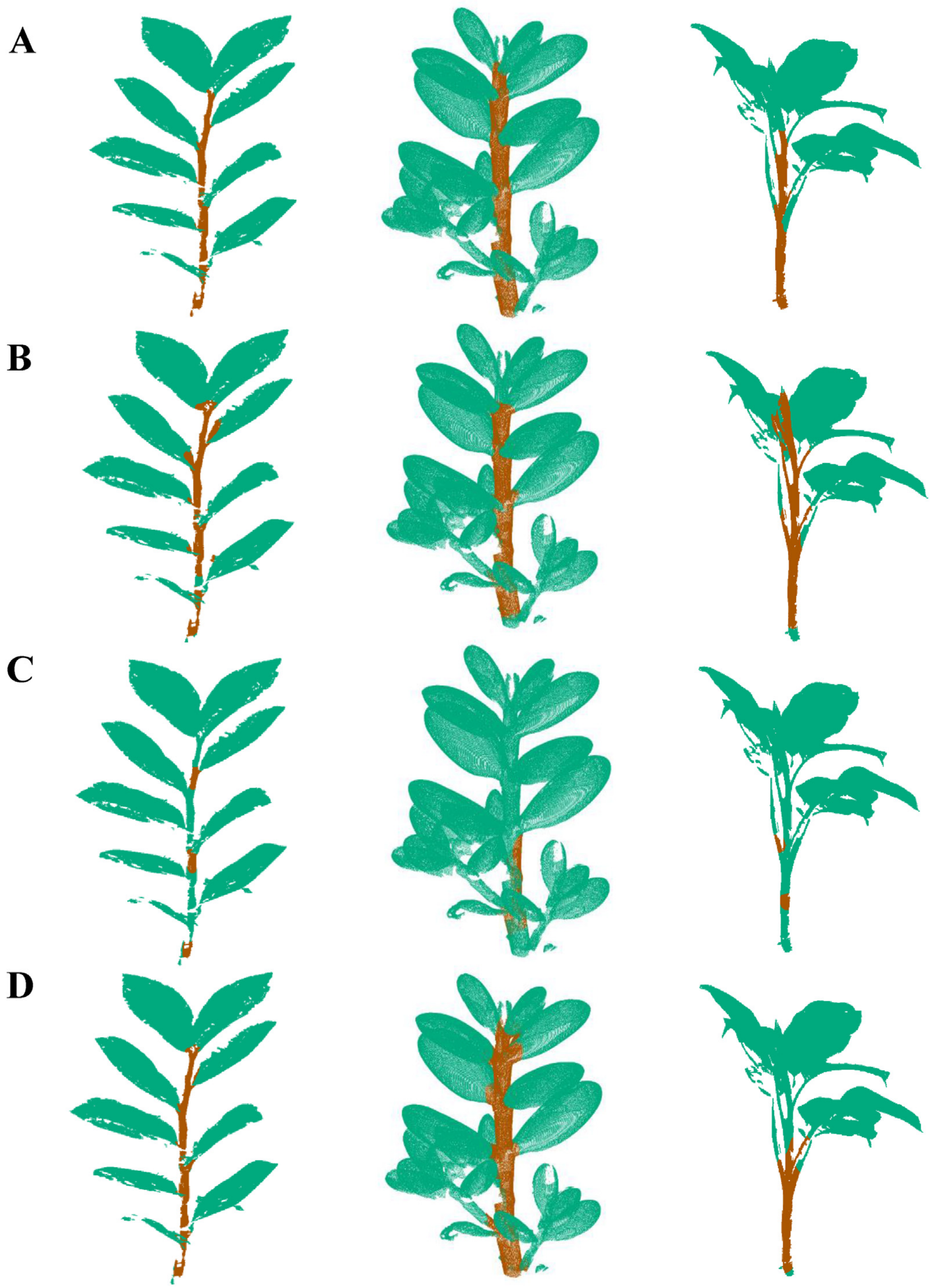

Fig. 13 – Classification results of standard classification (A), automated selection method (B), random selection method (C) and manual selection method (D) of *Z. zamiifolia*, *P. bracteosum* and *D. picta* (from left to right).

respectively, which are similar with the results of Paulus's method. Yet the kappa coefficients are more convincing and accurate for the data with high difference of numbers of different classes. The kappa coefficients in our study are from 0.78 to 0.83, which is at the same level as Ferrara's method and Tao's method.

There are two possible reasons for the efficiency of the proposed method. First, automated samples selection methods for stem and leaves reduce the time of manual feature extraction. Second, the point clouds of potted plants are relatively small size, which cost less processing time than the large data set.

Table 5 – The points numbers of classification results by using three methods.

Methods	*Z. zamiifolia*		*P. bracteosum*		*D. picta*	
	Number of leaf points	Number of stem points	Number of leaf points	Number of stem points	Number of leaf points	Number of stem points
Standard result	982236	61984	789775	81802	2601906	89625
Automated selection method	958139	86081	791318	80259	2428212	263319
Random selection method	1025454	18766	852936	18641	2678107	13424
Manual selection method	973135	71085	757831	113746	2614521	77010

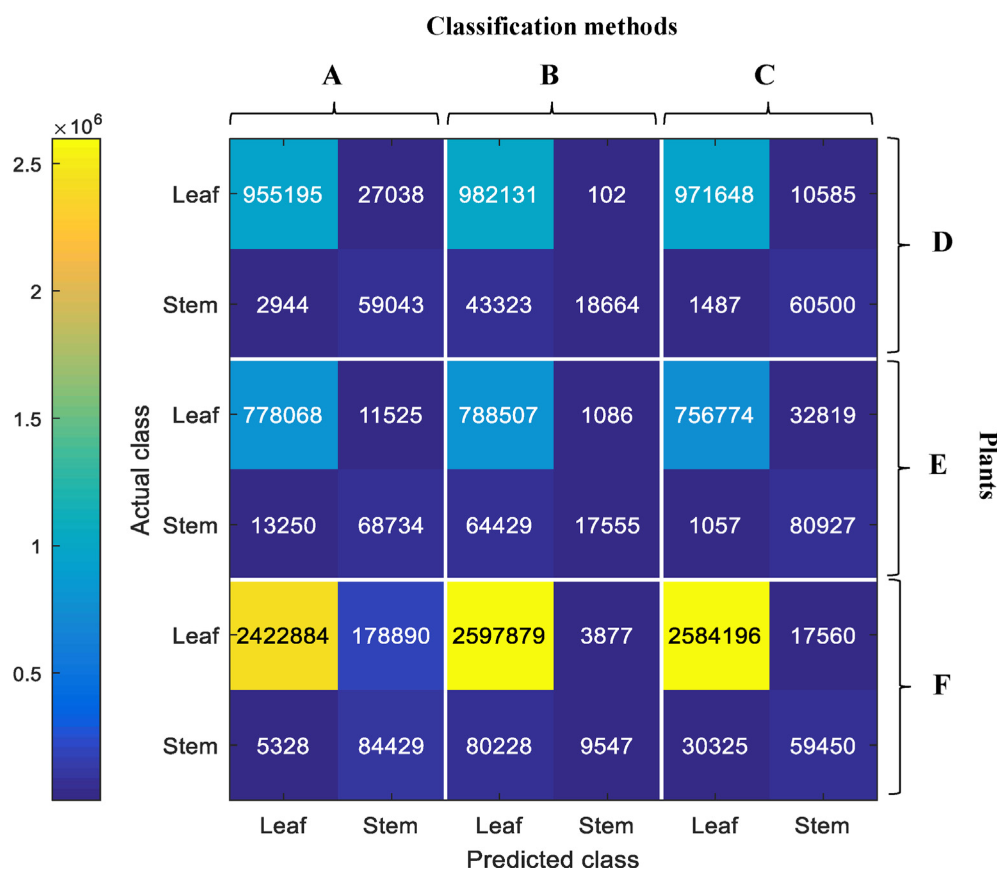

Fig. 14 – Confusion matrices of the experiments. Column: A. Automated selection method. B. Random selection method. C. Manual selection method. Row: D. *Z. zamiifolia*. E. *P. bracteosum*. F. *D. picta*. (The color scale represents the number of points.)

Table 6 – Kappa coefficients of the three methods.

Plant	Kappa coefficient		
	Automated selection method	Random selection method	Manual selection method
Z. zamiifolia	0.7825	0.4470	0.9031
P. bracteosum	0.8316	0.3254	0.8075
D. picta	0.4509	0.1779	0.7038

Table 7 – Time cost for classification of *Z. zamiifolia*, *P. bracteosum* and *D. picta* by using the automated selection method.

Plant	Time cost for Automated selection method
Z. zamiifolia	20.380 s
P. bracteosum	29.180 s
D. picta	30.186 s

There are also some possible reasons for the accuracy of the proposed method. First, automated processing avoids the error of manual selection. Second, the scanner used in our experiment is an accurate desktop scanner which helps to ensure the point cloud data is of good quality.

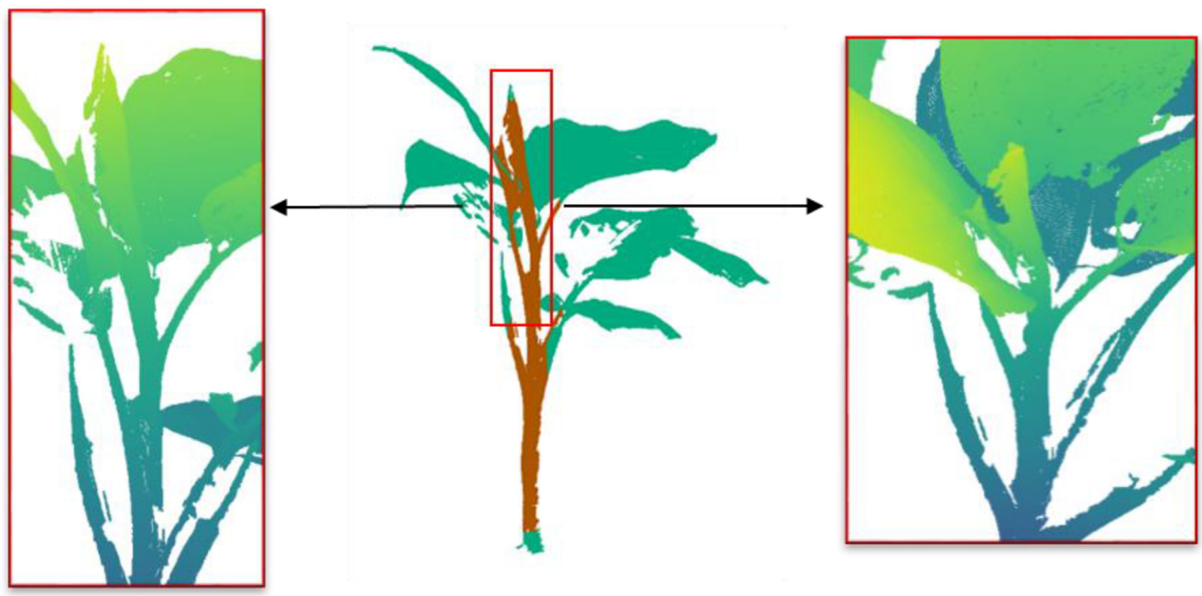

Fig. 15 – The uncertain part of *D. picta*.

Fig. 16 – New classification results of *D. picta* obtained by using the three methods. A. Standard classification results. B. The result of automated selection method. C. The result of random selection method. D. The result of manual selection method.

Table 8 – Kappa coefficients of the three methods with the processed *D. picta* data.

Plant	Kappa coefficient		
	Automated selection method	Random selection method	Manual selection method
D. picta	0.7851	0.3348	0.9210

In general, the proposed method can effectively and accurately classify stems and leaves of plants with good shapes. Furthermore, the proposed method can automatically select the samples to reduce subjective influence.

5. Conclusions

The point cloud data of plants provide a non-destructive solution for collecting, monitoring and analysing the status of plants and their organs. This study proposed a feasible and automated method for classifying the plant point cloud into leaves and stems. The training sets of the two classes can be automatically identified by using the 3D convex hull and the 2D projection distribution of the point cloud data. The results show that the proposed method has an advantage in overall performance for both accuracy and time-cost. The proposed method has the potential to improve the efficiency of research on monitoring or analysing plants. Future work could focus on improving the classification accuracy and the robustness via application to different plants.

CRediT author statement

Zichu Liu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Qing Zhang: Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. Pei Wang: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. Zhen Li: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft. Huiru Wang: Data Curation, Visualization.

Funding

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2015ZCQ-LY-02); the Fundamental Research Funds for the Central Universities (BLX201928); the State Scholarship Fund from China Scholarship Council (CSC No. 201806515050).

Declaration of competing interest

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

Acknowledgments

Thanks to the editor and the reviewers for their constructive comments and suggestions which help us to improve the manuscript.

REFERENCES

Alvarsson, J., Eklund, M., Andersson, C., Carlsson, L., Spjuth, O., & Wikberg, J. (2014). Benchmarking study of parameter variation when using signature fingerprints together with support vector machines. *Journal of Chemical Information and Modeling, 54*(11), 3211–3217.

Caccamo, G., Iqbal, I. A., Osborn, J., Bi, H., Aekley, K., Melville, G., et al. (2018). Comparing yield estimates derived from LiDAR and aerial photogrammetric point-cloud data with cut-to-length harvester data in a Pinus radiata plantation in Tasmania. *Australian Forestry, 81*, 131–141.

Claret, A. R., & Day, A. M. (1994). Experiments in the parallel computation of 3D convex hulls. *Computer Graphics Forum, 13*(1), 21–36.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Conde, L. (2011). Effects of phosphorus availability on plant growth and soil nutrient status in rice/soybean rotation system on new cultivated acid soils. *American Journal of Agriculture and Forestry, 2*, 309. https://doi.org/10.11648/j.ajaf.20140206.23

Davi, H., Dufrêne, E., Granier, A., Dantec, V. L., Barbaroux a, C., Françoisa, C., et al. (2005). Modelling carbon and water cycles in a beech forest: Part II.: Validation of the main processes from organ to stand scale. *Ecological Modelling, 185*, 387–405.

Dey, D., Mummert, L., & Sukthankar, R. (2012). Classification of plant structures from uncalibrated image sequences. In *Proceedings of IEEE workshop on applications of computer vision (WACV) 2012* (pp. 329–336). Piscataway: IEEE.

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874.

Ferrara, R., Virdis, S. G. P., Ventura, A., Ghisu, T., Duce, P., & Pellizzaro, G. (2018). An automated approach for wood-leaf separation from terrestrial LIDAR point clouds using the density based clustering algorithm DBSCAN. *Agricultural and Forest Meteorology, 262*, 434–444.

Fitter, A. H. (1987). An architectural approach to the comparative ecology of plant root systems. *New Phytologist, 106*(Supplement s1), 61–77.

Frasson, R. P. M., & Krajewski, W. F. (2010). Three-dimensional digital model of a maize plant. *Agricultural and Forest Meteorology, 150*, 478–488.

Hang, L., Lie, T., Steven, A. W., & Yu, M. (2017). A robotic platform for corn seedling morphological traits characterization. *Sensors, 17*(9).

Hétroy-Wheeler, F., Casella, E., & Boltcheva, D. (2016). Segmentation of tree seedling point clouds into elementary units. *International Journal of Remote Sensing, 37*, 2881–2907.

Huang, H. M., Zhao, Y. L., Xu, Z. G., Zhang, W., & Jiang, K. K. (2019). Physiological responses of Broussonetia papyrifera to manganese stress, a candidate plant for phytoremediation. *Ecotoxicology and Environmental Safety, 181*, 18–25.

Maltamo, M., Korpela, I., Tokola, T., Hyyppa, J., & Orka, H. O. (2010). Tree species classification using airborne LiDAR - effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica, 44*, 319–339.

Ma, L., Zheng, G., Eiter, J. U. H., Magney, T. S., & Moskal, L. M. (2016). Determining woody-to-total area ratio using terrestrial laser scanning (TLS). *Agricultural and Forest Meteorology, 228/229*, 217–228.

Méndez, V., Rosell-Polo, J. R., Pascual, M., & Escolà, A. (2016). Multi-tree woody structure reconstruction from mobile terrestrial laser scanner point clouds based on a dual neighbourhood connectivity graph algorithm. *Biosystems Engineering, 148*(148).

Morsdorf, F., Allgower, B., Hetherington, D., Danson, F. M., & Koetz, B. (2007). Forest canopy gap fraction from terrestrial laser scanning. *IEEE Geoscience and Remote Sensing Letters, 4*, 157–160.

Paulus, S., Dupuis, J., Mahlein, A. K., & Kuhlmann, H. (2013). Surface feature based classification of plant organs from 3D laser scanned point clouds for plant phenotyping. *BMC Bioinformatics, 14*, 1–12.

Sodhi, P., Vijayarangan, S., & Wettergreen, D. (2017). *In-field segmentation and identification of plant structures using 3D imaging. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.

Su, W., Zhu, D., Huang, J., & Guo, H. (2018). Estimation of the vertical leaf area profile of corn (Zea mays) plants using terrestrial laser scanning (TLS). *Computers and Electronics in Agriculture, 150*, 5–13.

Tao, S., Guo, Q., Xu, S., Su, Y., Li, Y., & Wu, F. (2015). A geometric method for wood-leaf separation using terrestrial and simulated lidar data. *Photogrammetric Engineering & Remote*

Sensing: Journal of the American Society of Photogrammetry, 81(10), 767–776.

Wahabzada, M., Paulus, S., Kersting, K., & Anne-Katrin, M. (2015). Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. *BMC Bioinformatics, 16*, 1–11.

Williams, L. E., & Ayars, J. E. (2005). Grapevine water use and the crop coefficient are linear functions of the shaded area measured beneath the canopy. *Agricultural and Forest Meteorology, 132*, 201–211.

Xiang, L., Bao, Y., Tang, L., Ortiz, D., & Salas-Fernandez, M. G. (2019). Automated morphological traits extraction for sorghum plants via 3D point cloud data analysis. *Computers and Electronics in Agriculture*, 162.

Yun, T., An, F., Li, W., Sun, Y., Cao, L., & Xue, L. (2016). A novel approach for retrieving tree leaf area from ground-based LiDAR. *Remote Sensing, 8*(11), 942.

Yun, T., Gao, M. X., Wang, Y. M., & Liu, X. F. (2013). Computational-geometry-based plant organs classification and foliage 3D reconstruction from point cloud data. *Sensors & Transducers, 156*, 217–226.

Zhang, X. G. (2000). Introduction to statistical learning theory and support vector machines. *Acta Automatica Sinica, 26*, 32–41.

Zheng, G., & Moskal, L. M. (2012). Computational-geometry-based retrieval of effective leaf area index using terrestrial laser scanning. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 3958–3969.

Zheng, G., Moskal, L. M., & Kim, S.-H. (2013). Retrieval of effective leaf area index in heterogeneous forests with terrestrial laser scanning. *IEEE Transactions on Geoscience and Remote Sensing, 51*(2), 777–786.

Zhou, S., Liu, H., & Ye, F. (2009). Variant of Gaussian kernel and parameter setting method for nonlinear SVM. *Neurocomputing, 72*(13), 2931–2937.