Research Article

Automatic Segmentation for Plant Leaves via Multiview Stereo Reconstruction

Jingwei Guo and Lihong Xu

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

Correspondence should be addressed to Lihong Xu; xulhk@163.com

This paper presented a new method for automatic plant point cloud acquisition and leaves segmentation. Quasi-dense point cloud of the plant is obtained from multiview stereo reconstruction based on surface expansion. In order to overcome the negative effects from complex natural light changes and to obtain a more accurate plant point cloud, the Adaptive Normalized Cross-Correlation algorithm is used in calculating the matching cost between two images, which is robust to radiometric factors and can reduce the fattening effect around boundaries. In the stage of segmentation for each single leaf, an improved region growing method based on fully connected conditional random field (CRF) is proposed to separate the connected leaves with similar color. The method has three steps: boundary erosion, initial segmentation, and segmentation refinement. First, the edge of each leaf point cloud is eroded to remove the connectivity between leaves. Then leaves will be initially segmented by region growing algorithm based on local surface normal and curvature. Finally an efficient CRF inference method based on mean field approximation is applied to remove small isolated regions. Experimental results show that our multiview stereo reconstruction method is robust to illumination changes and can obtain accurate color point clouds. And the improved region growing method based on CRF can effectively separate the connected leaves in obtained point cloud.

1. Introduction

In greenhouse crop cultivation, the three-dimensional structures of plants convey the real-time continuous information of the whole process of crop growth in a most direct way. Also, it is the most intuitive feature for plant phenotype analysis [1]. Therefore acquiring the 3D structure of plants by an automated, high-throughput, accurate, and rapid method not only helps to monitor crop growth, but also provides visual basis for optimizing the intelligent greenhouse environment control system; and finally, it has a significant meaning for plant phenotypic and genome-environment-phenotype synergistic analysis [2–6]. Owing to noninvasive and high-throughput characteristics, automated 3D-digitization technology has greatly promoted the accurate modeling of 3D structures of plants.

To date, there are several approaches for 3D structure reconstruction. For example, Microsoft Kinect provides depth images by utilizing coded infrared light patterns.

However, Kinect has limited range and can only be used indoors. Laser Scanning can obtain highly accurate and complete 3D structure but the devices are quite expensive for agricultural usage. Therefore the passive stereo vision technology, which puts low requirements on equipment and environment [7–10], has received wide attention. Biskup et al. [8] used a binocular stereo vision system to measure the structural parameters of plant canopies.

Binocular stereo vision system needs to manually calibrate the camera and its imaging angle is relatively fixed. Hence the method is sometimes inefficient and it is difficult for it to reveal the complex 3D structure of plants. Compared to binocular stereo, the multiview stereo reconstruction based on an image sequence is used more widely because it can perform automatic calibration of the camera and generate more intact 3D structure information. In the literature, multiview stereo algorithms are roughly categorized into three classes. The first class operates by first computing a cost function on a 3D volume and then extracting a surface
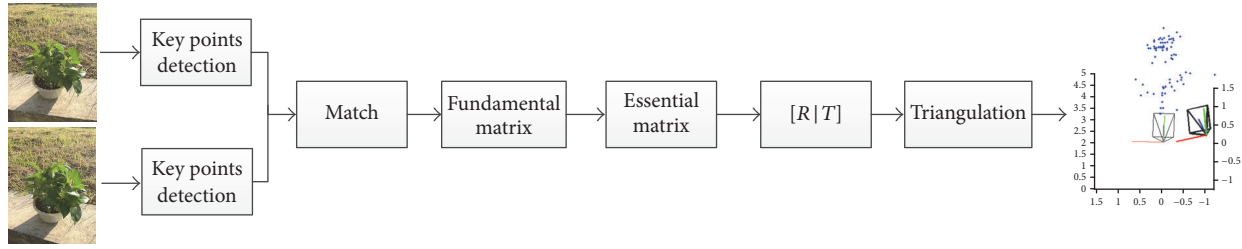
Figure 1: A simplified overview of the SFM process.

from this volume [11–14]. The second class is an image-space methodology that operates by fusing multiple depth maps [15–18]. The third class is surface expansion method which builds a quasi-dense point cloud from a set of matches [19–22]. The accuracy of volume-based method depends on the density of the grid, and therefore the computational and memory costs of volumetric grids are very high. The following two methods are both able to produce good results. In this paper, we propose a multiview stereo reconstruction method based on surface expansion which iteratively expands a sparse set of initial matches into a quasi-dense point cloud that represents surfaces of the scene. In order to overcome the negative effects from complex natural light changes and obtain more accurate plant point clouds, the Adaptive Normalized Cross-Correlation algorithm (ANCC) [23] is used in calculating the matching cost between points of interest in two images, which is robust to radiometric factors such as illumination direction, illuminant color, and imaging device changes. The ANCC method also can reduce the fattening effect that object boundaries are not reconstructed correctly; and this effect clearly deteriorates results of the Zero-mean Normalized Cross-Correlation (ZNCC)—the method used in comparison.

We then segment the plant leaves after getting the plant point cloud from the above multiview stereo reconstruction method. Generally, the segmentation of 3D point cloud can be roughly categorized into five classes [24]: edge-based methods, region-based methods, attributes-based methods, model-based growing methods, and graph-based methods. References [25–27] segmented a plant into two sets, one for the stems and the other for the leaves according to surface feature histograms, followed by a classification or a clustering method. Yin et al. [28] proposed an approach to accurately segment and reconstruct plant. However their approach requires manually cutting the plant leaves and scanning each leaf individually. Hétroy-Wheeler et al. [29] used spectral clustering to realize the segmentation of leaf point cloud, but it needed to manually specify the number of leaves before segmentation.

Based on the characteristics of the subject plants, an improved region growing method based on conditional random field (CRF) is proposed to segment the leaf point cloud. In a canopy point cloud, there may exist some connections between leaves, which easily disturb the segmentation. Therefore we perform some erosion operations around the leaf edge before segmentation. In the initial stage of segmentation, considering that different leaves usually have different surface normals and the points on the same leaf surface usually have similar normals, we use region growing algorithm based on local surface normal and curvature to segment leaves [30]. Usually the segmentation result contains some small isolated regions; therefore a fully connected CRF model defined with the complete set of points is built to further classify these regions to where they ought to belong. A highly efficient approximate inference algorithm based on mean field approximation is applied to the CRF model in which the pairwise edge potentials are defined by Gaussian kernel [31].

2. Multiview Stereo Reconstruction

The proposed multiview stereo approach consists of two steps: structure from motion (SFM: from images to sparse points) and dense matching (from sparse points to quasi-dense points). The SFM step is to obtain 3D sparse feature points location of the scene and camera poses (location and orientation). In the step of dense matching, we use the ANCC algorithm as a robust and accurate correspondence measure to match points of interest, which is robust to radiometric factors and can reduce the fattening effect.

2.1. Structure from Motion. An overview of the SFM process is presented in Figure 1. The first step is to find feature points in each image, where the SIFT key point detector [32] is used because of its invariance to image transformations. Other feature detectors could also be used instead. For instance, several detectors were evaluated in [33]. In addition to the key point locations themselves, SIFT provides a local descriptor for each key point. For each pair of images, key point descriptors are matched by using the approximate nearest neighbors package [34]. Then we robustly estimate a fundamental matrix for the pair using RANSAC. During each iteration of RANSAC, a candidate fundamental matrix is computed using the eight-point algorithm [35], followed by nonlinear refinement based on Levenberg–Marquardt algorithm. Afterwards, we convert the fundamental matrix to the essential matrix and then obtain camera poses [35]. At last, the 3D positions of features points can be acquired through the triangulation calculation. (The scale factor is not considered during triangulation calculation because in the following stage leaf segmentation can be performed without the knowledge of a scale factor.) [35].

After generation of the point cloud of features points between each pair of images, we should merge these point
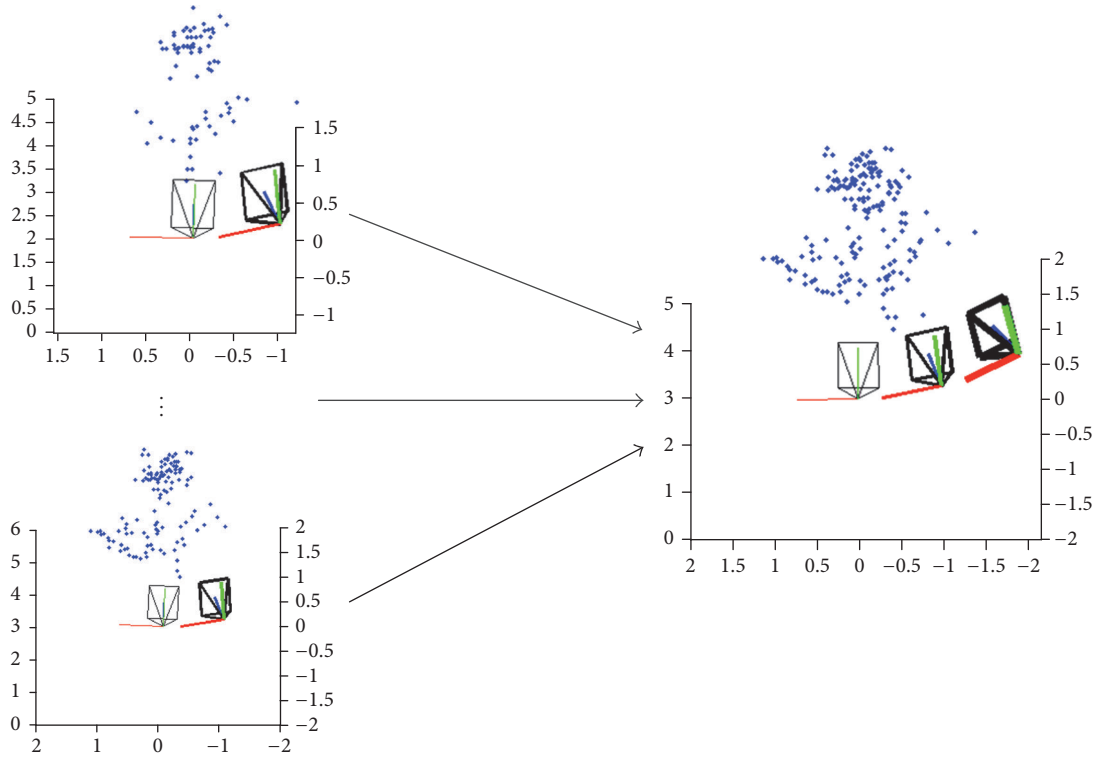
FIGURE 2: Merging point clouds in one coordinate system.

clouds because they are not in the same coordinate system, as shown in Figure 2. Finally, we get 3D positions of all the feature points in a single coordinate system with respect to all point clouds.

2.2. Dense Matching. In this stage, we expand a sparse set of initial matches into a quasi-dense point cloud representing the surfaces of the scene based on the best-first expansion strategy by Lhuillier and Quan [36], which is robust to initial seed match outliers and is efficient in terms of spatial-temporal complexity. In order to overcome the negative effects from complex natural light environments and to obtain a more accurate plant point cloud, the ANCC algorithm is utilized in the process as the matching cost to match points of interest instead of ZNCC. It is noteworthy that the ANCC method is robust to radiometric factors such as illumination direction, illuminant color, and imaging device changes. Moreover, it can reduce the fattening effect from which the Zero-mean Normalized Cross-Correlation (ZNCC) suffers. The best-first expansion strategy is presented in the table of Algorithm 1.

In Algorithm 1, $s(u)$ is a difference-based confidence measure to avoid propagating too uniform areas. The definition is

$$s(x) = \max\{|I(x + \Delta) - I(x)|, \Delta$$
$$\in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}\}. \tag{1}$$

$I(x)$ means the pixel grayscale value at x. $N(x, x')$ means the neighbors of the seed. Let $N(x) = \{u, u - x \in [-N, N]^2\}$, $N(x') = \{u', u' - x' \in [-N, N]^2\}$ denote all

```
Input: Seed
Output: Map
(1) Map ← ф
(2) while Seed ! = ф do
(3)     pull the ANCC-best match (x, x') from Seed
(4)     Local ← ф
(5)     for each (u, u') in N(x, x') do
(6)         if s(u) > t and s(u') > t and ANCC(u, u') > z
(7)         then store (u, u') in Local
(8)         end-if
(9)     end-for
(10)     while Local ! = ф
(11)         pull the ZNCC-best match (u, u') from Local
(12)         if (u, ∗) and (∗, u') are not in Map
(13)         then store (u, u') in Map and Seed
(14)         end-if
(15)     end-while
(16) end-while
```

ALGORITHM 1: The multiview best-first match expansion algorithm.

$(2N + 1) \times (2N + 1)$ neighboring pixels of pixels x and x', respectively. The possible matches limited by the discrete 2D disparity gradient are given as

$$N(x, x') = \{(u, u'), u \in N(x), u'$$
$$\in N(x'), \|(u' - u) - (x' - x)\| \le \varepsilon\}. \tag{2}$$

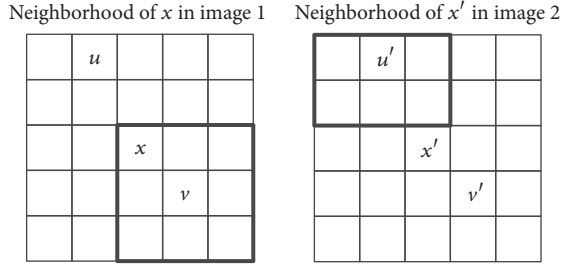Figure 3 illustrates a case for $N = 2$, $\varepsilon = 1$.

Neighborhood of x in image 1   Neighborhood of x' in image 2

Figure 3: Possible matches (u, u') and (v, v') around a seed match (x, x') come from its 5×5 neighbor $N(x)$ and $N(x')$ as the smallest size for discrete 2D disparity gradient limit. The match candidates for u (resp., v') are within the 3×3 (black framed) centered at u' (resp., v).

In Algorithm 1, $\text{ANCC}(u, u')$ is similarity measure to match points of interest in two images. Generally, the color image formation model at pixel p can be represented as follows:

$$\begin{pmatrix} R(p) \\ G(p) \\ B(p) \end{pmatrix} \longrightarrow \begin{pmatrix} \widetilde{R}(p) \\ \widetilde{G}(p) \\ \widetilde{B}(p) \end{pmatrix} = \begin{pmatrix} \rho(p)\, aR(p)^\gamma \\ \rho(p)\, bG(p)^\gamma \\ \rho(p)\, cB(p)^\gamma \end{pmatrix}, \quad (3)$$

where the brightness factor $\rho(p)$ depends on the angle between the direction of light and the direction of surface normal at that point. The global scale factors a, b, c depend on illumination color and γ depends on camera. The well-known similarity measure ZNCC that uses raw stereo images does not generate satisfactory result for two reasons. One is that various radiometric changes caused by $\rho, a, b, c,$ and γ are not taken into consideration. The other is that the object boundaries as the supporting windows from the left and right images do not appear exactly because of the view changes.

To keep the color values away from the impact of radiometric factors, we transform the nonlinear relationship between the corresponding pixels to the affine-transformed relationship that can be solved effectively by the NCC framework. First, we take logarithm to transform this nonlinear relationship into a linear one. Then, the color value can be represented by (use the red channel to describe)

$$R'_L(p) = \log \rho_L(p) + \log a_L + \gamma_L R_L(p),$$
$$R'_R(p') = \log \rho_R(p') + \log a_R + \gamma_R R_R(p'). \quad (4)$$

To delete the item $\log \rho_L(p)$, we subtract the average of the transformed color values in R, G, B channels. Then each color value becomes

$$R''_L(p) = \log \frac{a_L}{\sqrt[3]{a_L b_L c_L}}$$
$$+ \gamma_L \log \frac{R_L(p)}{\sqrt[3]{R_L(p)\, G_L(p)\, B_L(p)}}$$

$$\triangleq \alpha_L + \gamma_L K_L(p),$$
$$R''_R(p') \triangleq \alpha_R + \gamma_R K_R(p'). \quad (5)$$

We can see that $K_L(p)$ and $K_R(p')$ are not dependent on $\rho, a, b, c,$ and γ.

In order to reduce the fattening effect caused by outliers in the matching windows, each pixel t in $m \times m$ window $W(p)$ around the center pixel p is allocated with different weights. The weight is computed by using the bilateral filter as follows:

$$\omega(t) = \exp\left[-\frac{\|p - t\|^2}{2\sigma_d^2} - \frac{\|I(p) - I(t)\|^2}{2\sigma_s^2} \right]. \quad (6)$$

The weighted sum using the bilateral filter for the center pixel p is defined by the following equation, in which $Z(p)$ is the normalizing constant:

$$S(p) = \frac{\sum_{t \in W(p)} \omega(t)\, I(t)}{Z(p)}. \quad (7)$$

Then we subtract $S(p)$ for each channel, thereby resulting in the removal of α.

$$R'''_L(p) = R''_L(p) - S_L(p)$$
$$= \gamma_L \left(K_L(p) - \frac{\sum_{t \in W(p)} \omega_L(t)\, K_L(t)}{Z(p)} \right). \quad (8)$$

The similarity between p and p' is defined by the following equation that is close to NCC's framework:

$$\text{ANCC}_{\log \text{Chrom\_}R}(p, p')$$
$$= \frac{\sum_{i=1}^{M} \omega_L(t_i)\, \omega_R(t_i)\, R'''_L(t_i)\, R'''_R(t_i)}{\sqrt{\sum_{i=1}^{M} |\omega_L(t_i)\, R'''_L(t_i)|^2} \times \sqrt{\sum_{i=1}^{M} |\omega_R(t_i)\, R'''_R(t_i)|^2}}. \quad (9)$$

Because the intensity information of each channel is normalized in log-chromaticity, the discriminability is lower than the original color. We therefore combine log-chromaticity and original color for keeping the discriminability. The final similarity equation is defined as follows:

$$\text{ANCC}(p, p') = \theta \sum_{\xi} \frac{\text{ANCC}_\xi(p, p')}{3}$$
$$+ (1 - \theta) \sum_{k} \frac{\text{ANCC}_k(p, p')}{3}, \quad (10)$$

where $\xi \in \{\log \text{Chrom\_}R, \log \text{Chrom\_}G, \log \text{Chrom\_}B\}, k \in \{R, G, B\}$. θ is a balancing factor between the log-chromaticity color and the original RGB color.
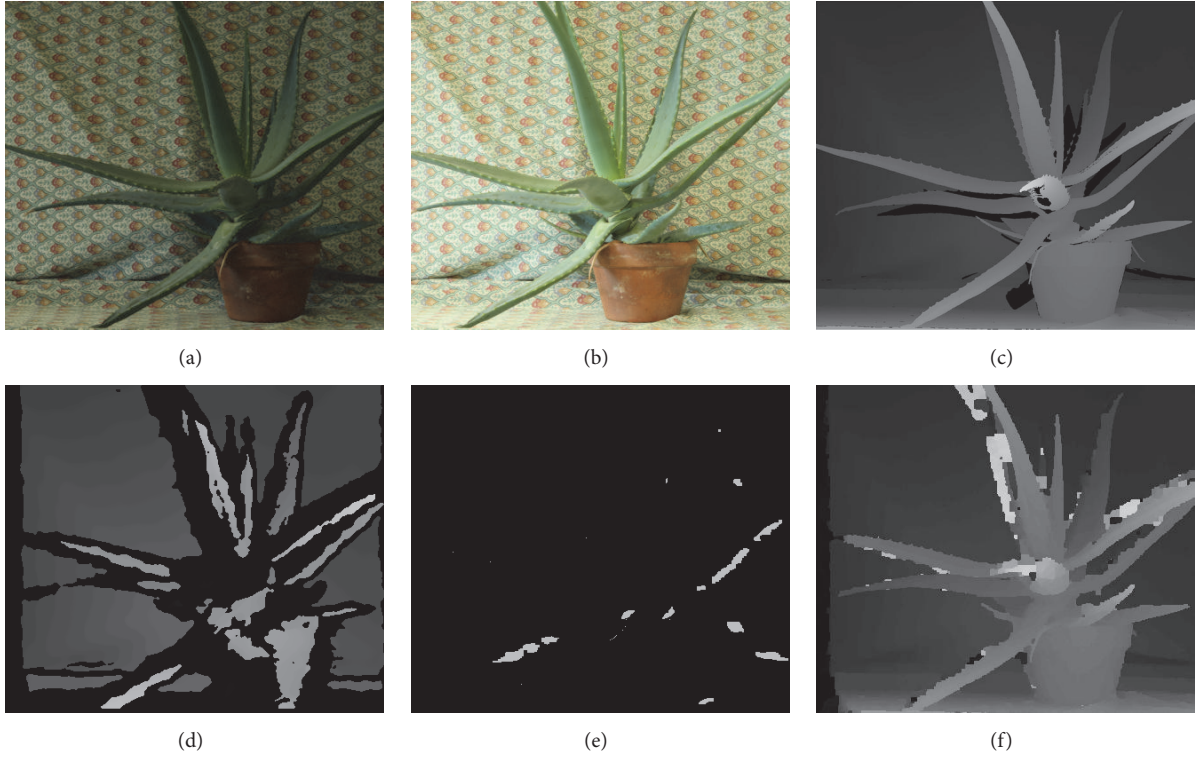
FIGURE 4: Results of test stereo algorithms on Aloe image pair in different lighting conditions. (a) The left image in lighting condition 1. (b) The right image in lighting condition 2. (c) The ground truth disparity map. (d) The disparity map of SGBM by OpenCV. (e) The disparity map of GC by OpenCV. (f) The disparity map of ANCC+GC.

To obtain the disparity map of the two images in different lighting conditions, we define stereo matching as a minimization problem of the following energy in the MAP-MRF framework:

$$E(f) = E_{\text{data}}(f) + E_{\text{smooth}}(f),$$

$$E_{\text{data}}(f) = \sum_P D_p(f_p),$$

$$E_{\text{smooth}}(f) = \sum_P \sum_{q \in N(p)} V_{pq}(f_p, f_q),$$

(11)

where $N(p)$ is the neighborhood of pixel p and f is the disparity value we are trying to find for each pixel p. $D_p(f_p)$ is the data cost function which is defined by $D_p(f_p) = 1 - \text{ANCC}(p, p+f_p)$. $V_{pq}(f_p, f_q)$ is the smoothness cost function based on truncated quadratic model which is defined by $V_{pq}(f_p, f_q) = \lambda * \min(|f_p - f_q|^2, V_{\max})$. The total energy $E(f)$ was optimized using the Graph Cuts (GC) algorithm [37] in order to find the disparity map f. Under different lighting conditions, it can be seen from Figure 4 that the algorithm using ANCC as the matching cost function is very robust to illumination changes compared with results of two classic algorithms SGBM and Graph Cuts by OpenCV library.

The point clouds from multiview stereo reconstruction in different matching method are illustrated in Figures 5 and 6. From the results we can see that many points are not reconstructed correctly (marked in red circles) based on the

ZNCC matching method because of the fattening effect. And more accurate points are acquired with our reconstruction method based on ANCC in Figure 6.

3. Leaves Segmentation from Point Clouds

To segment each individual leaf from point clouds, an improved region growing method based on fully connected CRF is applied in this section. The method can be divided into three steps: boundary erosion, initial segmentation, and refinement. At the beginning, the edge of each leaf point cloud is eroded to remove the connectivity between leaves. Then leaves will be initially segmented by region growing algorithm based on local surface normal and curvature. Finally an efficient CRF inference method based on mean field approximation is used to remove small isolated regions. The details are further explained as follows.

3.1. Boundary Erosion and Initial Segmentation. Because of the similarity of the colors of leaves, the color information is not reliable for segmentation. Considering the fact that the points on the same leaf surface usually have similar normals, and different leaves usually have different mean normals, the region growing algorithm is used. This method can find smoothly connected areas in point cloud based on point normals. First, the normal for each point is estimated by fitting a plane to some neighboring points. The neighborhood can be specified in two different methods, namely, K nearest

(a)   (b)   (c)

FIGURE 5: Multiview stereo reconstruction based on ZNCC: (a) bust; (b) *Scindapsus*; (c) *Pachira macrocarpa*.
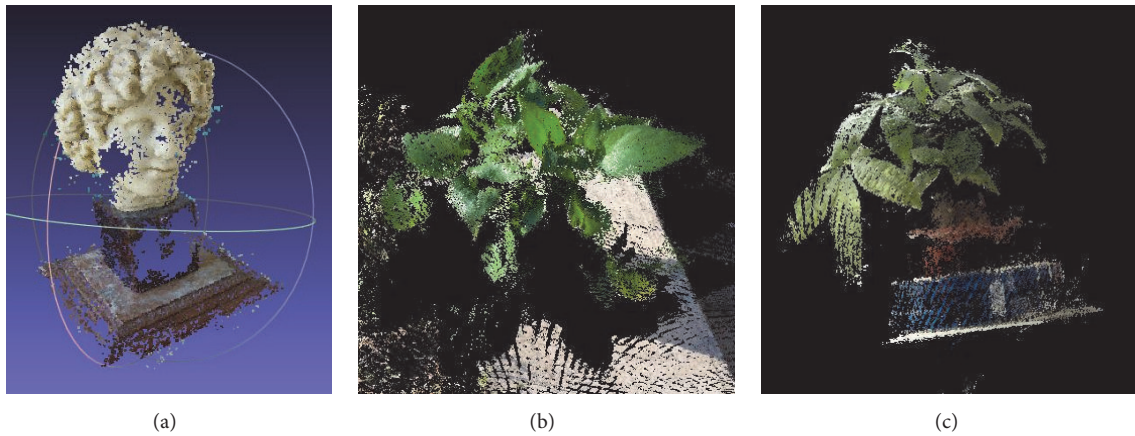

(a)   (b)   (c)

FIGURE 6: Multiview stereo reconstruction based on ANCC: (a) bust; (b) *Scindapsus*. (c) *Pachira macrocarpa*.

neighbors (KNN) and fixed distance neighbors (FDN). The next is the region growing algorithm based on normal and curvature, which is presented in Algorithm 2.

This algorithm is based on two constraints—local connectivity and surface smoothness. For local connectivity, the points in a segment should be locally connected. This can be realized by using only the neighboring points (through KNN or FDN) during region growing. For surface smoothness, the points in a segment should create a local smooth surface whose normals do not vary much. This constraint can be realized with a threshold θ_{th} on the angles between the current seed point and the points added to the region. Additionally, a threshold on curvature values γ_{th} can ensure that smooth areas are broken on the edges. The result of region growing segmentation without boundary erosion for a canopy point cloud is shown in Figure 7.

As it can be seen from Figure 7 some leaves are segmented into one part, which indicates relatively poor performance because some leaves are connected (marked by red circles) and these leaves have similar normals. To solve the problem, some erosion operations around the leaf edge are performed before segmentation. First, the boundary of the point cloud is extracted by detecting the boundary of the corresponding
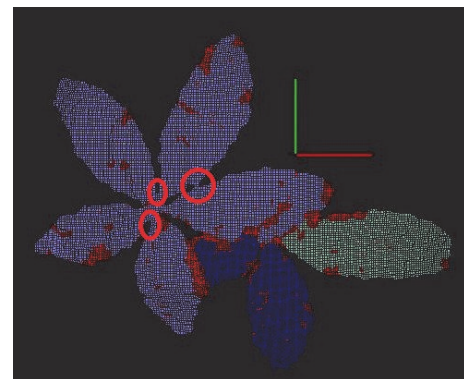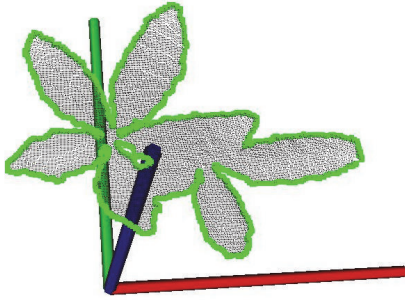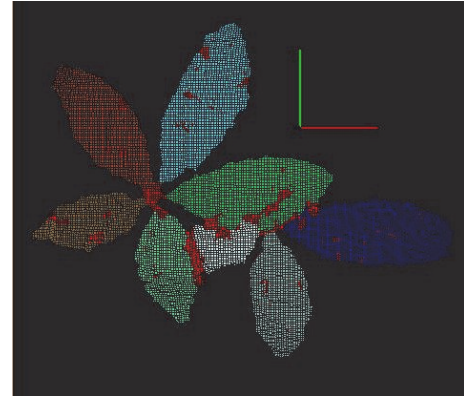

FIGURE 7: Region growing segmentation without boundary erosion for the point cloud from *Pachira macrocarpa*. θ_{th} = 3.25, γ_{th} = 1.0. The red points are small isolated regions because of too little points, which we will remove in the future refinement stage.

depth map. The boundary of the point cloud is illustrated in Figure 8. Then we construct *k-d* tree based on all points including the boundary ones. For every boundary point, we find the *k* nearest points nearby and then delete them. After

Input: Point cloud = {P}, point normals {N}, residuals {r}, neighbor finding function $\Omega(\cdot)$, residual threshold γ_{th}, angle threshold θ_{th}
Output: Leaf point cloud {R}
(1) **Initialize** Region List {R} $\leftarrow \Phi$, Available points list {A} $\leftarrow \{1 \cdots P_{count}\}$
(2) **while** {A} is not empty **do**
(3)   Current region {R$_c$} $\leftarrow \Phi$, Current seeds {S$_c$} $\leftarrow \Phi$
(4)   Point with minimum residual in {A} $\rightarrow P_{min}$
(5) $P_{min} \xrightarrow{\text{insert}} \{S_c\} \& \{R_c\}$
(6) $P_{min} \xrightarrow{\text{remove}} \{A\}$
(7)   **for** i = 0 to size ({S$_c$}) **do**
(8)   Find nearest neighbors of current seed point {B$_c$} $\leftarrow \Omega(S_c\{i\})$
(9)   **for** j = 0 to size ({B$_c$}) **do**
(10)   Current neighbor point $P_j \leftarrow B_c\{j\}$
(11)   **if** {A} contains P_j and $\cos^{-1}(|N\{S_c\{i\}\}, N\{P_j\}|) < \theta_{th}$
(12)   **then** $P_j \xrightarrow{\text{insert}} \{R_c\}$ $P_j \xrightarrow{\text{remove}} \{A\}$
(13)   **if** $r\{P_j\} < \gamma_{th}$
(14)   **then** $P_j \xrightarrow{\text{insert}} \{S_c\}$
(15)   **end-if**
(16)   **end-if**
(17)   **end-for**
(18)   **end-for**
(19)   Add current region to global segment list {R$_c$} $\xrightarrow{\text{insert}}$ {R}
(20) **end- while**
(21) Sort {R} according to the size of the region.
(22) **Return** {R}

ALGORITHM 2: The region growing algorithm for point cloud segmentation.

FIGURE 8: The boundary of the point cloud from *Pachira macrocarpa* leaves.

FIGURE 9: Region growing segmentation after boundary corrosion. $\theta_{th} = 3.25$, $\gamma_{th} = 1.0$.

erosion, we carry out the region growing algorithm for the point cloud and the new result is shown in Figure 9. It can be seen that the result is much better than previous result without boundary erosion and also leaves that were originally connected are now separated.

3.2. Segmentation Refinement. After initial segmentation, there still exist some small isolated regions (small red points region). Therefore we need to classify these regions and determine the actual regions they belong to. A common solution is to transform this problem as maximum a posteriori (MAP) inference in a conditional random field (CRF) defined over

pixels or image patches [38–41]. The CRF model contains pairwise potentials, which can simultaneously remove the noise and resolve ambiguous classification results.

In this subsection, a fully connected CRF is built, establishing pairwise potentials on all pairs of points from the canopy point cloud. Considering a random field X defined over a set of variables $\{X_1, \ldots, X_N\}$, the domain of each variable is a set of labels $L = \{l_1, l_2, \ldots, l_k\}$. Also a random field I is defined over variables $\{I_1, \ldots, I_N\}$. In our setting, I ranges over possible input point cloud of size N and X ranges

(a)   (b)   (c)   (d)
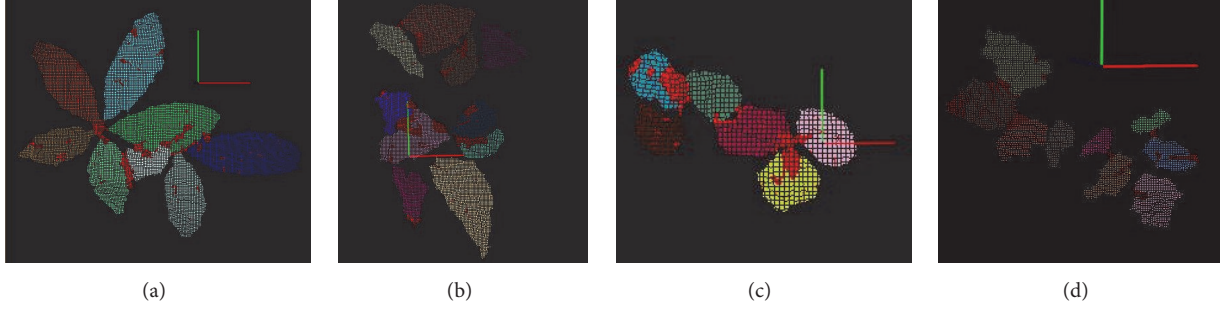
FIGURE 10: Initial segmentation results. (a) *Pachira macrocarpa*. θ_{th} = 3.25, γ_{th} = 1.0. (b) *Scindapsus*. θ_{th} = 4.25, γ_{th} = 1.0. (c) Strawberry. θ_{th} = 3.0, γ_{th} = 1.0. (d) Tomato. θ_{th} = 4.2, γ_{th} = 1.0.

over possible point-level labelings. I_j is the position vector of point j and X_j is the label assigned to point j. Hence, the corresponding Gibbs energy of the fully connected pairwise CRF model is formulated as

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j), \qquad (12)$$

where i and j both range from 1 to N. A conditional random field (I, X) is characterized by a Gibbs distribution $P(x \mid I) = \exp[-E(x)]/Z(I)$. The maximum a posteriori (MAP) labeling of the random field is $x^* = \arg\max_{x \in L^N} P(x \mid I)$. The unary potential is computed independently for each point with a classifier that produces a distribution over the label assignment x_i given point cloud features. The following is the function of the pairwise potentials in our model:

$$\psi_p(x_i, x_j) = \sum_m \mu^{(m)}(x_i, x_j) k^{(m)}(f_i, f_j), \qquad (13)$$

where $k^{(m)}$ is a kernel function and $\mu^{(m)}$ is a label compatibility function. A simple label compatibility function μ is given by the Potts model $\mu(x_i, x_j) = [x_i \neq x_j]$. We limit the choice of the kernel function to Gaussian kernel:

$$
\begin{aligned}
& k^{(m)}(f_i, f_j) \\
& = \omega^{(m)} \exp\left[-\frac{1}{2}(f_i - f_j)^T \Lambda^{(m)}(f_i - f_j)\right],
\end{aligned} \qquad (14)
$$

where $\omega^{(m)}$ is a linear combination weight and $\Lambda^{(m)}$ is a symmetric, positive-definite precision matrix, defining the shape of the kernel. The vectors f_i and f_j are feature vectors for points i and j in an arbitrary feature space, respectively. For our points labeling problem, the smoothness kernel $k(f_i, f_j) = \exp[-|p_i - p_j|^2/2\theta_\gamma^2]$ is used, defined in terms of point positions p_i and p_j to remove isolated regions.

Usually the high complexity of inference in fully connected models will limit its application. Hence, a highly efficient inference algorithm based on a mean field approximation to the CRF distribution [31] is adopted. This approximation yields an iterative message passing algorithm for approximate inference. As message passing in the presented model can be performed using Gaussian filtering in feature space, highly efficient approximations for high-dimensional filtering are utilized to reduce the complexity of message passing from quadratic to linear.

4. Results

The proposed method was applied to four sets of point clouds acquired from *Pachira macrocarpa*, *Scindapsus*, strawberry, and tomato plant.

4.1. Initial Segmentation. Results of the initial segmentation are shown in Figure 10. Some erosion operations were performed around the leaf edges beforehand in case that two leaves are divided into one part.

When a large θ_{th} of angle threshold is set, the algorithm tends to segment the point cloud into connected subsets. The smaller θ_{th}, the higher probability that a leaf is segmented into several clusters (see Figure 11).

4.2. Final Segmentation Results and Comparison with the Original Region Growing Algorithm. The results after inference of fully connected CRF are shown in the second row of Figure 12. It can be seen that all the small isolated regions are classified into its nearest leaves and all leaves are individually segmented.

For the original region growing algorithm, no matter how we adjust the values of θ_{th} and γ_{th}, there always exists the case that two or more leaves are segmented into one subset, which is shown in the first row (highlighted by red circles) of Figure 12. Meanwhile segmentation results contain many noise sections that are not classified. The results of our method are shown in the second row of Figure 12, and the results are superior than the first row.

5. Discussion and Conclusion

We proposed an automatic method to acquire and segment the point cloud of plant leaves. A multiview stereo reconstruction method based on surface expansion is first applied to obtain the 3D structure of plants. We use the Adaptive Normalized Cross-Correlation to match points of interest in two images, and it is robust to radiometric factors and can reduce the fattening effect. In the stage of leaf segmentation on point clouds, an improved region growing method based on fully connected conditional random field (CRF) is proposed to separate connected leaves with similar color. Our method requires no prior botanical knowledge; therefore it can be applied in a wide variety of cases. Qualitative
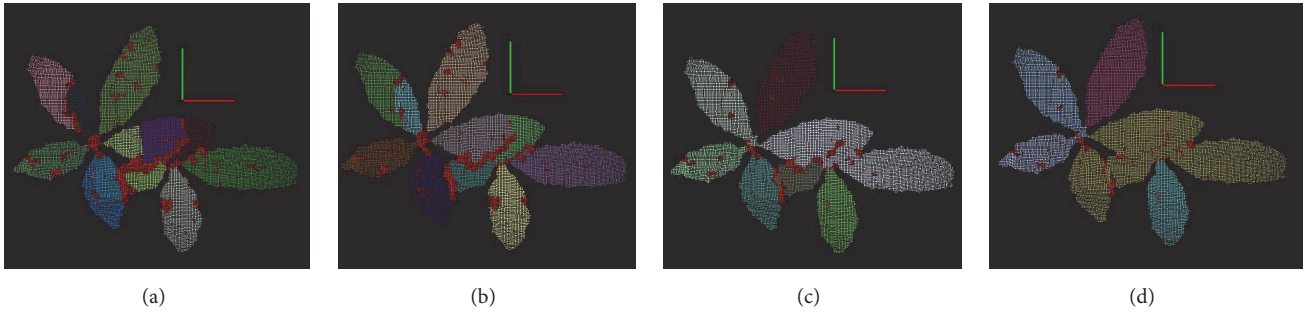
(a)   (b)   (c)   (d)

FIGURE 11: Initial segmentation results under different values of θ_{th}. (a) θ_{th} = 2.75. (b) θ_{th} = 3.00. (c) θ_{th} = 3.50. (d) θ_{th} = 3.75.

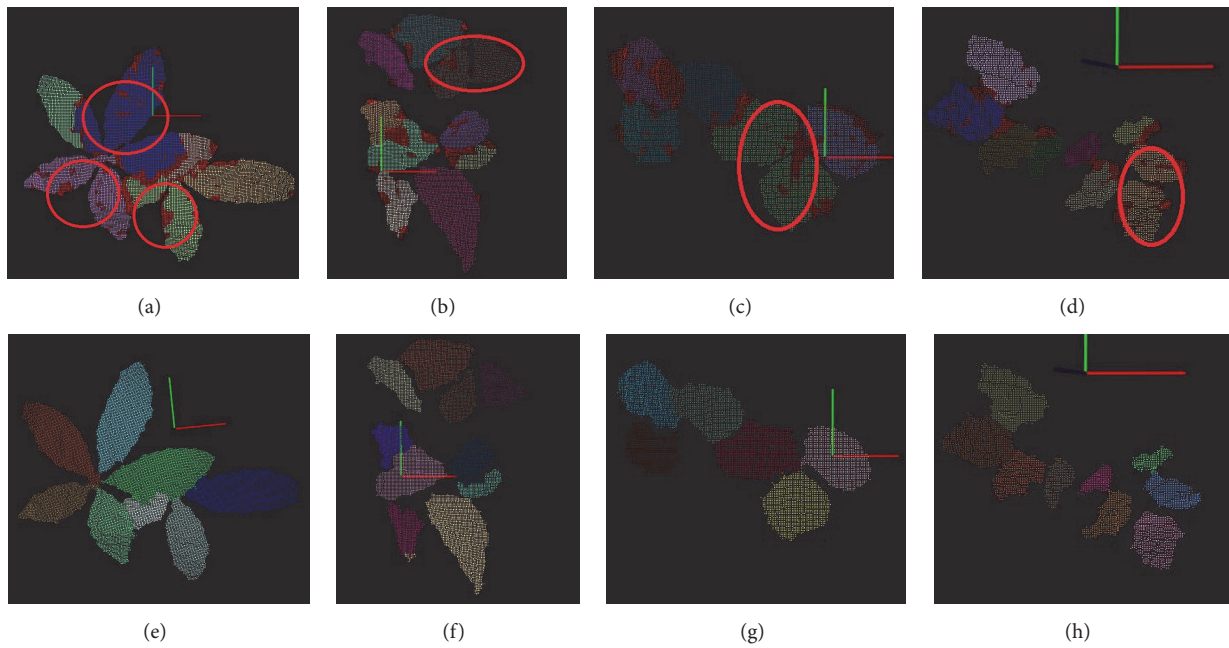(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)

FIGURE 12: Segmentation results. The original region growing algorithm: (a) *Pachira macrocarpa*. θ_{th} = 3.50, γ_{th} = 1.0. (b) *Scindapsus*. θ_{th} = 4.25, γ_{th} = 1.0. (c) Strawberry. θ_{th} = 3.7, γ_{th} = 1.0. (d) Tomato. θ_{th} = 4.2, γ_{th} = 1.0. Our algorithm: (e) *Pachira macrocarpa*. θ_{th} = 3.25, γ_{th} = 1.0. (f) *Scindapsus*. θ_{th} = 4.25, γ_{th} = 1.0. (g) Strawberry. θ_{th} = 3.0, γ_{th} = 1.0. (h) Tomato. θ_{th} = 4.2, γ_{th} = 1.0.

results on several kinds of plants show the effectiveness of our method. To prevent two leaves from being segmented into only one subset, some erosion operations around the leaf edge are performed before carrying out the region growing segmentation based on normal and curvature. A fully connected CRF model is built in which the pairwise edge potentials are defined by Gaussian kernel to remove small isolated regions after segmentation.

In the future, we will focus on developing advanced multiview stereo reconstruction algorithms to obtain more accurate point clouds for plant phenotype analysis. In addition, we will also study advanced point cloud segmentation algorithms to obtain other organs such as stems and fruits from plants, for realizing the final goal of accurate measurement on plant organs.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] V. Sarlikioti, P. H. B. De Visser, G. H. Buck-Sorlin, and L. F. M. Marcelis, "How plant architecture affects light absorption and photosynthesis in tomato: towards an ideotype for plant architecture using a functional-structural plant model," *Annals of Botany*, vol. 108, no. 6, pp. 1065–1073, 2011.

[2] J. Vos, J. B. Evers, G. H. Buck-Sorlin, B. Andrieu, M. Chelle, and P. H. B. De Visser, "Functional-structural plant modelling:

a new versatile tool in crop science," *Journal of Experimental Botany*, vol. 61, no. 8, pp. 2101–2115, 2010.

[3] W. Yang, Z. Guo, C. Huang et al., "Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice," *Nature Communications*, vol. 5, article 5087, 2014.

[4] N. Fahlgren, M. A. Gehan, and I. Baxter, "Lights, camera, action: high-throughput plant phenotyping is ready for a close-up," *Current Opinion in Plant Biology*, vol. 24, pp. 93–99, 2015.

[5] Y. Guanghui and F. Xuanjun, "Concept of phenomics and its development in plant science," *Molecular Plant Breeding*, no. 4, pp. 639–645, 2009.

[6] Y. Pan, "Analysis of concepts and categories of plant phenome and phenomics," *Acta Agronomica Sinica*, vol. 41, no. 2, pp. 175–186, 2015.

[7] N. Ivanov, P. Boissard, M. Chapron, and B. Andrieu, "Computer stereo plotting for 3-D reconstruction of a maize canopy," *Agricultural and Forest Meteorology*, vol. 75, no. 1–3, pp. 85–102, 1995.

[8] B. Biskup, H. Scharr, U. Schurr, and U. Rascher, "A stereo imaging system for measuring structural parameters of plant canopies," *Plant, Cell and Environment*, vol. 30, no. 10, pp. 1299–1308, 2007.

[9] P. Hu, Y. Guo, B. Li, J. Zhu, and Y. Ma, "Three-dimensional reconstruction and its precision evaluation of plant architecture based on multiple view stereo method," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 31, no. 11, pp. 209–214, 2015.

[10] L. Quan, P. Tan, G. Zeng et al., "Image-based plant modeling," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 599–604, 2006.

[11] S. N. Sinha, P. Mordohai, and M. Pollefeys, "Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[12] G. Vogiatzis, C. Hernández, P. H. S. Torr, and R. Cipolla, "Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2241–2246, 2007.

[13] K. Kolev, M. Klodt, T. Brox, and D. Cremers, "Continuous global optimization in multiview 3D reconstruction," *International Journal of Computer Vision*, vol. 84, no. 1, pp. 80–96, 2009.

[14] K. N. Kutulakos and S. M. Seitz, "Theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.

[15] M. Goesele, B. Curless, and S. M. Seitz, "Multi-view stereo revisited," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 2402–2409, June 2006.

[16] C. Zach, T. Pock, and H. Bischof, "A globally optimal algorithm for robust TV-L^1 range image integration," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[17] J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang, "Bundled depth-map merging for multi-view stereo," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2769–2776, IEEE, June 2010.

[18] L. Lucas, C. Loscos, and Y. Remion, "Multi- and stereoscopic matching, depth and disparity," in *3D Video*, chapter 7, pp. 137–155, John Wiley & Sons, Hoboken, NJ, USA, 2013.

[19] M. Habbecke and L. Kobbelt, "A surface-growing approach to multi-view stereo reconstruction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, Minneapolis, Minn, USA, June 2007.

[20] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 418–433, 2005.

[21] J. Čech and R. Šára, "Efficient sampling of disparity space for fast and accurate matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, IEEE, Minneapolis, Minn, USA, June 2007.

[22] P. Koskenkorva, J. Kannala, and S. S. Brandt, "Quasi-dense wide baseline matching for three views," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 806–809, Istanbul, Turkey, August 2010.

[23] Y. S. Heo, K. M. Lee, and S. U. Lee, "Robust Stereo matching using adaptive normalized cross-correlation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 807–822, 2011.

[24] A. Nguyen and B. Le, "3D point cloud segmentation: a survey," in *Proceedings of the 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM '13)*, pp. 225–230, IEEE, Manila, Philippines, November 2013.

[25] S. Paulus, J. Dupuis, A.-K. Mahlein, and H. Kuhlmann, "Surface feature based classification of plant organs from 3D laser-scanned point clouds for plant phenotyping," *BMC Bioinformatics*, vol. 14, no. 1, article no. 238, 2013.

[26] S. Paulus, J. Dupuis, S. Riedel, and H. Kuhlmann, "Automated analysis of barley organs using 3D laser scanning: an approach for high throughput phenotyping," *Sensors*, vol. 14, no. 7, pp. 12670–12686, 2014.

[27] M. Wahabzada, S. Paulus, K. Kersting, and A.-K. Mahlein, "Automated interpretation of 3D laserscanned point clouds for plant organ segmentation," *BMC Bioinformatics*, vol. 16, no. 1, article 248, 2015.

[28] K. Yin, H. Huang, P. Long, A. Gaissinski, M. Gong, and A. Sharf, "Full 3D plant reconstruction via intrusive acquisition," *Computer Graphics Forum*, vol. 35, no. 1, pp. 272–284, 2016.

[29] F. Hétroy-Wheeler, E. Casella, and D. Boltcheva, "Segmentation of tree seedling point clouds into elementary units," *International Journal of Remote Sensing*, vol. 37, no. 13, pp. 2881–2907, 2016.

[30] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 5, pp. 248–253, 2006.

[31] P. Krähenbühl and V. Koltun, "Parameter learning and convergent inference for dense random fields," in *Proceedings of the 30th International Conference on Machine Learning (ICML '13)*, pp. 1550–1558, IMLS, Atlanta, Ga, USA, June 2013.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[33] K. Mikolajczyk, T. Tuytelaars, C. Schmid et al., "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[34] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM*, vol. 45, no. 6, pp. 891–923, 1998.

[35] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003.

[36] M. Lhuillier and L. Quan, "Match propagation for image-based modeling and rendering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1140–1146, 2002.

[37] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[38] X. He, R. S. Zemel, and M. Á. Carreira-Perpiñán, "Multiscale conditional random fields for image labeling," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. II695–II702, July 2004.

[39] P. Kohli, L. Ladický, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *International Journal of Computer Vision*, vol. 82, no. 3, pp. 302–324, 2009.

[40] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, October 2007.

[41] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2–23, 2009.