SEGMENTATION OF APPLE POINT CLOUDS BASED ON ROI IN RGB IMAGES
/
基于彩色图像中感兴趣区域的苹果点云分割方法

Yuanxi Zhang[1)], Ye Tian[1)], Change Zheng*[2)], Dong Zhao[1)], Po Gao[1)], Ke Duan[1)]
[1)]School of Technology, Beijing Forestry University, Beijing, 100083/China,
[2)]Key Lab of State Forestry Administration on Forestry Equipment Automation, School of Technology, Beijing Forestry University, Beijing, 100083/China;
Tel: +8618601224519; E-mail: zhengchange@bjfu.edu.cn
DOI: https://doi.org/10.35633/inmateh-59-23

Keywords: *faster-RCNN, segmentation, apple tree, point clouds, unstructured scenes, ROI*

ABSTRACT

Autonomous harvesting and evaluation of apples reduce the labour cost. Segmentation of apple point clouds from consumer-grade RGB-D camera is the most important and challenging step in the harvesting process due to the complex structure of apple trees. This paper put forward a segmentation method of apple point clouds based on regions of interest (ROI) in RGB images. Firstly, an annotated RGB dataset of apple trees was built and applied to train the optimized Faster R-CNN to locate ROI containing apples in RGB images. Secondly, the relationship between RGB images and depth images was built to roughly segment the apple point clouds by ROI. Finally, the quality control procedure (QCP) was proposed to improve the quality of segmented apple point clouds. Images for training mainly included two lighting condition, two colours and three apple varieties in orchard, making this method more suitable for practical applications. QCP performed well in filtering noise points and achieved Purity as 96.7% and 96.2% for red and green apples, respectively. Through the comparison method, experimental results indicated that the segmentation method based on ROI is more effective and accurate for red and green apples in orchard. The segmentation method of point clouds based on ROI has great potential for segmentation of point clouds in unstructured scenes.

摘要

苹果的自动采摘和评估降低了劳动成本。在采摘过程中由于苹果树的复杂结构，分割消费级 RGB-D 相机获得的苹果点云是最重要且具有挑战性的一步。本文提出基于彩色图像中感兴趣区域的点云分割方法。首先制作了一个用来训练优化过的 Faster-RCNN 的苹果树数据集，定位出彩色图像中包含有苹果的感兴趣区域。然后，构建了彩色图与深度图之间的关系量，依照此关系量使用感兴趣区域对点云进行快速的粗分割。最后，针对粗分割的点云特征提出质量控制程序，提升苹果点云的分割质量。用来训练的彩色图像主要包含了果园中的两种光照条件以及三个品种、两种颜色的苹果，提升了此方法在实际场景中的适用性。质量控制程序的处理效果理想，对于红绿两色苹果分别得到了 96.7%和 96.2%的纯净度。通过对比实验，结果证明了基于彩色图像中感兴趣区域的苹果点云分割方法对于果园中红绿两色的苹果均有更高的效率和准确率，在非结构化场景中具有非常大的应用前景

INTRODUCTION

Harvesting and evaluation of apples is a labour-intensive task in a short period. In the task of reducing the labour cost, automatic system is of great value to agriculture (*Kitamura S. and Oka K., 2006; English A. et al., 2014; Bac CW et al., 2015*). The key of automatic system is the apple detection based on segmentation of point clouds. Since the point clouds is one of the most intuitive formats for a machine to perceive the world (*Charles RQ et al., 2017; Li Y et al., 2018; Pas AT and Platt R, 2015*), much more information such as the optimized end-effector pose and the position of apples can be obtained. Therefore, it is highly desirable to analyse the features (e.g., surface fitting and calculate the normal vector distribution) from segmented point clouds of apples. In structured scenes like buildings or other objects that contain neat planes and lines (*Yamauchi H et al., 2005; Himmelsbach M et al., 2010 Xu Y et al., 2018*;) achieved a good result of segmentation using geometry features, William R. introduced normal distribution feature representation to segment indoor point clouds. Moreover (*Green WR and Grobler H, 2015*), Huijun Y proposed an algorithm for segmenting fruit from background by converting octree cell into a splat and

approximating its local surface with MLS (*Huijun Y, 2011*). However, when these methods were used for real apple trees, the accuracy was not satisfactory.

The segmentation of apple point clouds is of considerable difficulty due to the limited size of apples, meanwhile, the occlusions from leaves or branches brought more difficulty for segmentation in unstructured scenes. Additionally, apples have different colours (such as red and green), making it difficult to segment the point clouds only based on colour information. Nguyen TT segmented the 3D information belonging to the red or bicolored apples from backgrounds (*Nguyen TT et al., 2014*). Combining Gaussian processes and a Gibbs-sampler with Euclidean distance between colours, Soria PR et al. segmented clusters of apples and estimated the centroids of red apples (*Soria PR et al., 2017*). Tao Y et al. adopted HSI to segment apple point clouds (*Tao Y et al., 2017*). Meanwhile, Bayesian discriminant method based on RGB model was performed on classing fruits (*Blasco J et al., 2003*). Linker R and Kelman E proposed detection of apples in night-time tree images by converting RGB to gray-level (*Linker R and Kelman E, 2015*). However, when it comes to the different colour condition like green apples in green leaves, the performance of above-mentioned methods is not satisfactory. Paulus et al. segmented point clouds of plant organs using an accurate laser scanner. But this paper got a poor result on apple point clouds using a consumer-grade RGB-D camera (*Paulus et al., 2013; Wahabzada M et al., 2015*). A support vector machine optimized by genetic algorithm (GA-SVM) classifier was constructed by Tao et al. to recognize apples. However, the GA-SVM didn't work well in different orchard scenes. Although there was some impressive progress in 3D deep convolutional neural networks trained by 3D datasets demanding great expense such as S3DIS and Scannet (*Wang W et al., 2017; Dai A et al., 2017*), it is too difficult to build a 3D point clouds dataset of real apple trees for this paper due to the complexity of the unstructured apple tree point clouds. 2D dataset, by contrast, is relatively easier to build, thus 2D deep learning has been widely adopted for many tasks like crops detection. Hasan MM et al. performed accurate detection of wheat spikes using Faster R-CNN and counted wheat spikes for yield estimation (*Hasan MM et al., 2018*).

This paper focuses on segmentation of apple point clouds in order to make full use of 3D data. The main contribution of this paper is using the optimized Faster R-CNN to segment apple point clouds, making segmentation effective and accurate. In this process, we built an apple dataset including 5660 images of different-colour apples in green leaves under different lighting conditions, 4810 images were labelled manually in standard format. Therefore, the dataset can be used for other works. Moreover, we proposed a quality control procedure (QCP) for filtering noise points, improving the quality of point clouds obviously.

MATERIALS AND METHODS

In this study, a new method for segmentation was proposed for apple tree point clouds. The Kinect V2 was used to obtain RGB-D information of apple trees as shown in Fig. 2. At the same time, a manually annotated RGB dataset, which contained the RGB images, was built and applied to train the optimized Faster R-CNN to locate the ROI containing apples in RGB images, and then, the relationship between RGB images and depth images was established by defining a variable called IDX, which was used to segment the point clouds. After roughly segmenting point clouds based on the ROI and IDX, QCP was put forward to ensure an accurate result. The main works and methods are shown in Fig. 1. In order to evaluate the performance of segmentation, the improved method based on Pixel Colour (MBPC) was introduced as comparison experiments.

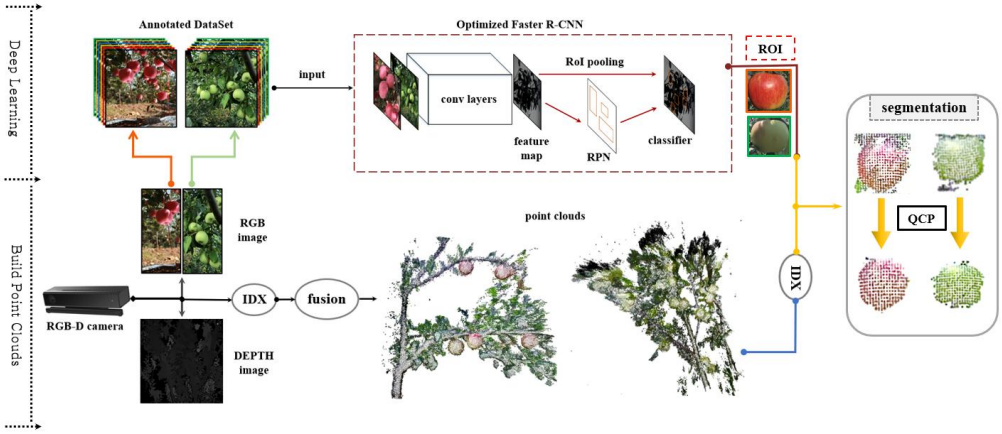


Fig. 1 - Flowchart of segmentation method using optimized Faster R-CNN and RGB-D camera. RGB images were used to train the optimized Faster R-CNN and fuse into point clouds. Combining ROI with IDX, the segmentation of raw apple point clouds was achieved, and the quality of segmentation was improved by QCP

Fig. 2 - Images of different apple trees in RGB dataset obtained on different lighting conditions (sunshine and shadow)

- **Dataset**

Apples have different colours, among them, the red and green apples are the most common colours, the apple varieties in RGB dataset included Fuji, Qinyang and Liaofu, which were collected in Heze and Jinan, Shandong province, China. These raw images were obtained by the Kinect V2 RGB sensor in two kinds of lighting conditions as shown in Fig. 2. The basic task in this paper is to detect the apples in RGB images; therefore, the apples in RGB images were manually labelled. In addition, the distance between the camera and the target was about 1-1.5m similar with practical harvesting applications. The Kinect V2 depth sensor uses a structured light technique to acquire the depth information of the apple trees, which is able to extract a set of discrete point clouds of apple trees. The resolution of each sensor is shown in Table 1.

Table 1

Main parameters of sensors used in data collection

Sensor	Resolution	Range of Detection
RGB	1920*1080	✗
Depth	512*424	0.5m~4.5m

The image annotations were manually labelled and saved in XML files in PASCAL VOC format. The annotations included two categories: red apple (RA) and green apple (GA), and the bounding boxes (bbox).

In order to reduce the possibility of over-fitting in training process (*Grinblat GL et al., 2016*), horizontal flip processing was used as data enhancement. Specifically, by performing horizontal symmetry processing on all RGB images, the amount of data was doubled. The enhanced RGB dataset contains a total of 5660 pictures in two categories: Red Apples in Green Leaves (RAGL) and Green Apples in Green Leaves (GAGL). Before training the optimized Faster R-CNN, 15% of the images from each category were randomly selected as validation set. The details of the RGB dataset are shown in Table 2. After the validation set was reserved, the apple tree point clouds were generated by using the RGB images aligned with the depth information, and then the Ground Truth (GT) dataset of apple point clouds was manually segmented according to the RGB images, as shown in Fig. 3. The remained RGB images were used as a training set of the optimized Faster R-CNN, the details of GT dataset are shown in Table 3.

Table 2

Number of images from RGB dataset for training and testing (Augmentation-performed)

Image	RAGL		GAGL	
Training	2000	bbox	2810	bbox
		12640		18304
Validation	300	✗	550	✗
Total	2300		3360	

Fig. 3 - Examples of GT dataset corresponding to the validation set. (Two colours)

Table 3

Details of GT dataset

GT	RA	GA
GT Num (piece)	1717	2350
Point Num (million)	1.88	2.29

- **The optimized Faster R-CNN**

Regions with convolutional neural networks (R-CNN) are widely used in object detection, using selective search methods to detect ROI, and then using CNN to classify these regions (*Girshick R et al., 2013*). After that, Fast R-CNN uses the ROI pooling to extract fixed-length feature vectors from the feature map, and backpropagates all network weights in training (*Girshick R, 2015*). When Faster R-CNN is proposed, the efficiency is further improved to 10ms for generating the candidate regions of an image by using the Region Proposal Networks (RPN) (*Ren S et al., 2015*).

In this paper, RGB images were used to detect apple regions, and the implementation of the RPN is the key work of this model. In addition, the sliding scan is performed by using the 3×3 sliding window on the 120×67 feature map of the fifth layer extracted by ZF network structure (*Zeiler MD and Fergus R, 2013*); then the optimized Faster R-CNN performs a convolution with a 1×1 kernel. As a result, a 256-dimensional feature vector was obtained. In addition, RPN relies on the anchor mechanism to generate candidate regions, then the R-CNN is used to perform apple detection and fine adjustment of the ROI. At the end, the apples in the image can be located by ROI.

For each ROI in RGB images detected by optimized Faster R-CNN, we provide a confidence C ∈ [0,1], 0 is for the lowest confidence and 1 is for the highest confidence. In the detection, the confidence thresholds were set to 0.5 for RAGL and 0.55 for GAGL, which were proved by experiments in both categories. More importantly, since training Faster R-CNN with RGB images is independent of point clouds, once the optimized Faster R-CNN is successfully trained for other objects, this model can be extended to more practice than apple.

- **Segmentation method based on ROI**

In this paper, the point clouds without colour were obtained from depth images as first step as follows: $\mathbf{V(u)} = D(\mathbf{u})\mathbf{K}^{-1}[\mathbf{u},1]$, where $\mathbf{u} = (u,v)$ is a separate pixel in a depth image, $D(\mathbf{u})$ is the depth value of $\mathbf{u}$ and $\mathbf{K}$ is the intrinsic calibration matrix of the Kinect depth camera. As a result, a specific depth measurement as a 3D vertex $\mathbf{V(u)}$ in the camera's coordinate space was obtained. Moreover, to get coloured point clouds, for each pixel $\mathbf{m} = (m_c, v_c)$ in RGB image, we used $[\mathbf{m},1] = \mathbf{CTV(u)}^*$ to search the corresponding 3D vertex to fuse coloured point clouds based on the pixel colour in RGB image, where $\mathbf{C}$ is the intrinsic calibration matrix of RGB camera, $\mathbf{T}$ is the external calibration matrix between two cameras and $\mathbf{V(u)}^* = \mathbf{V(u)} / D(\mathbf{u})$.

The second step in segmentation method based on ROI is to search the points in all the apple tree point clouds corresponding to coloured pixel in ROI. To avoid complex computation, the IDX containing constant relationship between RGB images and point clouds was defined as Equation (1), where $w(C)$ and $h(C)$ are the width and height of the RGB pixels respectively. Therefore, the IDX is the real number mapping each RGB pixel to its 3D vertex as well as a link between 2D RGB information and 3D point clouds information.

$$IDX = m_c w(C) + n_c, 1 < m_c < w(C), 1 < n_c < h(C) \quad (1)$$

As a result, the point clouds corresponding to colour pixels in ROI can be rapidly segmented by a series of IDX. In other words, based on IDX and the ROI output by the optimized Faster R-CNN in the RGB images, the point clouds of ROI were roughly segmented.

- **Quality control procedure (QCP).**

The ROI output from RGB images are a series of boxes. However, since the outline of an apple is a rough circle, a small amount of background will also be contained in the box, as shown in Fig. 4. Therefore, a

few noise points were included in the apple point clouds, which will bring great negative effects to practical applications demanding precise shape of apple, such as evaluating size, quality and yield. In addition, when using the robotic arm to pick apples, the normal vectors of the apple surface is critical to the pose of the arm, while the noise points are of great risk to lead to failure in apple grasping task, thus the noise points have to be filtered out to refine the point clouds. QCP was based on estimating surface normal vector and curvature using region growing method. Before this, we computed, using the cross product between two neighbouring vertices, whole corresponding normal vectors using Equation (2),

$$\mathbf{N(u)} = \left[\mathbf{n(V}(u+1,v)-\mathbf{V}(u,v))\times(\mathbf{V}(u,v+1)-\mathbf{V}(u,v))\right] \tag{2}$$

where $\mathbf{n(a)} = \mathbf{a} / \|\mathbf{a}\|_2$

After computing all surface normal vectors, through analyzing the apple point clouds, the $D(\mathbf{u})$ of noise points will have an obvious "drop" separating the noise point from the apple point clouds like the red lines shown in Fig. 4. In addition, the size of an apple is about 0.07m. So, given two vertices $\mathbf{V(u_1)}$ and $\mathbf{V(u_2)}$, we introduced the Euclidean Distance Score ($EDS \in [0,1]$) as shown in Equation (3) and validity distance d_N = 0.01m,

$$EDS(\mathbf{V(u_1)}, \mathbf{V(u_2)}) = (d_N - \|\mathbf{V(u_2)} - \mathbf{V(u_1)}\|_2) / d_N \tag{3}$$

In addition, the curvature of vertex $\mathbf{v(u_1)}$ is approximately equal to

$$c(\mathbf{V(u_1)}, \mathbf{V(u_2)}) = \left|\cos^{-1}(\mathbf{N(u_1)} \bullet \mathbf{N(u_2)} / \|\mathbf{N(u_2)}\|_2 \|\mathbf{N(u_1)}\|_2) / \|\mathbf{V(u_2)} - \mathbf{V(u_1)}\|_2\right|$$

The Curvature Score ($CS(\mathbf{V(u_1)}, \mathbf{V(u_2)}) \in [0,1]$) was defined as Equation (4) and the validity radius is r_N = 0.06m

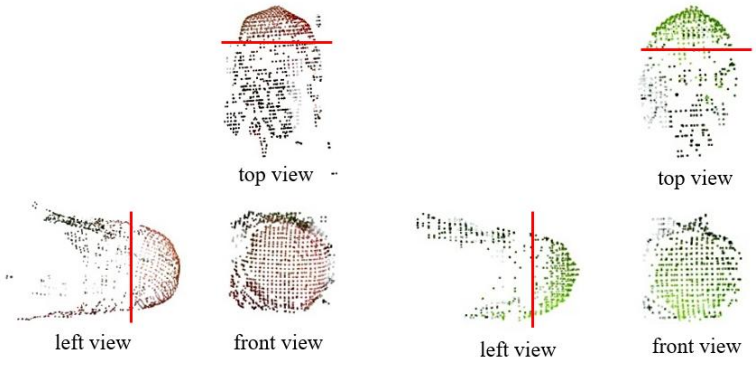
$$CS(\mathbf{V(u_1)}, \mathbf{V(u_2)}) = 1 - \left|1 - 1 / r_N c(\mathbf{V(u_1)}, \mathbf{V(u_2)})\right| \tag{4}$$

More importantly, through data statistics, we finally defined a QCP mask $M(\mathbf{V(u_1)}, \mathbf{V(u_2)}) \in [0,1]$ for a vertex $\mathbf{V(u_1)}$, combining surface normal vector with curvature to filter noise points with a threshold of 0.77 as shown in Equation (5): $M(\mathbf{V(u_1)}, \mathbf{V(u_2)}) \geq 0.77$ is for a point is valid, otherwise, the point is to be filtered when $M(\mathbf{V(u_1)}, \mathbf{V(u_2)}) < 0.77$.

$$M(\mathbf{V(u_1)}, \mathbf{V(u_2)}) = 0.65EDS + 0.35CS \tag{5}$$

By using QCP on roughly segmented apple point clouds, most of noise points can be filtered out.

The arithmetic logic of QCP is briefly showed in Algorithm 1.



(a) (b)

Fig. 4 - Two instances of the 'drop' in raw point clouds segmented by method based on ROI

- **Comparison methods**

Although the colour information has been extracted into the point clouds, it is saved in the RGB model. The Excessive Green Index (ExG) was successfully used to segment red Fuji apple (*Bulanon DM et al., 2002*). The ExG-ExR with zero threshold was proved to be 55% higher than the performance of others in segmenting green crops (*Meyer GE and Neto JC, 2008*). In this paper, ExG was used to segment point clouds of RAGL and ExG-ExR was used to segment point clouds of GAGL.

The detail of indices is shown in the Table 4.

Table 4

Segmentation algorithms for point colour detection

Colour indices	Formula
ExG	2G-R-B
ExG-ExR	3G-2.4R-B

Algorithm 1: QCP

- *While* $\{PtClouds\}$ is not empty *do*
  - $Current\{valid\} \leftarrow \varnothing$
  - $Current\{seeds\} \leftarrow \varnothing$
  - Point at the centre of ROI in $\{PtClouds\} \rightarrow d_c$
  - $\{valid\} \leftarrow d_c$
  - $\{seed\} \leftarrow d_c$
  - *for* i = 0 to size $(\{seeds\})$ *do*
    - *Freshfind* 4 nearest neighbours of current seed point $\{Nr\} \leftarrow \{N_4 seed(i)\}$
    - *for* j = 0 to size $\{Nr\}$ *do*
      - Current neighbour point $P_j \leftarrow N_r\{j\}$
      - Compute EDS$\{j\}$ & CS$\{j\}$ & M$\{j\}$
        - *If* M$\{j\} \geq 0.77$
        - $\{valid\} \leftarrow \{valid\} \bigcup P_j$
        - $\{Nextseeds\} \leftarrow P_j$
          - *If* M$\{j\} < 0.77$
            - $\{PtClouds\} \leftarrow \{PtClouds\} \setminus P_j$
          - end *If*
        - end *If*
    - end *for*
  - end *for*
  - $\{seeds\} \cap \varnothing \leftarrow \{Nextseeds\}$
- end *While*
- *return* R

- **Evaluation methods**

We evaluate the above-mentioned methods by comparing the Apple Point Clouds (APtC) from segmentation with each manually-segmented sample defined as GT^* in GT dataset. An intersection between GT^* and APtC segmented by the method based on ROI, is named $\mathbb{S}$, as shown in Equation (6),

$$\mathbb{S} = APtC \cap GT^* \tag{6}$$

conversely, there is a point set that does not belong to GT^* (NBGT). Due to the difference in size of apple, the number of points is also different. For this reason, the Overlap is defined as shown in Equation (7),

$$Overlap = \frac{\mathbb{S}}{GT^*} \tag{7}$$

which is the important criterion for measuring the segmentation effect. Given the influence of NBGT on the segmentation is also significant, Inaccuracy is defined as shown in Equation (8),

$$Inaccuracy = \frac{NBGT}{GT^*} \tag{8}$$

more importantly, the Purity is introduced here as shown in Equation (9), and the quality of Segmentation is evaluated by both $\mathbb{S}$ and NBGT.

$$Purity = \frac{\mathbb{S}}{APtC} \tag{9}$$

In addition, although training optimized Faster R-CNN is the first step of segmenting point clouds, the effect of training directly affects the result of segmentation of point clouds. If the accuracy of training effect can be improved, the result of segmentation will also be enhanced. Therefore, an evaluation of the training effect was introduced here.

Since the optimized Faster R-CNN used in this task will output a series of bounding boxes, and we hope that these boxes contain all the apples in an image that meet the requirements. Denoting boxes as apple or non-apple can yield three potential results, with the latter two being sources of error: true positive (TP)-Correctly classifying a region as an apple; false positive (FP)-Incorrectly classifying a background region as an apple or multiple detection of the same apple; false negative (FN)-Incorrectly classifying an apple as a background region. In contrast, true negative (TN)-Correct classification of background is not required in this binary classification problem where foreground is always determined for object detection. In order to quantify our training effect, three parameters are introduced as shown in Equation (10) *(Hasan MM et al., 2018*).

$$\begin{cases} precision = \dfrac{TP}{TP + FP} \\ recall = \dfrac{TP}{TP + FN} \\ F1\,score = 2\dfrac{precision * recall}{precision + recall} \end{cases} \tag{10}$$

RESULTS

- **Results of Segmentation method based on ROI**

In the testing process, since there are a large number of annotated apples, the validation set was divided into several groups in order to ensure the objectivity and robustness of the results, each group included 10 images for testing. When testing the performance of the model, 10 groups were randomly picked out. The recognition result of each group and the overall evaluation are shown in Table 5 and Table 6.

Table 5

Evaluation of RAGL/GAGL detection results of the optimized Faster R-CNN model

Group		Selected		Detection		TP		FP		FN	
RA	GA	RA	GA	RA	GA	RA	GA	RA	GA	RA	GA
r03	g01	65	70	62	68	60	67	2	1	5	3
r05	g05	61	73	60	69	59	68	1	1	2	5
r10	g09	52	65	50	64	49	62	1	2	3	3
r11	g15	69	67	65	64	65	61	0	3	4	6
r17	g16	40	74	38	71	38	71	0	0	2	3
r20	g23	70	64	66	63	66	62	0	1	4	2
r25	g29	55	73	54	70	53	70	1	0	2	3
r26	g30	54	50	53	48	52	48	1	0	2	2
r27	g40	67	69	65	68	65	66	0	2	2	3
r30	g51	30	66	30	63	30	63	0	0	0	3

Table 6

Confusion matrix of the optimized Faster R-CNN model

Category	Predicted group				Precision (%)	Recall (%)	F1 score
	RAGL	Background	GAGL	Background			
RAGL	537	26	-	-	98.9	93.4	96.1

Background	6	✗	-	-			
GAGL	-	-	638	33	98.5	95.1	96.8
Background	-	-	10	✗			

As the confusion matrix can be seen in Table 6, the optimized Faster R-CNN model trained with enhanced RGB dataset got precision over 98% and recall over 93% for recognizing both RAGL and GAGL data. The results of F1 score indicated that the optimized Faster R-CNN had great robustness in this task.

Based on this step, the point clouds data was segmented by combining the RGB-D mapping relationship with ROI output from the Faster R-CNN. The overall evaluation is shown in Table 7.

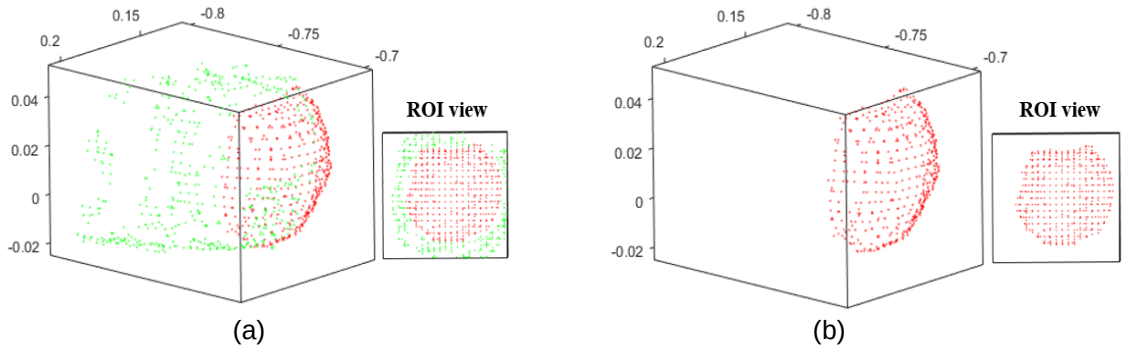
The Overlap can directly reflect the effect of segmentation of point clouds. Without QCP, the Overlap achieved by RAGL and GAGL represented satisfactory results of segmentation based on the optimized Faster R-CNN. Among them, RAGL's Overlap was lower. Through data analysis, in the recognition process using the Faster R-CNN model, a small number of red apples that were partially occluded by green leaves were not recognized as the TP results. Thus, the APtC correspondingly missed some points overlapping with GT^*, then the RAGL's Overlap was reduced. In fact, the evaluation of the APtC was consistent with the result of apple detection by the Faster R-CNN. Moreover, Inaccuracy of RAGL and GAGL were both more than 40%, reflecting that a lot of noise points corresponding to the corners of ROI existed in APtC. Therefore, both Purity of RAGL and GAGL were not satisfactory at all without QCP.

More importantly, Purity can directly reflect the performance of QCP, and QCP algorithm is the key of high Purity. Table 7 shows that the Purity (QCP performed) of RAGL and GAGL much better than the Purity without QCP. Even though QCP-performed Overlap was slightly reduced, the Inaccuracy was reduced by more than 40%, indicating that QCP is of great value. The detail of QCP's effect is shown in Fig. 5. A set of experimental results of the method based on ROI is shown in Fig. 6.

Table 7

Evaluation of RAGL/GAGL segmentation

Method	Category	Evaluation		
		Overlap (%)	Inaccuracy (%)	Purity (%)
RGB ROI	RAGL	95.1	48.7	66.1
	GAGL	95.7	44.8	68.1
RGB ROI (QCP-performed)	RAGL	93.5	3.2	96.7
	GAGL	94.5	3.7	96.2
MBPC	RAGL	96.8	3.8	96.2
	GAGL	97.2	236.3	29.1



(a) (b)

Fig. 5 - Effect of QCP-performed apple point clouds

(a) Raw apple point clouds segmented by RGB ROI, green points are noise that will be filtered by QCP, red points are target points of the apple. (b) QCP-performed apple point clouds of the side facing the camera

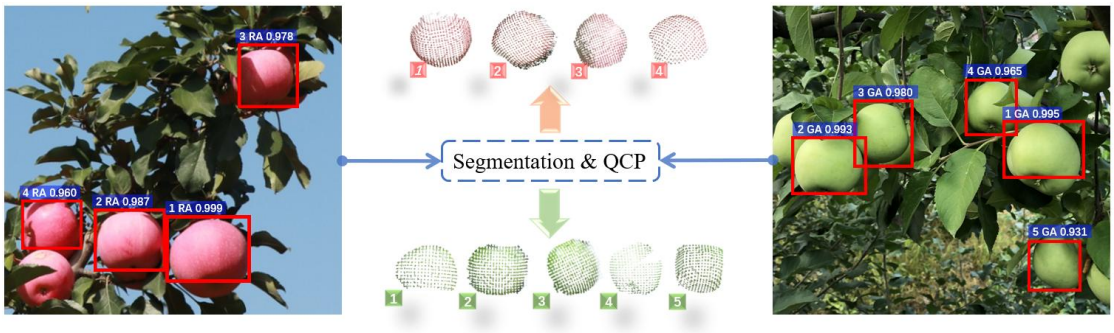
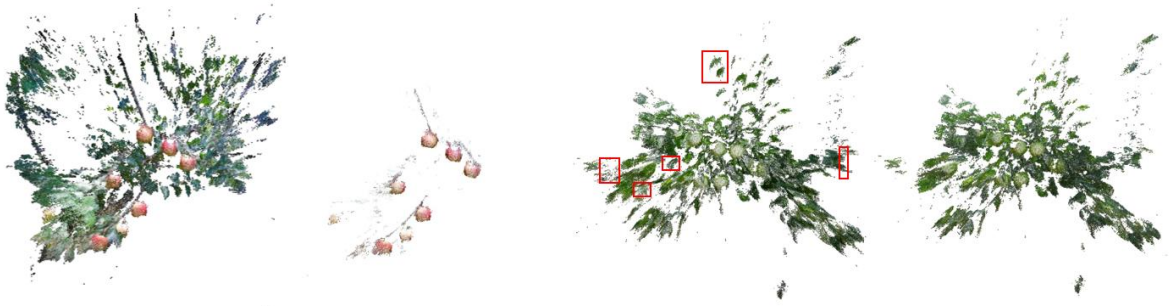


Fig. 6 - A group of results from the method based on ROI

- **Results of comparison methods**

At last, in the results of the MBPC, the segmentation of RAGL point clouds achieved excellent results due to the obvious colour difference between red apples and green leaves. As shown in Table 7, the Overlap of RAGL was even better than the method based on ROI without QCP, and MBPC achieved a great Inaccuracy. ExG can segment RAGL point clouds well, as shown in Fig. 7(a-b). In addition, Purity of RAGL using MBPC was much higher than the method based on ROI before QCP was performed. However, the QCP-performed Purity of the method based on ROI was the best result in segmentation.

When the GAGL was segmented, the Greenness difference between the leaves and the apples was not enough, making the result of segmentation unsatisfactory, the Inaccuracy is over 250%. It shows that points in NBGL were much more than the apple points. Therefore, the score of segmentation was very poor in Purity. As shown in the Fig. 7(c-d), lots of green points of apples and leaves both existed, and the segmentation of GAGL was not completed.



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

Fig. 7 - Results of the segmentation of MBPC

(a) Raw data of RAGL. (b) Results of RAGL data (c) Raw data of GAGL. (d) Results of GAGL data, except for the red marks, MBPC recognizes other parts as apple point clouds, the segmentation effect is bad

CONCLUSIONS

This paper put forward the segmentation method of apple point clouds based on ROI. An annotated RGB dataset of apple trees was built to train Faster-R-CNN to locate ROI in RGB images, which made this method robust in practical application. And then apple point clouds were roughly segmented by the relationship between point clouds and RGB images, and QCP was proposed to refine the apple point clouds. Some conclusions can be reached from the comparison experiments by the MBPC:

(1) The segmentation method based on ROI is simplified by combining the smart Faster R-CNN with the point clouds from consumer-grade RGB-D camera.

(2) QCP can greatly improve the quality of apple point clouds roughly segmented by ROI. Purity is raised from 66.1% to 96.7% for RAGL and from 68.1% to 96.2% for GAGL.

(3) The segmentation method based on ROI is effective for small-size point clouds in unstructured scenes.

ACKNOWLEDGEMENT

This study was supported by National Natural Science Foundation of China [31971668] and Beijing Natural Science Foundation [2182045].

REFERENCES

[1] Bac CW., Henten EJV., Hemming J., Edan Y., (2015), Harvesting Robots for High-value Crops: State-of-the-art Review and Challenges Ahead. *Journal of Field Robotics,* Vol.31, pp. 888-911, Ed.Wiley, USA;

[2] Blasco J. et al., (2003), Machine Vision System for Automatic Quality Grading of Fruit. *Biosystems Engineering,* Ed. Elsevier, pp.415-423, London/U.K.;

[3] Bulanon DM., Kataoka T., Ota Y. et al., (2002), AE-Automation and Emerging Technologies: A Segmentation Algorithm for the Automatic Recognition of Fuji Apples at Harvest., *Biosystems Engineering,* Ed. Elsevier, pp. 405-412, London/U.K.;

[4] Charles RQ., Hao S., Mo K., Guibas LJ., (2017), PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, *IEEE Conference on CVPR,* Ed. IEEE, pp. 77-85, PR/US;

[5] Dai A. et al., (2017), ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes, IEEE Conference on Computer Vision and Pattern Recognition, Ed. IEEE, pp. 2432-2443.;

[6] English A., Ross P., Ball D., (2003), Vision based guidance for robot navigation in agriculture., *IEEE International Conference on Robotics & Automation.*, pp. 1693-1698, Ed. IEEE, Hong Kong/China;

[7] Girshick R., et al., (2014), Rich feature hierarchies for accurate object detection and semantic segmentation, *IEEE Conference on Computer Vision and Pattern Recognition*, Ed. IEEE, pp. 580-587

[8] Girshick R., (2015), Fast R-CNN. Computer Science., *IEEE International Conference on Computer Vision,* Ed. IEEE, pp.1440-1448;

[9] Green WR., Grobler H., (2015), Normal distribution transform graph-based point cloud segmentation, *Pattern Recognition Association of South Africa & Robotics & Mechatronics International Conference*, Ed. IEEE, pp.54–59, Port Elizabeth / South Africa;

[10] Grinblat GL. et al., (2016), Deep learning for plant identification using vein morphological patterns. *Computers & Electronics in Agriculture.* Ed. Elsevier, pp.418-424., London/U.K.;

[11] Hasan MM. et al., (2018) Detection and analysis of wheat spikes using Convolutional Neural Networks, *Plant Methods,* Vol.14, Ed. BMC, Melbourne/Australia.;

[12] Himmelsbach M., Hundelshausen FV., Wuensche HJ., (2010), Fast Segmentation of 3D Point Clouds for Ground Vehicles, *Intelligent Vehicles Symposium*, Ed. IEEE, pp.560–565, California/US;

[13] Huijun Y., (2011), A Novel Algorithm for Segmenting Fruit from Unorganized Point Clouds, *International Conference on Virtual Reality Continuum & Its Applications in Industry,* Ed. ACM, pp.379-382, Hong Kong/China;

[14] Kitamura S., Oka K., (2006), Recognition and cutting system of sweet pepper for picking robot in greenhouse horticulture., *Mechatronics & Automation, IEEE International Conference*, Ed. IEEE, pp.1807-1812, Niagara Falls/Canada;

[15] Linker R., Kelman E., (2015), Apple detection in night-time tree images using the geometry of light patches around highlights. *Computers & Electronics in Agriculture.,* Ed. Elsevier, pp.154-162., London/U.K.;

[16] Li Y., Rui B., Sun M., Chen B., (2018), PointCNN, Shan Dong University/China;

[17] Meyer GE., Neto JC., (2002), Verification of colour vegetation indices for automated crop imaging applications., *Computers & Electronics in Agriculture.,* Ed. Elsevier, pp. 282-293, London/U.K.;

[18] Nguyen TT., et al., (2013), Apple detection algorithm for robotic harvesting using a RGB-D camera*, Computer Methods in Applied Mechanics & Engineering*, pp.06-10, Zurich/Swit;

[19] Pas AT., Platt R., (2018), Using Geometry to Detect Grasp Poses in 3D Point Clouds, *Robotics Research.*, pp. 307-324, Ed. Springer, Boston/US;

[20] Paulus et al., (2013), Surface feature based on classification of plant organs from 3D laser scanned point clouds for plant phenotyping, *BMC Bioinformatics,* Vol.14, Issue 1, pp.1-14, Ed. Springer Nature;

[21] Ren S., He K., Girshick R., Jian S., (2015), Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks., *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Ed. IEEE, pp.1137-1449;

[22] Soria PR., et al., (2017), Multi-view Probabilistic Segmentation of Pome Fruit with a Low-Cost RGB-D Camera, *Iberian Robotics Conference*, Sevilla/Spain;

[23] Tao Y., Zhou J., (2017), An Automatic Segmentation and Recognition Method of Apple Tree Point clouds in the Real Scene Based on the Fusion of Colour and 3D Feature, *ASABE Meeting,* Michigan/US;

[24] Zeiler MD., Fergus R., (2014), Visualizing and Understanding Convolutional Networks., *Lecture Notes in Computer Science*, Ed. Springer, pp.818-833.;

[25] Wahabzada M., et al., (2015), Automated interpretation of 3D laser scanned point clouds for plant organ segmentation, *BMC Bioinformatics,* Vol.16, Issue 1, pp. 248-258, Ed. Springer Nature;

[26] Wang W. et al., (2017) SGPN: Similarity Group Proposal Network for 3D Point Cloud Instance Segmentation, *Computer science,* Ed. Cornell University. New York/US.;

[27] Xu Y. et al., (2018), Unsupervised Segmentation of Point Clouds from Buildings Using Hierarchical Clustering Based on Gestalt Principles*, IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing,* Ed. IEEE, pp.1-17;

Yamauchi H. et al., (2005), Feature Sensitive Mesh Segmentation with Mean Shift., *International Conference on Shape Modeling & Applications*, Ed. IEEE, London/UK.