Computers and Electronics in Agriculture 142 (2017) 388–396

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking

Tao Yongting, Zhou Jun *

Jiangsu Key Laboratory for Intelligent Agricultural Equipment, Nanjing Agricultural University, Nanjing 210031, China

ARTICLE INFO

Article history:
Received 1 March 2017
Received in revised form 12 July 2017
Accepted 12 September 2017
Available online 26 September 2017

Keywords:
Agricultural robot
Apple detection
Point cloud data
Machine learning

ABSTRACT

Accurate apple recognition is a vital step in the operation of robotic fruit picking. To improve robot recognition ability and perception in three-dimensional (3D) space, an automatic recognition method was proposed to achieve apple recognition from point cloud data. First, an improved 3D descriptor (Color-FPFH) with the fusion of color features and 3D geometry features was extracted from the preprocessed point clouds. Then, a classification category was subdivided into apple, branch, and leaf to provide the system with a more comprehensive perception capability. A classifier based on the support vector machine, optimized using a genetic algorithm, was trained by the three data classes. Finally, the results of recognition and lateral comparison were obtained by comparison with the different 3D descriptors and other classic classifiers. The results showed that the proposed method exhibited better performance. In addition, the feasibility of estimating the occurrence of blocking using proposed method was discussed.

1. Introduction

Along with emerging concepts of digital agriculture and intelligent agriculture, the agricultural robot has evolved as a hot topic in agricultural applications (Mai et al., 2015). Consequently, the development of orchard harvesting robot technology has been maturing (Grift et al., 2008). However, in the unstructured, complex, and variable orchard environment, a harvesting robot is required to have the good perceptual capability as well as semantic information acquisition and modeling ability, such that semantic information is accurately extracted from sensing data, and then interpreted to robot movement instructions (Goncalves and Torres, 2010). For example, to achieve accurate and efficient grasping, a robot needs not only to perform rapid and accurate fruit recognition but also to achieve movement trajectory calculations and grasping strategy planning according to the spatial distribution of branches, leaves and fruits, including interfering conditions such as object blocking or overlapping. To achieve this objective, a number of studies have been conducted that have shown promising outcomes (Edan et al., 2009).

In recent years, fruit recognition research has gradually developed from single digital image processing technology into an interdisciplinary research into artificial intelligence and machine learning (Zhao et al., 2016). Research outcomes have also shown that both the accuracy of fruit recognition and performance by a detection system can achieve significant improvement by applying a machine learning method. In addition, with the development of depth information, three-dimensional (3D) data of a scene can be directly and accurately obtained by laser radar or depth camera, which helps the system to more effectively recognize and detect the object and help the robot to perform the grasping operation (Jiang et al., 2016).

In summary, for the purpose of using 3D information more effectively to help fruit picking robots recognize objects accurately, this study proposed an improved 3D descriptor, which not only fully described the 3D geometric characteristics of an object but also improved robustness based on color characteristics. Then, an automatic recognition method was proposed using a support vector machine (SVM), optimized by a genetic algorithm (GA), to recognize apples, branches and leaves in a scenario, thereby benefiting the efficiency of the fruit-harvesting robot.

2. Related work

With respect to fruit recognition and localization, significant achievements have been realized based on two-dimensional (2D) images from a monocular or binocular vision system. For example, (Wachs et al., 2010) have used two levels of visual characterizations for extractions based on maximal interaction information of infrared and color images, which are then used to recognize green apple in a canopy foreground. With recognition rates for

* Corresponding author.
E-mail address: zhoujun@njau.edu.cn (J. Zhou).

https://doi.org/10.1016/j.compag.2017.09.019
0168-1699/

the two levels at 54% and 74%, respectively. Ma et al. (2013) have established a combined target fruit recognition algorithm using a quantum GA and fuzzy neural network, which is capable of global search and self-adaptivity. Specifically, apples with full colors, the recognition rate is 100% and this rate remains high at 96.86% for fruit with uneven illumination, having time consumption of 1.72 s. Moreover, in a study by Ji et al. (2012), they proposed a region growing segmentation method to process 2D image of apple tree. Meanwhile, a SVM classifier was applied to recognize apple, the recognition success rate reached 89%. However, the location of fruits could be further addressed and the success rate needed improved. Jia et al. (2015) have proposed a novel recognition method based on K-means cluster segmentation and optimized radial basis function (RBF) neural networks using GA and least mean square algorithm, for the purpose of further improving the recognition accuracy and speed. The overall recognition rate reached 96.95%, and slightly dropped to 95.38% and 96.17% in conditions of blocking and overlapping, respectively. Although 2D image-based fruit recognition and localization technology have produced promising results, the complex spatial structure is not revealed in 2D images, thus the absence of spatial data interferes with grasp strategy planning in the next step.

The technology of 3D point cloud data acquisition and 3D reconstruction enables accurate restoration of spatial characteristics of fruit trees and provides spatial distribution data for the harvesting robot. Thus, this technology is of great significance to the operation of fruit-harvesting robot. Wahabzada et al. (2015) have used a laser scanner equipped on a robotic arm to obtain point cloud data of grapes, wheat, and barley, and conducted target segmentation and recognition using K-means cluster algorithm. However, due to the absence of RGB information, the related accuracy was undesirable. Instead, Nguyen et al. (2014) have obtained point cloud data with an RGB-D camera, and achieved fruit segmentation and extraction based on a Euclidean clustering segmentation algorithm. Then, achieved fruit recognition and localization. However, the method requires complicated depth and color filtering in order to obtain good results. The self-adaptivity of robotic recognition and interference resistant capability, researchers have implemented algorithms in pattern recognition and machine learning in fruit recognition and localization. The 3D shape of each fruit's point cloud was extracted and its 3D spatial position information and radius were obtained by using the RANSAC algorithm to construct apple model (Mai et al., 2015), thus the recognition and location of fruits were completed. But the step of recognition needed manual instruction and intervention. (Sa et al., 2017) took peduncle of sweet pepper as the study object, performed flower-stalks recognition using support vector classifier based on HSV color features and 3D geometric features extracted from point cloud data. The contribution of color features was analyzed. The outcome was promising, yet improvements with respect to classifier selection and parameter optimization are necessary.

3. Materials and methods

3.1. Framework

Accurate apple recognition and the spatial distribution of fruits and branches was achieved using an RGB-D camera to obtain 3D point could data of apple trees. Thereafter, an automatic apple recognition method was proposed (Fig. 1). Specifically, a RGB-based region growing segmentation method was applied first to obtain separate point cloud data for fruits, branches, and leaves. Then, the 3D features of which were extracted to describe data for different fruit tree parts; the descriptor was composed of color and 3D geometric information. Next, the descriptor was input to the support vector machine to train the classifier, and key parameters of the classifier optimized via GA, such that an intelligent recognition system for classification and recognition of different fruit tree parts was built. Furthermore, unblocked and blocked apple samples were classified to assist the robot in determining whether recognized fruits were blocked during grasping to better instruct the harvesting robot.

3.2. Point cloud obtainment and pre-processing

In this study, Kinect v2 was selected to obtain point cloud data (Yang et al., 2015), as it not only processed excellent performance in depth image accuracy and resolution but also had the ability to function under natural light as well as being low cost. Kinect software for Windows SDK 2.0 was used to compile a point cloud obtainment system in the Visual Studio 2013 platform. This system successfully communicated with Kinect v2 and achieved point cloud data acquisition and storage. Moreover, the system removed redundant information and background in a scene in the process of obtaining data by setting the depth threshold, which simplified the data and improved the data processing efficiency. The parameters of the Kinect v2 was shown in Table 1.

Point cloud data obtained under natural light inevitably contained some outlier points and noises, caused by measurement error or light interference. These errors resulted in complexity in the local point cloud features, such as in surface normals and curvature estimations, which increased computational errors. Therefore, two data preprocessing methods were adopted in this paper, outlier points filtering which can remove the outlier points and noise points; VoxelGrid based downsampling which can simplify point cloud from millions points to thousands points without changing the original shape (Rusu and Cousins, 2011).

3.3. Point cloud segmentation

In this paper, a RGB-based region growing segmentation method was applied to extract independent data of apples, branches and leaves from a scene. This method was based on the color difference $CD(P_1, P_2)$ and Euclidean distance $Dis(P_1, P_2)$ between the seed point and its neighbor point. The definition is shown as follow:

$$CD = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \tag{1}$$

$$Dis = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2} \tag{2}$$

As shown in Fig. 2, the process starts from growing the first point of the input point cloud. This point is called seed and added to the seed set S_s. For every neighbor point, the color difference is calculated between its color value and the current seed's. If the value is less than the color difference threshold T_c which indicates two points colorimetrical similarity, then the point is added to the current region. Meanwhile, the distance value between the neighbor and seed point is calculated. If the value is greater than the distance threshold T_d, then this point is added to the seed set. The procedure iterates reading unlabeled points, growing a new region and not terminating until all points are labeled.

Segmentation example results showed that the scene was divided into three types of data, apple, branch and leaf (Fig. 3), with kind of data labeled and stored for training the classifier.

3.4. 3D features computation

In recent years, in-depth studies have been conducted into extracting descriptive 3D feature with good performance. Local

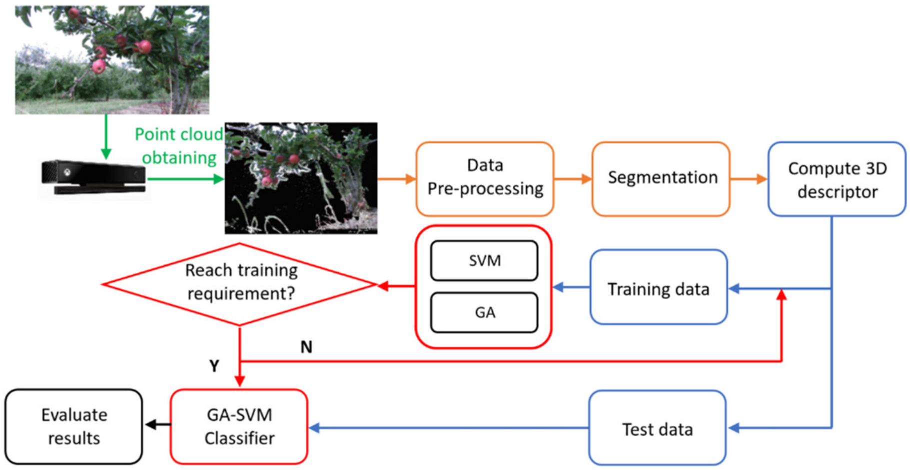


Fig. 1. Flowchart for apple recognition.

Table 1
Parameters of Kinect v2.

Kinect v2	Camera parameter				Experiment parameter	
	FOV	Range (m)	Resolution (RGB)	Resolution (Depth)	Distance to object(m)	Depth threshold (m)
	70° × 60°	0.5–4.5	1920 × 1080	512 × 424	0.8–1.5	0.5–2.0

Algorithm of RGB-based region growing segmentation

Inputs: Point cloud {P}; Color difference function *CD* and threshold T_c; Euclidean distance function *Dis* and threshold T_d; Neighbor searching function *KNN*(.).

Initialize: Region array {R}←Φ; Seed set S_s←Φ.

While any point P_a in {P} has not been labelled, push it in S_s, initialize a new region R_a, labeled P_a as belong to R_a

    while S_s is not empty

        for each neighbor point P_n

            if $Dis(P_a, P_n)>T_d$ or P_n is labelled

                continue

            if $CD(P_a, P_n)<T_c$

                Add P_n to R_a

            end if

        end for

    end while

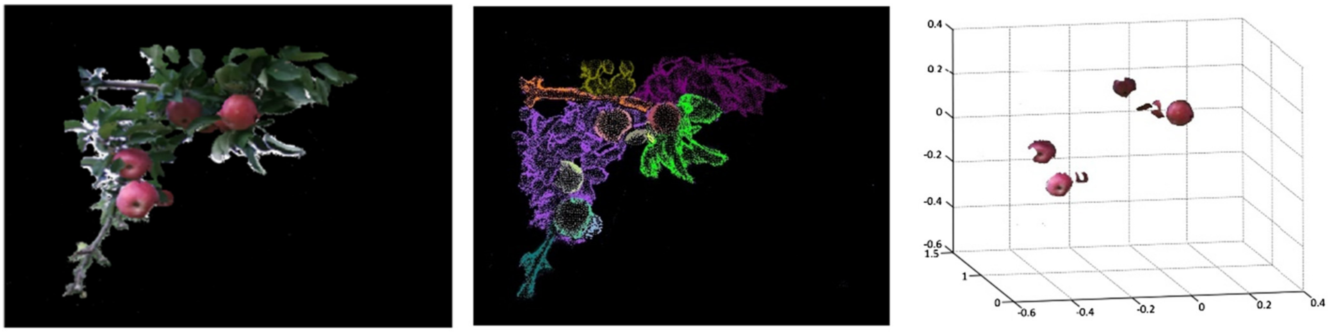
    add R_a to {R}

end while

return {R}

Fig. 2. Algorithm of RGB-based region growing segmentation.

descriptors, such as PFH (Point Feature Histogram) (Rusu et al., 2008), FPFH (Fast Point Feature Histogram) (Rusu et al., 2009) and SHOT (Signatures of Histograms of Orientations) (Tombari et al., 2010), which are extracted from the key points in the object's point cloud and more suitable for the instance recognition and object classification. The global descriptors, such as VFH (Viewpoint Feature Histogram) (Rusu et al., 2010), ESF (Ensemble of Shape Functions) (Wohlkinger and Vincze, 2011) and GRSD (Global Radius-Based Surface Descriptor) (Marton et al., 2010), which are extracted from the entire point cloud of the object and more suitable for the matching. With regard to apple trees, although geometric characteristics were, to a certain extent different for fruit, branches, and leaves, recognition error still existed under some circumstances because of small marginal differences in surface characteristic and similarity in size of significantly blocked fruit and leaves. In view of this problem, an improved feature extraction method was proposed in this study to extract the RGB and HSI (hue, saturation, and intensity) (Weeks and Hague, 1997) color components from point cloud data. These components were then combined with the 3D descriptor FPFH to jointly describe the point cloud data of fruit trees, thereby enhancing the recognition system performance.



(a) (b) (c)

Fig. 3. Result of segmentation (a) pre-processed point cloud, (b) segmented point cloud represented by different color and (c) individual point clouds of apple.

3.4.1. Color feature computation

Because of color differences between fruit, branches, and leaves, a method for extracting color features to identify each class was feasible. Although RGB information is able to describe an object's color, considering the effects of natural sunlight during outdoor operations, HSI information was further computed by RGB values as well (Li et al., 2002). HSI color space was consistent with human eye color perception and showed fairly good light interference resistant capability. Therefore, 6 color components in RGB and HSI color space were extracted to represent the color features of each point in object point cloud data.

3.4.2. 3D geometric feature computation

In the 3D space of point cloud data, surface normal and curvature estimations are two basic notations of geometric characteristics around a certain point. These values were easily calculated, yet much information was not obtained. By means of a parameterized query of the spatial differences between a point and its adjacent area, the FPFH descriptor formed a multidimensional histogram to describe the geometric properties within the k neighborhood of a point. This offered the advantages of rotation invariance and good robustness under different sampling densities and adjacent noise levels. The relative deviation between normal n_1 and n_2, which corresponded to two points p_1 and p_2, was calculated by defining a fixed local coordinate system on one point, as shown in Fig. 4. A set of angle values were used to indicate the deviations between the two points using *UVW* coordinates, as shown in Eq. (3).

$$\begin{aligned} \alpha &= \mathrm{V} \cdot n_2 \\ \varnothing &= \mathrm{U} \cdot \frac{p_1 - p_2}{d} \\ \theta &= \arctan(\mathrm{W} \cdot n_2, \mathrm{U} \cdot n_2) \\ \mathrm{d} &= \|p_2 - p_1\|_2 \end{aligned} \tag{3}$$

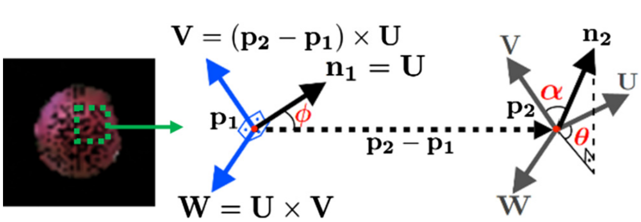


Fig. 4. *UVW* coordinate.

First, the values of α, φ, and θ between each query point and neighbor point in k nearest neighbors were calculated. Then, this feature set, representing the neighbor area, was obtained by constant iteration of feature weights in the neighbor area. Each value was divided and statistically calculated, thereby eventually obtaining a histogram representing the distribution features of the point, which was essentially a 33Dimensional eigenvector.

The calculation procedure for the Color-FPFH descriptor was used after color features and 3D geometric features were computed. Then, the 6-dimensional color feature set and 33-dimensional FPFH feature set were combined and normalized, generating a 39-dimensional Color-FPFH point cloud descriptor (Fig. 5). For the purpose of comparison, FPFH, SHOT, VFH and ESF features were also extracted from the target point cloud, their basic attributes are shown in Table 2. These features were calculated based on PCL (point cloud library) (Rusu and Cousins, 2011) using default parameters. All features were then imported into the classifier for training, an experiment was then conducted to confirm the performance of Color-FPFH descriptor in apple recognition.

3.5. Optimized support vector classifier based on GA

Similar to the multilayer perception network and RBF (Radial Basis Function) network, SVM exhibited outstanding performance in pattern classification recognition and nonlinear regression. The core idea was to set up a classification hyperplane as the decision surface, thereby maximizing the isolation edge between positive and negative examples. SVM showed fairly good generalization ability in pattern classification and strong robustness and calculation ability, such that it was used in many classification problems. After analyzing the 3D features, a support vector classifier based on RBF was adopted in this study. Despite RBF-SVM having high accuracy in dealing with discrete data classification, its core parameters, such as the penalty parameter c and kernel parameter g, had great influence on classifications. Thus, selecting the optimal parameters was a key step in using SVM classifier. Generally, manually calculating optimal parameters in the multidimensional data structure was clearly not realistic, and using a cross validation method to search for the global optimal solution resulted in long computation time because of high features dimensions and large amount of data. Consequently, a heuristic parameters optimization method, based on GA was adopted in this study (Lessmann et al., 2006). The system was able to adapt itself to optimal parameters until finding the globally optimal solution, thus improving the accuracy of classification predictions. The GA-SVM algorithm flow is shown in Fig. 6. The advantage of combining GA and SVM was discussed in detail in the next section via comparison of experiment results (Li and Kong, 2014).

```
Algorithm of Color-FPFH Descriptor
procedure GetFeatures
                                                        Color Feature
    ▶R,G,B value extraction
        feature(1) (2) (3) ← R, G, B value
    ▶H,S,I value computation
        feature(4) (5) (6) ← H, S, I value
                                                        3D Geometry Feature
    ▶Estimating the normals and α, ϕ, θ value computation
    ▶Dividing feature and statistical calculation in neighbor
        feature(7) – (39) ← Fast Point Feature Histogram
    ▶Merge color and geometry features
    ▶Normalization processing
return feature
end procedure
```

Fig. 5. Color-FPFH Descriptor calculation procedure.

Table 2
Basic attributes of 3D descriptors (N = number of points in input point cloud, Normal = whether to calculate the normal of the points).

Name	Type	Size	Normal
Color-FPFH	Local	N * 39	Yes
FPFH	Local	N * 33	Yes
SHOT	Local	N * 352	Yes
VFH	Global	308	Yes
ESF	Global	640	No

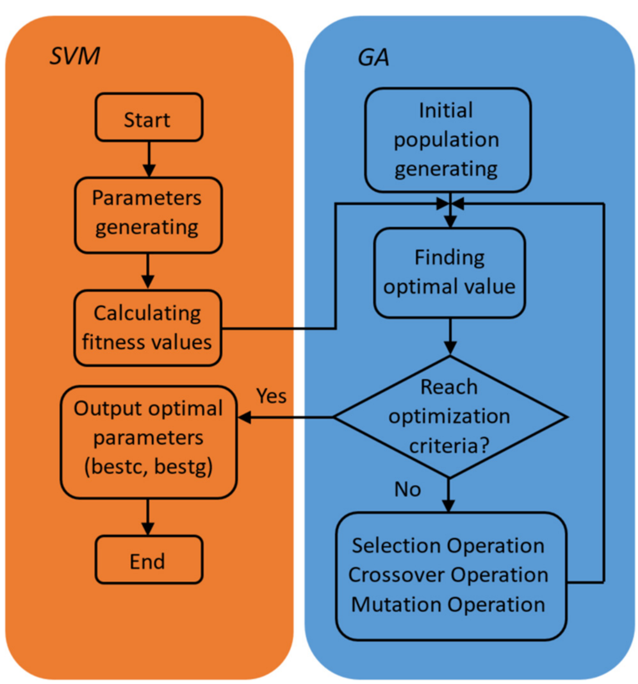


Fig. 6. Flowchart of GA-SVM.

4. Experiment in recognition

In this study, experimental data was collected in an apple orchard in Cobb County, Atlanta, USA during September 2016. Data acquisition was described in Section 3.2, and point cloud data of mature Fuji apples were collected. The fruit tree point cloud obtaining system was used to obtain 52 point cloud data from 8 apple trees from different views under natural sunlight conditions. After preprocessing, the obtained apple tree point cloud data were segmented as mentioned in Section 3.3, obtaining three kinds of point cloud data for apples, branches, and leaves, and the data then stored, as shown in Fig. 7.

4.1. Training the GA-SVM classifier

First, training and test data sets was created; data distribution is shown in Table 3. Although blocking classification was roughly conducted for apples, only three types of data were labeled. Labeling blocked apples was, to gain accurate recognition of blocked apples to validate method the robustness, and also, to prepare for experiments in estimating whether apples were blocked or not. The training steps of GA-SVM classifier with optimal parameters were as follows:

(1) Following the above processing, the features of three classes' of point clouds were extracted. As the proposed Color-FPFH descriptor was for the local descriptor, it only reflected the relationship between a certain point and its neighbor. Although its single feature dimension was 39, $n * 39$ feature dimensions were obtained if the features of point cloud with n points were calculated. Obviously, calculating and training the features of a point cloud with hundreds of points would increase system consumption and, also, the redundant features compromised classification results. Therefore, as it is necessary to extract key points from the object point cloud, NARF key points were applied in this study.
(2) Features of all sample data were calculated and stored according to the flowchart as shown in Fig. 4.
(3) All labeled data were imported into SVM for training, and the GA was then used to optimize penalty parameter c and kernel parameter g, thereby obtaining global optimal parameters. In this study, the optimal parameters obtained from the Color-FPFH descriptor training were c = 17.35, and g = 0.46, respectively.

4.2. Results of recognition

The automatic apple recognition system based on GA-SVM was established in the above section, and all test data was then input

Fig. 7. Point cloud data instance of apples, branches and leaves.

Table 3
Samples dataset.

Category	Train (80%)	Test (20%)	Total number
Apple (Blocked area < 50%)	156	39	195
Apple (Blocked area < 50%)	82	20	102
Branch	54	13	67
Leaf	181	45	226

into the system to obtain prediction results. The recognition results obtained from the Color-FPFH descriptor are shown in Table 4. The recall, precision, and accuracy were computed as below to fully evaluate the system performance:

$$\text{Recall} = \frac{T_P}{T_P + F_N}, \ \text{Precision} = \frac{T_P}{T_P + F_P}, \ \text{Accuracy} = \frac{T_P + T_N}{P + N} \quad (4)$$

where P is the number of total positive samples, N is the number of total negative samples, T_P is the number of true positives (correct recognition), F_P is the number of false positives (false recognition), F_N is the number of false negatives (miss recognition) and T_N is the number if true negatives (correct rejection). Specifically, recall indicated the proportion of correctly recognized samples among all samples, which represented the system recognition ability to identify positive samples. Precision reflected the ratio of true positive examples to the total number of recognized positive samples. The accuracy was used to reflect recognition system ability in judging the whole sample. For apples, the proposed recognition method showed good outcomes for the these three performance measurements. The system benefitted from the large proportion of apple samples and embodied the robustness of the Color-FPFH descriptor. In addition, optimization of parameters based on GA also played a key role. Compared to apples, the recognition of branches and leaves showed relatively low recall and precision due to the small sample number and over-fitting or under-fitting that resulted from indistinct classifying boundaries and large differences among individual point clouds.

A set of apple data from the same scenario was randomly selected, for prediction by the system (Fig. 8), and the system correctly recognized the four unblocked or slightly blocked apples, indicating high recognition accuracy, but, of three seriously blocked apples in this scenario, the system identified only two of them. Blocking could led to inaccurate acquisition of actual color information in conditions of low light while obtaining a point cloud, which therefore reduced the distinguishing ability of color features. As the geometric features of apples become unstable as a result of blocking, apples might be recognized as branches or leaves. Regardless, the experiment results showed that the proposed recognition method possessed adaptivity to orchard scenarios. Generally, its outstanding performance in recognition of the objects of interest, --apples, met the needs of the fruit harvesting robot.

4.3. Evaluation

4.3.1. Evaluation of features

The recognition ability of the 3D descriptor proposed in this study was evaluated by adopting and comparing the other four 3D descriptors mentioned in Table 2 for the recognition of apples, branches, and leaves (Table 5). The Color-FPFH descriptor outperformed the other methods in all three kinds of data recognition. Compared to FPFH and SHOT, Color-FPFH demonstrated significantly increased accuracy in apple recognition after including color characteristics, which proved the feasibility of combining color and 3D geometric features. In terms of the global descriptors VFH and ESF, they did not show any advantages because of complicated calculations, using simple descriptions of the geometric characteristics. The result might have been biased to evaluate just one model or feature by purely relying on recognition accuracy. Therefore, the contribution of each descriptor to the recognition system was evaluated by introducing the a precision-recall curve.

P-R curves for three kinds of data recognition using different 3D descriptors showed that the P-R curve can visually display the precision and recall ability of each recognition system on overall samples (Fig. 9). Thus, P-R curves were generally used to evaluate the performance of a classifier, as well as the performance of a feature in the classifiers. When one P-R curve was entirely surrounded by another, the latter was asserted to have better performance than the former. When two P-R curves intersected, the Area Under Curve (AUC) value was calculated to evaluate their performance.

Table 4
Accuracy, recall and precision rates for each organ of apple tree.

Category	Correct samples	Total samples	Recall	Precision	Accuracy
Apple	53	59	89.83%	94.64%	92.30%
Branch	8	13	61.53%	47.05%	88.03%
Leaf	33	45	73.33%	75.00%	80.34%

Fig. 8. Apple recognition result in point cloud.

Table 5
Classification accuracy rates by different 3D description.

Category	Color-FPFH	FPFH	SHOT	VFH	ESF
Apple	92.30%	75.21%	73.50%	75.21%	86.32%
Branch	88.03%	81.20%	82.90%	78.63%	88.03%
Leaf	80.34%	68.38%	70.08%	55.56%	74.35%

The higher the AUC value, the better the performance. The AUC value of Color-FPFH was the highest in all three kinds of data, indicating its superior performance in recognition and object description (Fig. 9).

These experimental results were also evaluated by comparing the computation time used by different descriptors (Table 6). Training and prediction results were all conducted in computers with a core i7 3.2 GHz CPU. The computational speed of the Color-FPFH descriptor ranked in the middle, while VFH and FPFH with lower dimensions showed higher speeds. However, these two descriptors showed the worst performance in terms of recognition accuracy. Overall comparisons showed that Color-FPFH exhibited the best performance, which further indicated the superiority of the proposed descriptor.

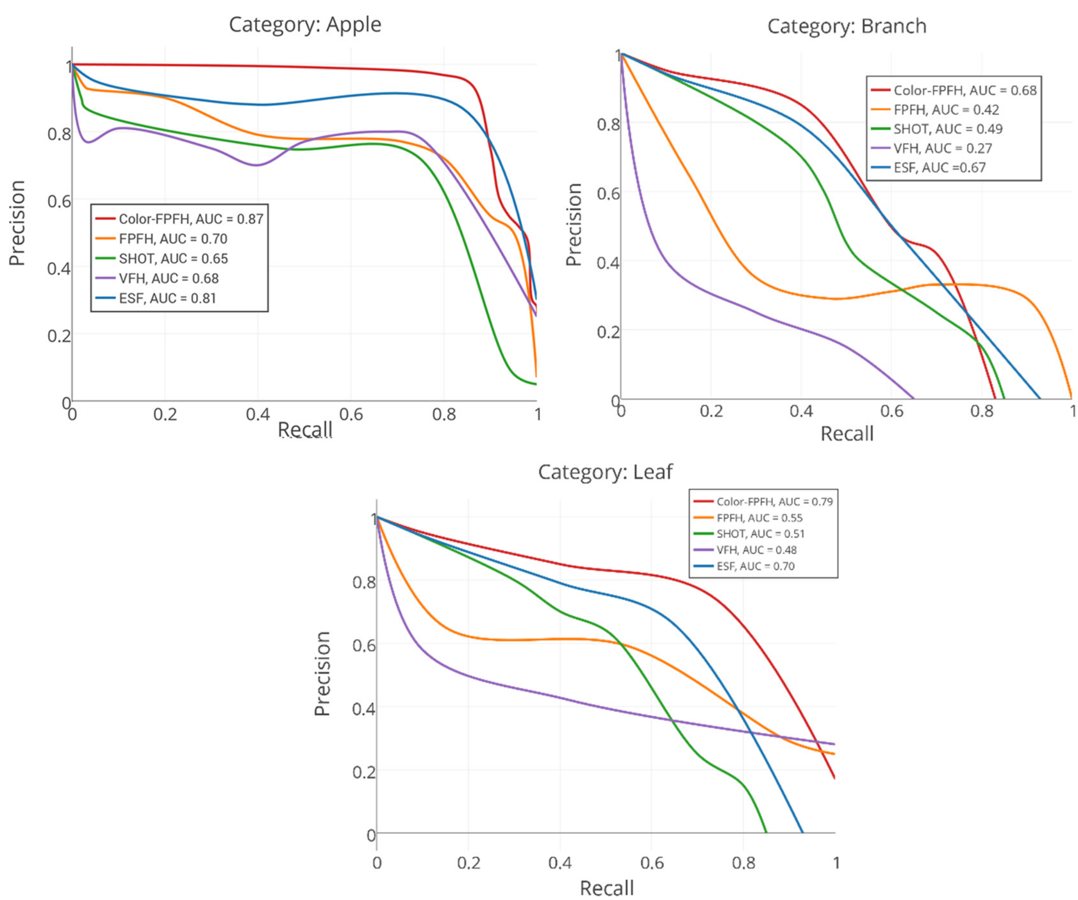


Fig. 9. Precision-Recall curve of different 3D descriptor.

Table 6
Computation time of each descriptor.

Descriptor	Color-FPFH	FPFH	SHOT	VFH	ESF
Feature dimensions	n * 39	n * 33	n * 352	308	640
Training time (sec)	452	367	2751	175	482
Testing time (sec)	12	9	114	5	21

Table 7
Classification accuracy rate by different classifier.

Category	GA-SVM	SVM	KNN	RF
Apple	92.30%	87.18%	80.34%	83.76%
Branch	88.03%	84.61%	90.59%	85.47%
Leaf	80.34%	75.21%	74.35%	77.78%

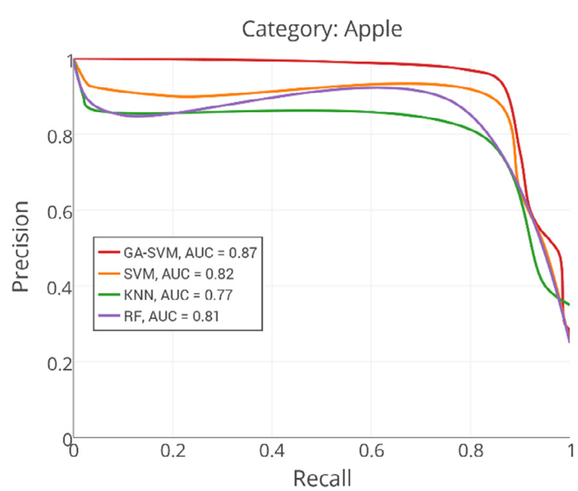


Fig. 10. Precision-Recall curve of apple recognition by different classifier.

4.3.2. Evaluation of classifiers

The proposed Color-FPFH descriptor was shown to have better robustness in fruit recognition in apple tree point cloud data, as well as great advantages in recognition accuracy and system computation time. Moreover, the performance of the selected optimized support vector classifier GA-SVM based on GA needed to be evaluated. Although high recognition accuracy, precision and recall in apples were obtained, the lateral comparison provided more comprehensive data. Therefore, the Color-FPFH feature was adopted to train classifiers of several classic machine learning algorithms (i.e. KNN (K Nearest Neighbor), SVM and RF (Random Forest)), which were then used to test the sample data.

The GA optimized SVM has higher recognition accuracy compared not only to the traditional SVM algorithm, but the other two algorithms as well (Table 7). Then, the P-R curves of the interested object -- apple, based on different classifiers, were plotted (Fig. 10). The AUC value of GA-SVM is the highest, which indirectly reflected the superior performance of the GA-SVM classifier and thus shown the feasibility and necessity of proposing a GA optimized SVM classifier.

Table 8
Accuracy of blocked apple estimation.

Descriptor	Color-FPFH	FPFH	SHOT	VFH	ESF
Accuracy	73.46%	57.38%	61.82%	42.38%	78.29%

5. Estimating on blocking of apples

The objective of this study was not only to realize accurate recognition of apples using a RGB-D camera, but also to discuss how a recognition system can improve the efficiency of fruit harvesting robots. Generally, the images of fruit trees collected by robots from one view were complicated, such as by object blocking, overlapping and objects in great depth. Although an automatic recognition system with good robustness enabled the robot to recognition of all objects from one certain view in regardless of object blocking or overlapping, it needed to exclude seriously blocked objects during its movement planning to avoid grasping failures because of inaccurate positioning or branch interference. As an example, while almost all apples were recognized in the results in Fig. 8, only the four in the front could be successfully grasped. The remaining two apples need to be excluded in movement planning to ensure a good success rate. The seriously blocked apples might have become visible from another view when the robot moved, in which case, after recognition and blocking estimation, the robots would make a decision whether to harvest these apples or not.

In summary, the proposed algorithm for apple recognition was used in this experiment to allow a robot to estimate conditions with interfere harvesting. First, apple sample data was subdivided and labeled. Sample apple blocked less than 50% were taken as positive examples and those with more than 50% blockage as negative. These data were then input into the GA-SVM classifier for training, and prediction results produced.

The proposed 3D descriptor did not stand out in recognition of the same type of samples, but it still excelled in terms of accuracy (Table 8). On one hand, the color characteristics played no more of an important role in the same data type than that in different data type, with the color characteristics of seriously blocked apples certainly different from those unblocked or slightly blocked. This was the fundamental reason why Color-FPFH showed higher recognition accuracy than the other two local descriptors. On the other hand, although a local descriptor, Color-FPFH described the

Fig. 11. Blocked apple detected in point cloud.

geometric characteristic of an object, but under fitting could easily happen during the recognition process because of the high feature-similarity degree in the same data type. However, because of the good robustness of ESF in geometric characteristic description of target objects, the ESF descriptor showed the best performance in terms of recognition accuracy.

Improvements should be made for accurately determining the condition of fruit blocking. The key of this study was to discuss the feasibility, that is, to estimate the occurrence of blocking on the basis of accurate recognition of apples using features and classifiers. Seriously blocked apples were labeled with a red bounding box, such that the robot would perform localization and movement planning, giving normal apple, with an (orange bounding box) in priority for harvesting (Fig. 11).

6. Conclusion

In this study, because a harvesting robot needs good performance ability for fruit recognition and spatial perception, an apple tree's point cloud data was used as the processing object. After analyses, an improved 3D descriptor was proposed that used color fusion and 3D geometry information as features for describing apples, branches, and leaves. An automatic recognition classifier based on a GA-optimized support vector machine was trained to classify apples, branches, and leaves from an apple tree's point cloud data. The results showed high recognition accuracy and performance by the proposed method, and the feasibility of using the proposed descriptor was discussed for estimating blocked apples. The results of this study could contribute to providing a reference for fruit recognition and information extraction in agricultural robots.

In the further research, the classifier will be further optimized and the computing speed improved. At the same time, improved 3D descriptors will be considered for more agricultural applications.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31471419), Doctoral Program foundation of Ministry of Education of China (No. 2013009711043).

References

Edan, Y., Han, S., Kondo, N., 2009. Automation in agriculture, Springer handbook of automation. Springer, pp. 1095–1128.

Goncalves, P., Torres, P., 2010. Learning approaches to visual control of robotic manipulators. In: The Second International Conference on Advanced Cognitive Technologies and Applications. Citeseer, pp. 103–108.

Grift, T., Zhang, Q., Kondo, N., Ting, K., 2008. A review of automation and robotics for the bioindustry. J. Biomech. Eng. 1, 37–54.

Ji, W., Zhao, D., Cheng, F., Xu, B., Zhang, Y., Wang, J., 2012. Automatic recognition vision system guided for apple harvesting robot. Comput. Electr. Eng. 38, 1186–1195.

Jia, W., Zhao, D., Liu, X., Tang, S., Ruan, C., Ji, W., 2015. Apple recognition based on K-means and GA-RBF-LMS neural network applicated in harvesting robot. Trans. Chinese Soc. Agric. Eng. 31, 175–183.

Jiang, L., Koch, A., Zell, A., 2016. Object recognition and tracking for indoor robots using an rgb-d sensor, Intelligent Autonomous Systems 13. Springer, pp. 859–871.

Lessmann, S., Stahlbock, R., Crone, S.F., 2006. Genetic algorithms for support vector machine model selection. In: Neural Networks, 2006. IJCNN'06. International Joint Conference on. IEEE, pp. 3063–3069.

Li, J.-F., Wang, K.-Q., Zhang, D., 2002. A new equation of saturation in RGB-to-HSI conversion for more rapidity of computing. In: Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on. IEEE, pp. 1493–1497.

Li, X., Kong, J., 2014. Application of GA-SVM method with parameter optimization for landslide development prediction. Nat. Hazards Earth Syst. Sci. 14, 525.

Ma, X., Liu, G., Zhou, W., Feng, J., 2013. Apple recognition based fuzzy neural network and quantum genetic algorithm. Trans. Chin. Soc. Agric. Mach. 44, 227–251.

Mai, C., Zheng, L., Sun, H., Yang, W., 2015. Research on 3D reconstruction of fruit tree and fruit recognition and location method based on RGB-D Camera. Trans. Chinese Soc. Agric. Mach. 46, 6.

Marton, Z.-C., Pangercic, D., Rusu, R.B., Holzbach, A., Beetz, M., 2010. Hierarchical object geometric categorization and appearance classification for mobile manipulation. In: Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on. IEEE, pp. 365–370.

Nguyen, T.T., Vandevoorde, K., Kayacan, E., De Baerdemaeker, J., Saeys, W., 2014. Apple detection algorithm for robotic harvesting using a rgb-d camera. In: International Conference of Agricultural Engineering, Zurich, Switzerland.

Rusu, R.B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. In: Robotics and Automation, 2009. ICRA'09. IEEE International Conference on. IEEE, pp. 3212–3217.

Rusu, R.B., Bradski, G., Thibaux, R., Hsu, J., 2010. Fast 3d recognition and pose using the viewpoint feature histogram. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on. IEEE, pp. 2155–2162.

Rusu, R.B., Cousins, S., 2011. 3d is here: Point cloud library (pcl). In: Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, pp. 1–4.

Rusu, R.B., Marton, Z.C., Blodow, N., Beetz, M., 2008. Persistent point feature histograms for 3D point clouds. In: Proc 10th Int Conf Intel Autonomous Syst (IAS-10), Baden-Baden, Germany, pp. 119–128.

Sa, I., Lehnert, C., English, A., McCool, C., Dayoub, F., Upcroft, B., Perez, T., 2017. Peduncle detection of sweet pepper for autonomous crop harvesting—combined color and 3-D information. IEEE Robot. Automat. Lett. 2, 765–772.

Tombari, F., Salti, S., Di Stefano, L., 2010. Unique signatures of histograms for local surface description. In: European Conference on Computer Vision. Springer, pp. 356–369.

Wachs, J.P., Stern, H., Burks, T., Alchanatis, V., 2010. Low and high-level visual feature-based apple detection from multi-modal images. Precision Agric. 11, 717–735.

Wahabzada, M., Paulus, S., Kersting, K., Mahlein, A.-K., 2015. Automated interpretation of 3D laserscanned point clouds for plant organ segmentation. BMC Bioinformatics 16, 248.

Weeks, A.R., Hague, G.E., 1997. Color segmentation in the HSI color space using the K-means algorithm, Electronic Imaging'97. International Society for Optics and Photonics, pp. 143–154.

Wohlkinger, W., Vincze, M., 2011. Ensemble of shape functions for 3d object classification. In: Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on. IEEE, pp. 2987–2992.

Yang, L., Zhang, L., Dong, H., Alelaiwi, A., El Saddik, A., 2015. Evaluating and improving the depth accuracy of Kinect for Windows v2. IEEE Sens. J. 15, 4275–4285.

Zhao, Y., Gong, L., Huang, Y., Liu, C., 2016. A review of key techniques of vision-based control for harvesting robot. Comput. Electron. Agric. 127, 311–323.