GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition

Xiaowei Wang[1], Shoulin Yin[1,*], Ke Sun[1,*], Hang Li[1], Jie Liu[1], and Shahid Karim[2]

[1] Software College, Shenyang Normal University Shenyang, P.C 110034 - China
[2] Department of Computer Science, ILMA University, Karachi, Pakistan
[*] Corresponding author. E-mail: 352720214@qq.com; ssxbsk@126.com

Received: Oct. 10, 2019, Accepted: Jan. 05, 2020

In this paper, we propose a new apple segmentation and recognition method based on improved Gaussian ker-nel combining fuzzy c-means and convolutional neural network. The importance of determining data distribution characteristics is analyzed. The convolution neural network with good self-learning ability is used to super-vised learn the image and extract the image features. Meanwhile, the modified fuzzy c-means is used for feature clustering analysis. We modify the selection of radial width to improve Gaussian kernel function and use it for support vector machine, which will classify the extracted features. Finally, experiments on Fuji apple images demonstrate that the robustness stability and accuracy of the proposed algorithm is better than other state-of-the-art representative methods.

Keywords: Image segmentation; Fuzzy c-means; Gaussian kernel; Convolutional neural network; Support vector machine; Apple recognition

http://dx.doi.org/10.6180/jase.202009_23(3).0020

1. Introduction

With the development of science and technology, the research of fruit picking robot in fruit trees is becoming more and more perfect. As a part of the picking robot, the speed, accuracy and adaptability of object detection of visual identification system to the surrounding environment have a great impact on the working efficiency and working hours of the robot. Stable object recognition allows robots to work long hours, reducing labor costs and improving productivity. Image segmentation [1, 2] is the preprocess of image classification, scenario analy-sis, object detection, image 3D reconstruction tasks. And they are widely used in medical image analysis, traffic control, weather forecast, geological exploration, face and fingerprint identification, and many other fields [3–5].

Deep learning is a new field in machine learning research. Its motivation is to build a neural network to simulate human brain for analytical learning including a series of machine learning algorithms. High-level abstrac-tions in data are modeled by adopting a deep architecture composed of multiple nonlinear transformations. In deep learning frameworks, convolutional neural network (CNN) is widely used in image classification, object detection, semantic segmentation, face recognition and other fields due to its special structure of local weight sharing, good fault-tolerant ability, parallel processing ability and self-learning ability [6–8]. Kuang [9] proposed an image segment method based on CNN and SVM, which applied the depth features generated by CNN into segmentation. Using linear SVM for training, object and background could be separated to a large extent.

Traditional image segmentation methods mainly include threshold method, boundary detection method and region-based method [10–13]. However, in the complex scene, the actual segmentation effect is not ideal. Therefore, Moreno [14] introduced a novel objective function based on Dice coefficient. Moreno [15] used a hybrid chromatic distance inspired in the human vision system which shifted from the chromatic to the grey-scale distance depending on the pixel's luminance value. Zhang [16] presented a novel level set method for im-age segmentation in the presence of intensity inhomogeneity. However, there are still problems by using the above methods, such as amount of training samples and time consumption.

Therefore, we propose an improved Gaussian kernel function and support vector machine for apple image segmentation and recognition. The learning layer in CNN network can extract the image features more effec-tively and avoid the complex feature extraction and data reconstruction process in the traditional recognition algorithm. The difference between FCM clustering and traditional clustering lies in that it changes the confusing phenomenon in traditional classification. An object can belong to multiple classes at the same time in different degrees. Due to the characteristics of image data information, the information does not necessarily belong to a fixed category, but what the user needs to find is the most interesting and intuitive data. This satisfies the idea of FCM clustering, which can cluster the most similar feature data into a category according to the features ex-tracted by CNN.

The structure of our paper is organized as follows. In Section 2, the proposed apple detection and organiza-tion method is presented. Experiment analysis of our method is presented in Sections 3. Finally, Section 4 pre-sents the conclusion.

2. Proposed image segmentation method

2.1. Feature extraction by CNN

CNN is a kind of feed-forward neural network in deep learning, which is widely used in the field of computer vision due to its strong self-learning ability and unsupervised learning mode [17]. At present, CNN has been ap-plied into the target recognition based on the similarity of the underlying features such as color, texture and shape for detection without considering the high-level semantic features of the image. It can appear the problem of "semantic gap". CNN simulates the process of cellular visual information and obtains richer semantic fea-tures through multi-level network learning. Compared with the ordinary neural network model, CNN can obtain more comprehensive features by deepening the depth of the network.

This paper adopts the GoogLeNet model to extract the features of images. The network adopts the idea of sparse learning including many CNN blocks such as convolution, normalization and pooling. There are many CNN models, and different network architectures have different processing effects. LeNet network is one of the most classic CNN that can recognize handwritten data sets, but it consumes a lot of memory (without GPU con-figuration). The AlexNet network model has lower layers and higher error rate. The VGG16 network model has an uniform structure and is effective in image classification and other tasks. Compared with the above network models, the GoogLeNet model has higher network layers and richer features.

GoogLeNet network consists of 8 layers. From first layer to fifth layer, it includes convolution layer, excitation layer, pooling layer, etc. The sixth, seventh, eighth layer are the full connection layer. GoogLeNet adopts a mod-ular structure for addition and modification with a total of 22 layers. In addition, in order to avoid the disap-pearance of the gradient, the network also adds two additional auxiliary normalization processes for conducting the gradient forward. The image feature vector selected in the experiment is the data in the 20th layer. After testing, the feature data can fully display the features of the image.

GoogLeNet model was used to extract image features. Firstly, the pixel value matrix of each image is ob-tained as the input of the network, which is successively processed through convolution layer, excitation func-tion, pooling layer, convolution, pooling, and then input to the full connection layer to finally obtain the feature vector representation. In the convolution layer, several trainable filtering matrices are convolved with the origi-nal image. Convolution is the inner product of the filter matrix and the original image pixel matrix. The input of pooling layer generally comes from the previous convolution layer, whose main function is to provide strong robustness, reduce the number of parameters and prevent the occurrence of over-fitting.

Support vector machine (SVM) algorithm is in characteristic space of the training sample to separate two classes of samples which there are no errors of getting maximum separation hyper-plane. And mathematically, it is expressed as a convex quadratic programming problem [14].

The optimal hyper-plane SVM method is proposed based on optimal hyper plane under linear separable condition. Set a sample set of linear separable.

$$(x_i, y_i), i = 1, 2, ..., n, ; x \epsilon R, y, \epsilon \{-1, 1\} \tag{1}$$

The classification plane equation is:

$$\varpi . x + b = 0 \tag{2}$$

This plane separates two sample classes without error and makes sample closest to classification plane have the biggest distance to the classification plane. Namely, the classification interval is the biggest, which is equivalent to minimize the $\|w\|^2$, w is the normal vector of classification plane. However, classification plane us re-quired to classify all samples correctly. Constraint condition is:

$$y_i(\varpi . x_i + b) - 1 \geq 0 \tag{3}$$

Therefore, meeting the above conditions is the optimal classification plane. In the two kinds of samples, the point closest to classification plane and parallel to the optimal classification hyper-plane H1, H2, the training sample is called support vector.

The optimal classification surface can be expressed as the optimization problem of the following constraints, namely, the function is evaluated under the constraint of equation 3.

$$\varphi(\varpi) = 0.5 \left\|\varpi\right\|^2 = 0.5(\varpi\varpi) \tag{4}$$

So it can define the following Lagrangian functions.

$$L(\varpi, b, a) = 0.5 \left\|\varpi\right\|^2 - \sum_{i=1}^{n} a_i \left[y_i(\varpi.x + b) - 1\right] \tag{5}$$

Where $a_i > 0$ is Lagrangian coefficient.

The original problem can be transformed into the following simple dual problem:

$$maxQ(a) = \sum_{i=1}^{n} a_i - 0.5 \sum_{i,j=1}^{n} a_i, a_j, y_i, y_j(x_i, x_j) \tag{6}$$

Where $\sum_{i=1}^{n} y_i a_i = 0, a_i \geq 0$

The above is discussed optimal and generalized linearity classification function of nonlinear SVM. To solve a problem of optimal linear classification in feature space, it just needs to know the inner product operation in space. According to the generalized linear discriminant function, to solve a nonlinear problem, it can try to trans-form it by nonlinear transformation for another space linear problem. In this transformation space, we can find the optimal or the most generalized classification plane. Consider Mercer condition, for any symmetric function $K(x, x^{'})$, it is a sufficient and necessary condition for the integral operation in the characteristic space. For any $\int \varphi^2(x)dx < \infty$ and $\varphi(x) \neq 0$, so $\int\int K(x, y)\varphi(x)\varphi(y)dxdy > 0$. If inner product replaces the dot product of the optimal classification, which is equivalent to change the original eigenspace to a new feature space, while the support vector machine is,

$$maxQ(a) = \sum_{i=1}^{n} a_i - 0.5 \sum_{i,j=1}^{n} a_i, a_j, y_i, y_j K(x_i, x_j) \tag{7}$$

Where $\sum_{i=1}^{n} y_i a_i = 0, C \geq a_i \geq 0$. C is a constant. It has the effect of controlling the punishment degree of the wrong classification sample, and balances the proportion of the wrong sample and the complexity of the algo-rithm.

The corresponding discriminant function is also changed as:

$$\int(x) = sgn[\sum_{i=1}^{n} a_i^* y_i K(x_i + x) + b^* \tag{8}$$

In here, a_i^* is the optimal solution; b^* is the classification threshold, which is obtained by a support vector and also can be obtained by any pair of support vectors in two classes.

For a given training set, the data distribution of the geometric feature could be known in advance including Gaussian distribution, circular distribution, ring distribution, columnar distribution, etc. For this kinds of regular data sets, it can construct the corresponding kernel function to train the SVM to improve the generalization ability of SVM. However, it does not know the distribution features of the training data set in advance, so it needs to construct some algorithms, and give data distribution approximation as an important reference index of kernel function and parameters selection of SVM so as to improve the generalization ability of SVM. For those training data set cannot determine the geometric distribution, in the process of SVM training, the Gaussian kernel function or polynomial kernel function is usually used.

To avoid high dimensional operation, we use a kernel function $\Re(x, y)$ satisfying the Mercer condition. The inner product operation of high-dimensional space can be realized by simple operation of low dimensional space. It does not need to consider the form of nonlinear mapping.

$$\Re(x, y) =< \zeta(x), \zeta(y) > . \tag{9}$$

The Gaussian kernel is defined as:

$$\Re(x_i, y_i) = exp(-\frac{\left\|x_j - v_i\right\|^2}{\sigma^2}) \tag{10}$$

2.2. Determining membership and cluster centers

We first define the object function of FCM.

$$J(U, V) = 2\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m (1 - \Re(x_j, v_i)) \tag{11}$$

Where $\sum_{i=1}^{c} u_{ij} = 1$

Let the partial derivative of u_{ij} for $J(U, V)$ be zero.

$$\frac{\partial J}{\partial u_{ij}} = 0 => (1 - \Re(x_j, v_i)) - \lambda_i = 0 \tag{12}$$

So

$$u_{ij} = \frac{(\lambda_i)^{1/(m-1)}}{\left[1 - \Re(x_j, x_i)\right]^{1/(m-1)}} \tag{13}$$

Let the partial derivative of cluster center v_i for $J(U, V)$ be zero.

$$\frac{\partial F}{\partial v_i} = 0 <=> 2\sum_{j=1}^{n} u_{ij}^m \Re(x_j, v_i)(-\frac{x_j - v_i}{\sigma^2}) = 0 \tag{14}$$

We can obtain the following equation.

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^m \Re(x_j, v_i) x_j}{\sum_{j=1}^{n} u_{ij}^m \Re(x_j, v_i)} \tag{15}$$

(a) sample 1 (b) sample 2 (c) sample 3

Fig. 1. Fuji apple.

2.3. Radial width selection of Gaussian kernel

σ is the width parameter of function to control radial range of function. For clustering problems, if the clus-tering structure in feature space of sample points is compact, the smaller σ can ensure the effect of the sample points clustering; If the clustering structure is dispersed, the membership function is fuzzy, so the larger can help to obtain the explicit membership function distribution.

Assuming sample set $X = x_j, ..., x_n$, we define the center of sample $\bar{x} = (\sum_{i=1}^{n} x_i)/n$, the distance be-tween sample point and sample center is $d_i = \|x_i - \bar{x}\|$, so the average distance of sample set is $\bar{d} = (\sum_{i=1}^{n} d_i)/n$.

The distance variance from the center of the sample to the sample point is as the radial width.

$$\sigma = \sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1) \qquad (16)$$

The distance variance between the sample points represents the degree of polymerization around the clustering, which is the compaction degree of the sample points cluster structure, and the distance variance of the sample points can be roughly measured.

3. Experiments and Analysis

To demonstrate the efficiency of the proposed image segmentation algorithm, experiment is carried out on re-mote sensing images. The segmentation performance of the proposed method is compared with ACP [18], MSRT [19] and KWFLICM [20]. These algorithms are implemented in Matlab. Fig. 1 is the Fuji apple sample of fruit trees. The computing platform is configured with Intel Core I7 4.0GHz CPU, 16G Memory and NVIDIA GTX 780 GPU. The algorithm is compiled by MATLAB 2017(a). The image size in experiment is 512×512pixel.

Using the algorithms mentioned above, we also obtain the compared results as shown in Figs. 2-4.

The results showed that our method could robustly get the segmentation results from many large complex scenes. To analyze the effect of proposed segmentation method, we made a comparison with the existed ap-proaches. Our method can segment the fruit images correctly. The segmentation accuracy of the proposed method is 90.7% and the average computing time is about 15 seconds. This comparison showed that our pro-posed method can get a better result.

Supposing TP, FP, and FN denote the number of correctly recognized apples, the number of false recognized apples, and the number of unrecognized apples respectively in the test images. The recognition rate and error rate can be formulated as: We also adopt IoU (Intersection over Union) to evaluate the recognition effect.

$$reg_{rate} = \frac{TP}{TP+FP}.e_{rate} = \frac{FP}{TP+FN} \qquad (17)$$

Tables 1 is the comparison results with different methods. From this table, we can see that the proposed method has better result than other methods.

Table 1. Recognition rate of different methods.

Method	e_{rate}	reg_{rate}	IoU
ACP	0.46	84.2%	46.7%
MSRT	0.41	89.8%	58.3%
KWFLICM	0.38	90.8%	67.2%
Proposed method	0.22	96.5%	76.5%

We also conduct the time comparison with the above methods. Tables 2 shows the time consumption. It can be seen that proposed method obtains the better segmentation performances. This validates our visual inspection.

4. Conclusion

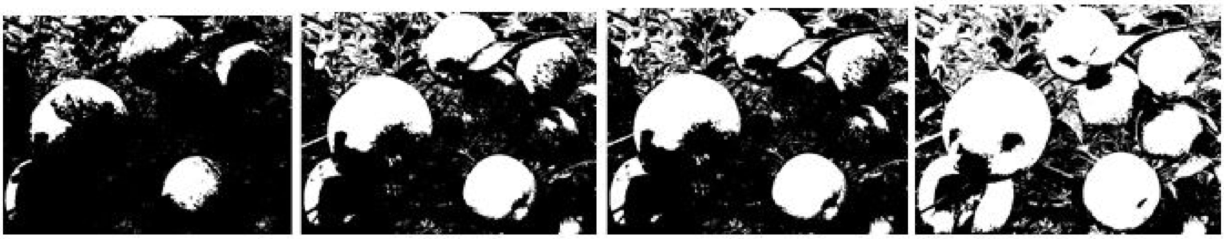
This paper proposes a new segmentation method using an modified Gaussian kernel function and CNN for apple

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

Fig. 2. Comparison of the sample 1 segmentation results for the proposed method, ACP, MSRT, and KWFLICM method. From left to right. (a) MSRT method. (b) ACP method. (c) KWFLICM method. (d) Proposed method.

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

Fig. 3. Comparison of the sample 2 segmentation results for the proposed method, ACP, MSRT, and KWFLICM method. From left to right. (a) MSRT method. (b) ACP method. (c) KWFLICM method. (d) Proposed method.



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

Fig. 4. Comparison of the sample 3 segmentation results for the proposed method, ACP, MSRT, and KWFLICM method. From left to right. (a) MSRT method. (b) ACP method. (c) KWFLICM method. (d) Proposed method.

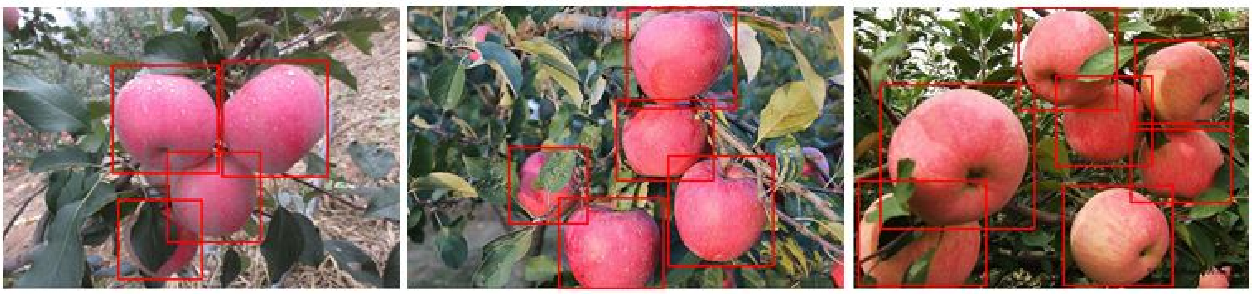


Fig. 5. Fuji apple sample recognition result with proposed method.

Table 2. Time complexity of different methods.

Method	Sample1	Sample2	Sample3
ACP	0.38s	0.41s	0.55s
MSRT	0.34s	0.32s	0.50s
KWFLICM	0.32s	0.29s	0.48s
Proposed method	0.26s	0.14s	0.35s

recognition. By the improved Gaussian kernel function, region can be segmented from large complex area in seconds. Experiments show that the proposed method can obtain a better results than other methods. This method can identify the small and medium sized apple images accurately and has good real-time performance. In the future, I will use the deep learning methods to further perfect our method and apply them into practical engineering.

Acknowledgement

This work is supported by the Natural Science Fund Guiding Program in Liaoning Province (Grant No.20180520024)

References

[1] Shoulin Yin, Jie Liu, and Hang Li. A Self-Supervised Learning Method for Shadow Detection in Remote Sensing Imagery. *3D Research*, 9(4), dec 2018.

[2] Lin Teng, Hang Li, and Shoulin Yin. Modified pyramid dual tree direction filter-based image denoising via curvature scale and nonlocal mean multigrade remnant filter. *International Journal of Communication Systems*, 31(16), nov 2018.

[3] Lin Teng and Hang Li. CSDK: A Chi-square distribution-Kernel method for image de-noising under the Internet of things big data environment. *International Journal of Distributed Sensor Networks*, 15(5), may 2019.

[4] Ping Wang, Xiaokai Dong, Zhaoqing Qiang, and Huizhen Jia. The auto-obtaining of the knowledge about developing hailstone weather. In *Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, volume 4, pages 394–398, 2007.

[5] Zhigen Fei and Junjie Guo. A new hybrid image segmentation method for fingerprint identification. In *Proceedings - 2011 IEEE International Conference on Computer Science and Automation Engineering, CSAE 2011*, volume 3, pages 382–386, 2011.

[6] Lin Teng, Hang Li, Shoulin Yin, and Yang Sun. Improved krill group-based region growing algorithm for image segmentation, oct 2019.

[7] Lin Teng, Hang Li, Shoulin Yin, Shahid Karim, and Yang Sun. An active contour model based on hybrid energy and fisher criterion for image segmentation, jan 2020.

[8] Shoulin Yin and Ye Zhang. Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images, apr 2019.

[9] Hulin Kuang, Cairong Liu, Leanne Lai Hang Chan, and Hong Yan. Multi-class fruit detection based on image region selection and improved object proposals. *Neurocomputing*, 283:241–255, mar 2018.

[10] Jeng Shyang Pan, Jun Bao Li, and Zhe Ming Lu. Adaptive quasiconformal kernel discriminant analysis. *Neurocomputing*, 71(13-15):2754–2760, 2008.

[11] Hang Li, Lin Teng, and Shoulin Yin. A new bidirectional research chord method based on bacterial foraging algorithm. *Journal of Computers (Taiwan)*, 29(3):210–219, jun 2018.

[12] Shou Lin Yin, Jie Liu, and Lin Teng. A new krill herd algorithm based on SVM method for road feature extraction. *Journal of Information Hiding and Multimedia Signal Processing*, 9(4):997–1005, 2018.

[13] Shoulin Yin, Ye Zhang, and Shahid Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model. *IEEE Access*, 6:26069–26080, may 2018.

[14] R. Moreno, M. Graña, D. M. Ramik, and K. Madani. Image segmentation on spherical coordinate representation of RGB colour space. *IET Image Processing*, 6(9):1275–1283, 2012.

[15] Kaihua Zhang, Lei Zhang, Kin Man Lam, and David Zhang. A Level Set Approach to Image Segmentation with Intensity Inhomogeneity. *IEEE Transactions on Cybernetics*, 46(2):546–557, feb 2016.

[16] Peng Li, Zhikui Chen, Laurence T. Yang, Liang Zhao, and Qingchen Zhang. A privacy-preserving high-order neuro-fuzzy c-means algorithm with cloud computing. *Neurocomputing*, 256:82–89, sep 2017.

[17] Fazel Keshtkar, Wail Gueaieb, and Mehedi Masud. Knowledge-based image segmentation using swarm intelligence techniques. *International Journal of Innovative Computing and Applications*, 4(2):75–91, may 2012.

[18] Meng Jian and Cheolkon Jung. Interactive Image Segmentation Using Adaptive Constraint Propagation. *IEEE Transactions on Image Processing*, 25(3):1301–1311, mar 2016.

[19] R. S. Medeiros, J. Scharcanski, and A. Wong. Image segmentation via multi-scale stochastic regional texture appearance models. *Computer Vision and Image Understanding*, 142:23–36, jan 2016.

[20] Kashif Hussain Memon, Sufyan Memon, Muhammad Ali Qureshi, Muhammad Bux Alvi, Dileep Kumar, and Rehan Ali Shah. Kernel Possibilistic Fuzzy c-Means Clustering with Local Information for Image Segmentation. *International Journal of Fuzzy Systems*, 21(1):321–332, feb 2019.