Research Article
Method of Fruit Image Segmentation by Improved K-Means

Fangzheng Wang

Yancheng Institute of Health Sciences, Yancheng 224005, China

Abstract: The clustering algorithm of K-means is a widely used clustering algorithm, which characteristic is efficient and simple to implement. In this study, it takes the clustering algorithm of K-means as the starting point, which also explains the improvement of the clustering algorithm of K-means clustering, discussing the application of K-means on the realization of fruit image segmentation.

Keywords: Clustering algorithm, fruit image segmentation, K-means

INTRODUCTION

In recent years, with the further development and advance of computer technology, more and more color images are used, therefore, the research on color image segmentation has become a major researching focus, which aims at the specific requirements from users according to the specific issues, scholars from all over the world have put forward many classic ways about segmentation algorithm, so far, the segmentation of color image histogram can be divided into the following aspects: The threshold method, feature space clustering method, region based method, edge detection method, fuzzy method and neural network method and so on. In previous studies, the research that is based on clustering algorithm of K-means and the improved image segmentation technology has been widely got concern (Zhou *et al.*, 2013). The clustering algorithm of K-means is an unsupervised clustering method in dynamic algorithm, which has a certain adaptability, but the result of clustering can be easily influenced by the initial center of clustering.

MATERIALS AND METHODS

The basic theory of clustering algorithm of K-Means: The clustering algorithm of K-means is a kind of technology based on the average value. Taking K as the parameter, then dividing the sample point set $s = \{x_1, x_2,......, x_n\}$ into classes $\{\{c_1\}, \{c_2\}, ..., \{c_k\}\}$, at the same time, assuming that the center of clustering is Z_p, the clustering error E can be defined as follows:

$$E = \sum_{i=1}^{K} \sum_{X_j \in C_i} \left| X_j - Z_p \right|^2 \text{, among them}$$

$$j = 1, 2,......, n, p = 1, 2,......, k \qquad (1)$$

The clustering algorithm of K-means can try to reduce the value of E by using iterative method, so that the cluster itself can be as compact as possible, which can be as far apart as possible from the other clustering. The steps of algorithm are as follows:

- From the sample point $s = \{x_1, x_2,......, x_n\}$, there are K initial cluster centers that can be randomly selected.
- Putting the sample point set $s = \{x_1, x_2, ..., x_n\}$ into clustering according to the center of Z_1, $Z_2,......, Z_p$, clustering, K classes of clustering including $\{\{c_1\}, \{c_2\},...... \{c_k\}\}$ can be acquired, the method to determine ci can be as the following: for any $xj \in s$, if $(x_j - z_i)$ $2 \leq (Xj-zp)$ 2, $p \neq i$, $p = 1, 2, ..., k$, then $x_j \in c_i$.
- Adjusting the center of clustering, the new class center can be acquired, namely, $Z_i^* = \dfrac{1}{n} \sum_{X_j} X_j$.

  Among them, n_i is the number of samples that c_i contained.
- According to formula (1) calculating error function E, until the value of E can not be obviously changed any longer or there is no change in the member of clustering.

The algorithm is as shown in Fig. 1.

RESULTS AND DISCUSSION

The improvement of clustering algorithm of K-Means: There is an important step in clustering algorithm of K-means, which is the determination of K value and the selection of the initial clustering center. Since the traditional clustering algorithm of K-means is generally to select the number of the categories of the clustering as well as its center randomly (Chen *et al.*, 2013). There is a strong correlation between the
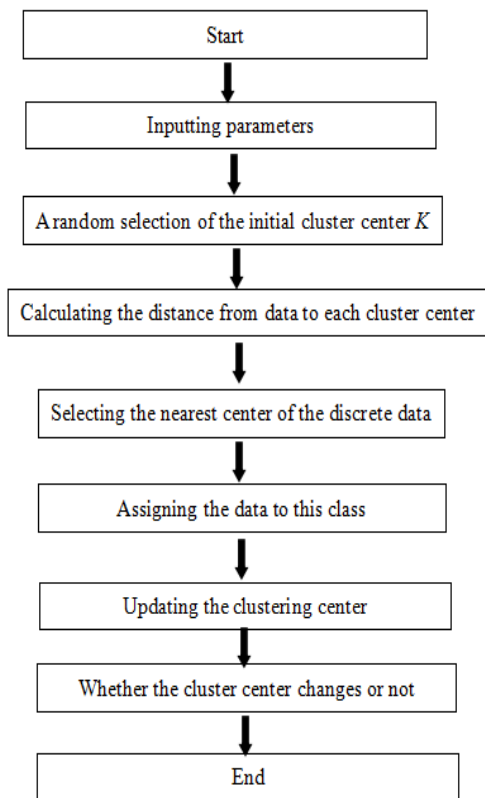
Fig. 1: The flow chat of classical K-means algorithm

clustering result and the location of the sample, moreover, the performance of the cluster is close related to the selection of the initial clustering center. If the selection of *K* sample is not reasonable, it will increase the complexity of operation, which will mislead the process of clustering, thus the clustering results can not be acquired reasonably.

In this study, using the algorithm of K-means to have initial segmentation for the image of fruit, only according to the pixel image color information of the similar color of the fruit can it be divided into the same class, which is not to have an accurate segmentation for the image of fruit (Song *et al*., 2013). After finishing the initial segmentation, it needs to carry out regional merger according to the color and spatial information of the image of fruit, so as to get the final segmentation result. On choosing the number of categories, the choice of a fixed number of cluster can meet the requirements of the initial segmentation, which also can have high operation efficiency.

This study firstly puts the image of fruit into blocks, whose size is 16×16, since the size of the sub-block can basically guarantee each block have smaller

proportion in the image of fruit, so the change in color of each piece of block should not be too large, which is convenient for the clustering analysis. Once it is specified each small block as 3, then the value of *K* is 3, namely, the value of characteristics of each small block value is 3. Besides, the selection of clustering centers is another important factor that can affect the results of clustering. In this study, because the image of fruit is divided into many small pieces, in the sub-block, there is little color change among them, so we can choose the average pixel value as the initial clustering centers that the analysis on the image of fruit can be carried on. If the current representative object is replaced by the non-representative objective, then, the cost function can calculate the absolute error value. While the total cost of exchanging is the total sum of the costs that are caused by all non-representative objects (Cai *et al*., 2008). If the total cost is negative, then *E* will be decreased, cluster center can be replaced by any random cluster point; if the total cost is positive, then the current cluster center is acceptable, there is no change in this iteration.

The improved clustering algorithm of K-Means: According to the research results made by Ohta and some other staff, in this study, it takes component II as the one-dimensional characteristic quantity of color image's pixel, using component II to replace the gray value of the clustering of K-means in image segmentation. Indicating the number of pixels with T, then $I^P_1 \in \{0, 1, \ldots, Lmax-1\}$ ($p=1, 2, \ldots, T$), I^P_1 is the gray value of the pixel in set Ai (j). Lmax is the brightness grade of II, λ_i is the average value of No. i class after No. j iterations, $A_i(j)$ can represent the set of pixel of No. *i* class after No. *j* iterations. As shown in Table 1.

By using algorithm of K-means, the initial image of fruit can be segmented, due to the reasons that many small blocks are divided in advance, many divided small areas are submerged by the target area, which is the result of the over segmentation. There are still many similarities between each region, those regions without great obvious differences should be merged into a larger region (Wang *et al*., 2012). After the process of carrying on the following steps with the initial clustering segmentation, the final segmentation results can be obtained. When regions carry on merging, the size of the minimum region is critical for segmentation. The measurement of regional distance is an important standard for regional merging, the method of distance measurement can directly determine the final

Table 1: The number of clusters and the processing time when different image segmentation algorithms are selected

The pixel of the image	The algorithm of K –means		Fast fuzzy algorithm of K –means		The improved algorithm of K-means	
	k	t/s	k	t/s	k	t/s
300×300	5	4.01	5	3.23	5	2.37
400×286	3	3.68	3	2.72	3	1.54
800×600	5	24.17	5	15.25	5	10.80

Table 2: Comparison of the speed

The number of samples (block)	The number of clusters	K-means	The algorithm in this study
256×256	3	0.726524s	0.53083s
300×300	3	1.011003s	0.521945s

segmentation result and image region merging results, which can become the necessary condition for the merger of the two regions: Since the two regions are similar in color and close to each other in space. Therefore, in this study, it mainly uses the adjacent color region to carry on regional merging. Through the experiment, we find out when threshold value T is 0.16% of the pixel number of the segmentation image of fruit, it can get better segmentation effect. As shown in Table 2.

CONCLUSION

The method of clustering of K-means is an unsupervised dynamic algorithm. The result of clustering of K-means can be influenced by the number of cluster center and the initial cluster center, which also can be affected by the geometry statues of samples at the same time. In view of the above problems, in this study, it proposes an improved clustering algorithm of K-means, which is based on rough set theory to determine the initial class number and class center of the clustering of K-means. By using the researching results of made by Ohta and some other staff, selecting the first component as the one-dimensional characteristic quantity of color image's pixel that can effectively represent the characteristic of color in the set to replace the gray value of the image segmentation of classical clustering of K-means, it can greatly reduce the amount of computing. At the same time, the distance measurement can adopt the characteristic distance as the method for measuring the clustering of K-means, which can improve the applicability and accuracy of the algorithm.

REFERENCES

Cai, J., X. Zhou and Y. Li, 2008. Recognition of mature oranges in natural scene based on machine vision. T. Chinese Soc. Agric. Eng., 24: 175-178.

Chen, K., X. Zou, J. Xiong *et al.*, 2013. Improved fruit fuzzy clustering image segmentation algorithm based on visual saliency. T. Chinese Soc. Agric. Eng., 29: 157-165.

Song, H., C. Zhang, J. Pan, X. Yin and Y. Zhuang, 2013. Segmentation and reconstruction of overlapped apple images based on convex hull. T. Chinese Soc. Agric. Eng., 29: 163-168.

Wang, X., X. Han and H. Mao, 2012. Vision-based detection of tomato main stem in greenhouse with red rope. T. Chinese Soc. Agric. Eng., 28: 135-141.

Zhou, W., J. Feng, G. Liu and X. Ma, 2013. Application of image registration technology in apple harvest robot. T. Chinese Soc. Agric. Eng., 29: 20-26.