Clustering Techniques based Crops Image Segmentation

K.Muthukannan
Research Scholar, Anna University of
Technology, Tirunelveli, Tamilnadu, India

P.Latha
Associate Professor, Government College of
Engineering, Tirunelveli, Tamilnadu, India

ABSTRACT

Segmentation of an image entails the division or separation of the image into regions of similar attribute. The most basic attribute for segmentation of an image is its luminance amplitude for a monochrome image and color components for a color image. The main objective of this paper is to segment the natural crops images by using clustering techniques which is produced very good results. The clustering based image segmentation in the field of agriculture imaging(crops image segmentation) based on its color, size, shape, contrast and etc. and this algorithm is going to be produced more advantages such as less execution time, more accuracy.

Keywords

Edge detection, K-Means clustering, Optimal Fuzzy C-Means clustering, Segmentation.

1. INTRODUCTION

The image segmentation is a key process of the image analysis and the image comprehension. Because of the influence of the complicated background, the object characteristics diversity and the noise, the image segmentation is the difficult and hot research issues on the image processing. The process of partitioning a digital image into multiple regions (sets of pixels) is called image segmentation. Actually, partitions are different objects in image which have the same texture or color.

The result of image segmentation is a set of regions that collectively cover the entire image, or a set of contours extracted from the image. All of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristics. Some of practical applications of image segmentation are: image processing, computer vision, face recognition, medical imaging, digital libraries, image and video retrieval, etc [8]. Image segmentation methods fall into five categories: Pixel based segmentation [7], Region based segmentation [6], Edge based segmentation, Edge and region Hybrid segmentation and Clustering based segmentation [1]. In this study an image

segmentation of crops images has been proposed, this implementation result is based on clustering procedure.

The K-Means clustering technique is a well-known approach that has been applied to solve low-level image segmentation tasks. This clustering algorithm is convergent and its aim is to optimize the partitioning decisions based on a user-defined initial set of clusters that is updated after each iteration. This procedure is computationally efficient and can be applied to multidimensional data but in general the results are meaningful only if homogenous non-textured color regions define the image data.

Fuzzy clustering techniques have been effectively used in image processing, pattern recognition and fuzzy modeling. The OFCM was proposed by Gath and Geva and Xuejian Xiong,

Kap Luk Chan. An effective method for unsupervised optimal fuzzy clustering based on a generalized objective function is discussed in this paper [2].In conclusion, the aim of this work is three fold: 1) Edge detection using prewit and canny edge detection method 2) Segmentation based on k-means clustering and optimal fuzzy c-means clustering procedure 3) Calculating the performance by manual evaluation method.

The rest of this paper is organized as follows: In Section 2, give the details of the proposed approach concept. In Section 3, give the details of the image segmentation using K-Means clustering and the K-Means clustering algorithm details are discussed. In Section 4, give the details of the Fuzzy clustering and also give the details of the color image segmentation using Optimal Fuzzy C-Means clustering. The experimental results are shown in Section 5. Finally, some conclusions and future research of the experimental results were in Section 6.

2. THE PROPOSED APPROACH CONCEPT –DETAILS

The overall concept that is the framework for any vision related algorithm of image segmentation is almost the same. First, the digital images are acquired from the environment using a digital camera. Then image-processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. Figure 1 depicts the basic procedure of the proposed vision-based image segmentation of natural images (crops) algorithm in this research [1].

The proposed approach step - by - step of the image segmentation processes is illustrated in Algorithm 2.1. In the initial step, the RGB images of all the leaf samples were picked up. Some real samples of those diseases are shown in Figure 2.
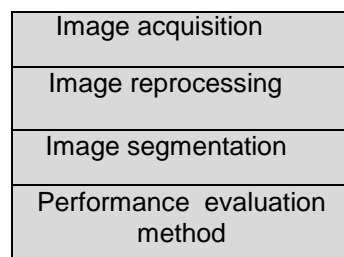

Fig 1: The basic procedure of the proposed clustering based crop image segmentation

It is obvious from Figure 2 that leaves belonging to pepper, tomato, potato and rice images have significant differences form greasy spot leaves in terms of color and *texture*. However, the leaves had very small differences as discernible to the human eye, which may justify the misclassifications based on naked eye.

2.1. Algorithm

The basic steps describe the proposed algorithm.

1. RGB image acquisition
2. Create the color transformation structure
3. Convert the color values in RGB to the space specified in the color transformation structure
4. Apply K-means clustering
5. Apply OFCM clustering
6. Performance evaluation method

In details, in step 2 a color transformation structure for the RGB leaf image is created, and then, a device-independent color space transformation for the color transformation structure is applied in step 3. Steps 2 and 3 are inevitable for carrying out step 4. In this step the images at hand are segmented using the K-Means clustering technique [9; 4]. The Proposed approach for crops image segmentation based on clustering and edge detection method. The following sections discuss the clustering techniques and show the implementation results using popular edge detection methods.

3. IMAGE SEGMENTATION USING K-MEANS CLUSTERING

The K-Means clustering technique is a well-known approach that has been applied to solve low-level image segmentation tasks. This clustering algorithm is convergent and its aim is to optimize the partitioning decisions based on a user-defined initial set of clusters. The applications of the clustering algorithms to the segmentation of complex color-textured images are restricted by two problems. The first problem is generated by the starting condition (the initialization of the initial cluster centers), while the second is generated by the fact that no spatial (regional) cohesion is applied during the space partitioning process. The k-means clustering was proposed by Bo Zhao, ZhongxiangZhu, Enrong Mao and Zhenghe Song [16].

There are two preprocessing steps that are needed in order to implement the K-means clustering algorithm: The phase starts first by creating device-independent color space transformation structure. In a device independent color space, the coordinates used to specify the color will produce the same color regardless of the device used to draw it. Thus, we created the color transformation structure that defines the color space conversion. Then, we applied the device-independent color space transformation, which converts the color values in the image to the color space specified in the color transformation structure. The color transformation structure specifies various parameters of the transformation. A *device dependent color space* is the one where the resultant color depends on the equipment used to produce it. For example the color produced using pixel with a given RGB values will be altered as the brightness and contrast on the display device used. Thus the RGB system is a color space that is dependent.

3.1. The Algorithm

K-means is one of the most popular clustering algorithms. It is simple and fairly fast. K-means is initialized from some random or approximate solution. Each iteration assigns each point to its nearest cluster and then points belonging to the same cluster are averaged to get new cluster centroids. Each iteration successively improves cluster centroids until they become stable. Formally, the problem of clustering is defined as finding a partition of D into k subsets such that

$$\sum_{i=1}^{n} \zeta (t_i; C_j) \tag{1}$$

is minimized, where Cj is the nearest cluster centroid of ti. The quality of a clustering model is measured by the sum of squared distances from each point to the cluster where it was assigned [26; 21; 6]. This quantity is proportional to the average quantization error, also known as distortion [19; 24]. The quality of a solution is measured as:

$$q© = 1/n \sum_{i=1}^{n} \zeta (t_i; C_j) \tag{2}$$

This can be computed from R as,

$$q(R, W) = \sum_{j=1}^{k} W_j \sum_{i=1}^{d} R_{ij} \tag{3}$$

In general, spatial partitioning methods are implemented using iterative frameworks that either attempt to minimize the variation within the clusters or attempt to identify the optimal partitions based on a set of Gaussian Mixture Models. In this paper focus the implementation of the K-Means algorithm, although the methodology detailed in this paper can be applied to other clustering schemes such as fuzzy clustering [3] or competitive agglomerative clustering [2]. The K-Means is a nonhierarchical clustering technique that follows a simple procedure to classify a given data set through a certain number of K clusters that are known a priori. The K-Means algorithm updates the space partition of the input data iteratively, where the elements of the data are exchanged between clusters based on a predefined metric (typically the Euclidian distance between the cluster centers and the vector under analysis) in order to satisfy the criteria of minimizing the variation within each cluster and maximizing the variation between the resulting K clusters. This clustering algorithm, in its standard formulation consists mainly of four steps that are briefly described below:

3.2. Steps of the classical K-Means clustering algorithm

K-means is one of the most popular clustering algorithms. It is simple and fairly fast. K-means is initialized from some random or approximate solution. This clustering algorithm, in its standard formulation consists mainly of four steps that are briefly described below:

1. Initialization – generate the starting condition by defining the number of clusters and randomly select the initial cluster centers.
2. Generate a new partition by assigning each data point to the nearest cluster center.
3. Recalculate the centers for clusters receiving new data points and for clusters losing data points.
4. Repeat the steps 2 and 3 until a distance convergence criterion is met.

The *K*-means clustering is a partitioning method for grouping objects so that the within-group variance is minimized. By minimizing dissimilarity of each subset locally, the algorithm will globally yield an optimal dissimilarity of all subsets [15].

The algorithm, as applied to image threshold, is given by the following steps:

1). Initialize the (*K*) class centers. For simplicity, un equal-distance method is used to define the initial class centers:

$$Center_i^0 = GL_{min} + [(i-i/2) (GL_{max} - GL_{min})/k)]$$
$$i = 1, 2....k \tag{4}$$

Where $Center_i^0$ is the initial class center for the ith class, GL_{max} and GL_{min} are the maximum and minimum of the gray value GL in the sample space, respectively.

2) Assign each point to its closest class center. The criterion to assign a point to a class is based on the Euclidean distance in the feature (*GL*) space using:

$$Distance_{i,j} = abs\ (GL_j - Center_i)$$
$$i = 1,2,\ldots, K; j = 1,2,\ldots, N. \qquad (5)$$

Where $Distance_{i,j}$ is the distance from the jth point to the ith class, and N is the total number of points in the sample space.

3). Calculate the (*K*) new class centers from the mean of the points that are assigned to it. The new class centers are calculated by

$$Center_i{}^m = 1/N_i \quad \sum_{i=1}^{N_i} GL_j$$
$$j = 1, 2, \ldots. K. \qquad (6)$$

Where N_i is the total number of points that are assigned to the ith class in step 2.

4) Repeat step 2 if any class centers change, otherwise end the circulation.

5) The threshold value is defined as the average of the Kth class center and the (*K*-1) th class center:

$$Threshold = 1/2(center_k + center_{k-1}). \qquad (7)$$

4. FUZZY CLUSTERING

Clustering involves the task of dividing data points into homogeneous classes or clusters so that items in the same class are as similar as possible and items in different classes are as dissimilar as possible. Clustering can also be thought of as a form of data compression, where a large number of samples are converted into a small number of representative prototypes or clusters. Depending on the data and the application, different types of similarity measures may be used to identify classes, where the similarity measure controls how the clusters are formed. Some examples of values that can be used as similarity measure include distance, connectivity, and intensity. Areas of application of fuzzy cluster analysis include data analysis, pattern recognition, and image segmentation [15].

4.1. The unsupervised optimal fuzzy c-means clustering

Simple and well-known Fuzzy clustering algorithms, which are also widely used, is the fuzzy c-means method. In fact, there are two main shortcomings in the FCM algorithm. First, the numbers of resulting clusters need to be specified in advance, which in practice, such as in the unsupervised classification, this is out of the question. Secondly, the FCM is very limited according to the restriction to spherical clusters. Hence, hyper ellipsoidal clusters can lead to wrongful clustering. An unsupervised optimal fuzzy clustering (UOFC) algorithm [13], which can also be regarded as an improvement of the FCM, is proposed here to overcome these disadvantages [14].

4.2. The optimal fuzzy clustering

Let us consider a collection of n patterns constituting vectors in the p-dimensional space of real numbers, namely x1,x2,……xn∈Rp,forming the input data set X.then the new modified generalized objective function proposed based on [9] is given as follows:

$$J(U, V; X) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m \{(1-g) * (\| x_k - V_i \|_1)^l$$
$$+ g *(\sum_{j=1}^{r} \{(xk - V_i) \bullet s_{ij} \}^2)\}, l \geq 1 \qquad (8)$$

Where, c is the number of the clusters, and other notations are described as follows .V_i, (i=1, 2… c) is the prototype of the ith cluster. If one pattern x_k belongs to the ith cluster, that means the distance between x_k and V_i is smallest. The exponent parameter m is used to control the influence of intermediate membership values on the objective function and $1 < m < \infty$.U = {μik} is the fuzzy membership matrix. Where, μ_{ik} denotes the grade of membership of the kth pattern in the ith cluster and it should satisfy the following two conditions:

$$\bullet \sum_{K=1}^{n} \mu_{ik} = 1 \text{ for all k.}$$

$$\bullet\ n > \sum_{K=1}^{n} \mu_{ik} > 0 \text{ for all i.}$$

g ∈ [0,1] is a weighted value whose role is to keep the balance between two basic components in the above equation. Depending on the data set, a change of g could affect the resulting shapes of the obtained clusters.
The symbol (•) means the inner product.
It can be seen that Eq. (5) there are two components:

The first one characteristic the distance between the prototype V_i and the kth pattern x_k, which presents the dissimilarity between V_i and x_i. Unlike the simple Euclidean distance measure used in the FCM algorithm, UOFC algorithm adopted a more general l-norm distance measure. Apparently, if l=2, the distance measure is the well-known Euclidean distance. The advantage is that it needs not to be restricted in to the spherical clusters. Actually the UOFC algorithm can be applied in to arbitrary-shaped clusters.

The second term represents a linear variation which goes through the prototype Vi and is spanned by the collection of r linearly independent vectors $s_{i1}, s_{i2},\ldots,s_{ir}$. These r vectors are the eigenvectors of the generalized within cluster scatter matrix E_i, corresponding to its first r largest eigen values which give the cohesiveness of the cluster.

$$E_i = \sum_{K=1}^{N} (\mu_{ik})\ m\ (x_k - V_i)\ (x_k - V_i)^T \qquad (9)$$

These r eigenvectors, seen as the r principle eigenvectors determining the whole cluster approximately, give the most important directions, along which most of the patterns x_k,(k =1, 2,….n) in the i-th cluster scatter. By introduction this special term, the principle directions of the cluster are emphasized .As a result, the speed of searching the prototype of the cluster is improved. Especially for a large number of input patterns, the value of r can be increased to significantly elevate the convergence speed of this clustering algorithm. Differentiating the objective function J with respect to each Vi and μ_{ik} ,we can obtain eqs.(10),(11) used for updating the membership degrees and the prototypes in an iterative procedure until the difference between the new membership matrix and the old one in the previous iteration step is less than a given tolerance bound.

$$V_i{}^t = \sum_{k=1}^{n} (\mu_{ik}{}^{t-1})^m\ x_k / \sum_{k=1}^{n} (\mu_{ik}{}^{t-1})^m \qquad (10)$$

$$\mu_{ik}{}^t = 1 / \{\sum_{j=1}^{r} \delta_{ik}{}^t / \delta_{jk}{}^t \}^{(1/(m-1))} \qquad (11)$$

Where t is the iterative step number, and

$$\delta_{ik}{}^t = (1-g) * (\| x_k - V_i{}^t \|_1)^l + r$$

$$g*(\sum\{(x_k - V_i^t)\bullet s_{ij}^{\ t}\}^2 \qquad (12)$$
$$J=1$$

Obviously, these two values are the necessary conditions for J to have a local minimum. However, minimization of J with Eq. (5) forms a class of constrained nonlinear optimization problems whose solution is unknown. This problem arises by placing the l-norm in the objective function. For the present research, only l=2 norm is considered for testing the UOFC algorithm while other norm measures will be studied in future work.

It can be seen that the two main advantages of the UOFC algorithm are
(1) Its additional linear in the generalized objective function.
(2) The l-norm distance measure used.
By minimizing the objective function J, we can quickly group great number of input patterns along with their r largest scattering direction. The computational precision can then be improved while the time and memory requirement can be greatly decreased [15].

4.3.The algorithm

Fuzzy partition is carried out through an iterative optimization:
(1) Choose initial cluster centroids (seeds) Vi.
(2) Compute the degree of membership of all feature vectors in all the clusters:

$$u\_\{ij\} = \frac{(1/d^2(Xj, Vi))^{\wedge}(1/(q-1))}{\text{Sum }(k=1, K)\ (1/d^2(Xj, Vk))^{\wedge}(1/(q-1))}$$

(3) Compute new centroids Vi_new:

$$Vi\_new = \frac{\text{Sum }(j=1, N)\ (u\_\{ij\})^{\wedge}q\ Xj}{\text{Sum }(j=1, N)\ (u\_\{ij\})^{\wedge}q}$$

(4) When the movement of centroids (relative changes) is less than a predetermined threshold MOVETHRS, stop the iteration. Otherwise go to step 2.
(5) Finally, a data point Xj is assigned to cluster i if the fuzzy membership u_{ij} >= u_{kj} for all k clusters.

5. EXPERIMENTAL RESULTS

The crops image is used for this experiment because the image has different regions. The result of the experiments is used to find the accuracy values. The results that got by using prewit edge detection method is shown in the Fig. 3,canny edge detection is shown in Fig,4, and the original crops images are shown in Fig.2.
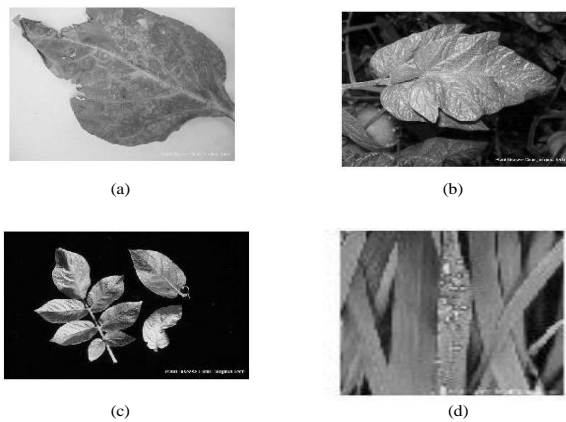

Fig 2: The original crops images,(a).pepper image (b).tomato image(c).potato image (d).rice image

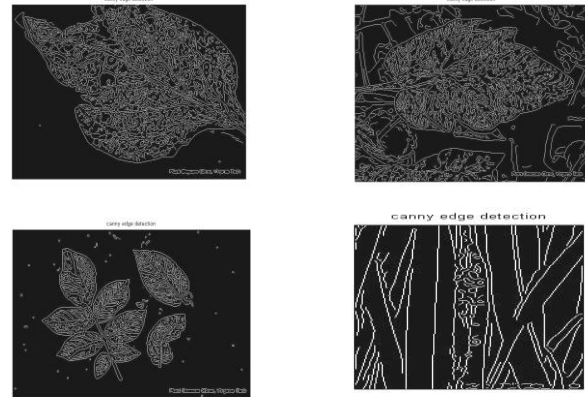**Fig.3: The edge detection results using prewit operators**

Fig.4: The edge detection results using canny operators

This is the preprocessing steps for image segmentation. Here, the popular edge detection techniques such as prewit and canny edge detections are used and results are shown in figure 3 & figure 4.The above results are shows the very good edge detection in the crops images and it also provides the very good edge detection in the disease area.

5.1. Performance analysis method

The two main considerations in defining the accuracy measure are (1).workable in cases where not all types of objects are present in each image; (2).able to count the correct and false results separately for each type of object. Suppose the image contains N types of object, then the accuracy measure is computed by

$$Accuracy = \sum_{i=1}^{N} \frac{Correct\ segmented\ Pixels\ in\ i^{th}\ object}{Total\ number\ of\ Pixels\ in\ i^{th}\ object} \qquad (13)$$

Five types of regions are distinct: (1) correct segmented; (2) over segmented; (3) under segmented; (4) missed and (5) noise. This is the simple method for analyzing the performance of the clustering techniques.
The previous theoretical studies show that the accuracy value of the OFCM clustering is around 85%.It is a very good performance compared with other segmentation techniques.

6. CONCLUSION

In this paper, a comparative study of two clustering techniques was discussed. The K-Means clustering and Optimal Fuzzy C-Means clustering techniques were chosen for evaluation. The most popular edge detection techniques such as prewit and canny edge detection is used for the proposed approach and also implemented on four different crops images like pepper, tomato, potato and rice images. The experimental results

indicate that the proposed approach is a valuable approach, which can significantly support an accurate image segmentation of crops images in a little computational effort. By the detailed theoretical study, the performances of a traditional OFCM could be improved. Perhaps, the performances of most of the algorithm primarily depend on the chosen cluster centers. By properly fixing the cluster centers through domain knowledge, or some other mean, the convergence time along with the accuracy could be improved. So here the point is to find a better mechanism to fix the cluster centers by prior knowledge.

An extension of this work will focus on developing hybrid algorithms such as fuzzy based clustering algorithms in order to decrease the error rate of the final segmentation process underscoring the advantages of hybrid algorithms; also, we will dedicate our future works on detect the disease from the crops images.

7. REFERENCES

[1] Al-Bashish, D., M.Braik and S. Bani-Ahmad, 2011.Detection and classification of leaf diseases using K- means-based segmentation and neural-networks-based classification. Inform. Technol. *J.*, 10: 267-275.

[2] Ahmed M N, Yamany S M, Mohamed N, etal.A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRIdata [J].IEEE Trans on Medical Imaging, 2002, 21(3):193-199

. [3] Aldrich, B.; Desai, M. (1994) "Application of spatial grey level dependence methods to digitized mammograms," Image Analysis and Interpretation, 1994, Proceedings of the IEEE Southwest Symposium on, vol., no., pp.100-105, 21-24 Apr 1994.

[4] Ali, S. A., Sulaiman, N., Mustapha, A. and Mustapha, N., (2009).K-means clustering to improve the accuracy of decision tree response classification.Inform.Technol.

J.8:12561262.DOI:0.3923/itj.2009.

[5] J.Bezdek, "A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithm", IEEE Trans.Pattern. Anal. Mach. Intel, 2:1-8, 1980.

[6] GAO Xin-bo.Fuzzy cluster analysis and its applications [M].Xi'an, China: House of Xidian University, 2004. (in Chinese)

[7] Gath, I. and Geva A. B. (1989)," Unsupervised Optimal fuzzyclustering",IEEE Transaction on Pattern Analysis Machine Intelligence, 11(7): 773-781.

[8] Ford, A. and Roberts, A., (2010) Color Space Conversions,August11,1998.

[9] Hartigan, J. A.; Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C (Applied Statistics)

[10] F.Meyer,"Color image segmentation," In IEEE International Conference on Image Processing and its Applications, pages 303–306, May 1995.Maastricht, The Netherlands.

[11] A.Moghaddamzadeh and N. Bourbakis, "A Fuzzy Region Growing Approach for Segmentation of Color Images," Pattern Recognition, 30(6):867- 881, 1997.

[12] G.A. Ruz, P.A.Estevez, and C.A. Perez, "A Neurofuzzy Color Image Segmentation Method for Wood Surface Defect Detection," Forest Prod. J. 55 (4), 52-58, 2005.

[13] Otsu, N. (1979). "A threshold selection method from gray- level histograms". IEEE Trans. Sys., Man Cyber. 9:

[14] A.A. Younes, I. Truck, and H. Akdaj, "Color Image Profiling Using Fuzzy Sets," Turk J Elec Engin, vol.13, no.3, 2005.

[15] Rafael C.Gonzalez and Richard E.Woods, Digital Image processing (second edition) Pearson Education Asia Limited and Publishing House of Electronics Industry, 2007.

[16] ZhongxiangZhu ,Bo Zhao ,Enrong Mao and Zhenghe Song," Image segmentation based on Ant colony optimization and K-means clustering", Proceedings of the IEEE international conference on Automation and logistics, August 18-21,2007,jinan,china.