Fuzzy C-Means Clustering Using Principal Component Analysis for Image Segmentation

Anju Bala[1], Aman Kumar Sharma[2]
[1]Pursuing a Doctoral degree program in Computer Science, Himachal Pradesh University, Shimla (HP)
[2]Chairman of Computer Science Department, Himachal Pradesh University, Shimla (HP)

Abstract:- Nowadays, Image segmentation is the area in which most of the research is carried out. It is considered as one of the most crucial fields in image analysis. It is used to divide an image into meaningful regions and thus extract the region of interests. These regions are considered as objects. Fuzzy c-means (FCM) clustering is one of the best clustering method used for image segmentation, but have a drawback of unknown cluster number. This paper focuses on this drawback of FCM and to overcome it, the Principal component analysis (PCA) is used. PCA is used for detection of cluster numbers for FCM because of its dimension reduction capability. The cluster number is the important factor on which the clustering result depends. Experimental results show that the proposed method efficiently calculate the cluster number for different test images and gives effective results.

Keywords:- *Clustering, Cluster Number, Fuzzy C-Means, Image Compression, Image Segmentation, Principal Component Analysis.*

I. INTRODUCTION

Image segmentation segments the image into meaningful objects or region of interests. It makes the feature extraction easy and lowers the computational complexity of image by compressing it while maintaining the quality of the image. Thus, it simplifies the image into something more useful and meaningful [1].

Clustering term was first given by Tryon in 1939 [2]. It is a process of classifying an image into regions based on homogeneity or heterogeneity of some feature criterion. The classified regions of clustering process are called clusters. These clusters are regions which have minimum intra-cluster distance and maximum inter-cluster distance [3]. Clustering algorithms are categorized into hard and soft clustering. Hard clustering refers to K-means clustering, where each pixel belongs specifically to a single cluster. It has a drawback that it cannot converge optimally in the general case. Soft clustering refers to Fuzzy c-means clustering, which was proposed by Dunn in 1972 [4]. It is the extended version of k-means in which there is a membership function which states the membership of each pixel to the different clusters. That means in it a pixel can belong to two or more clusters. Although the computation cost of FCM is higher than the k-means, however, it gives optimal results [5].

The FCM uses the membership function and an objective function to perform cluster classification. This classification is based on some features, and an image has lots of features. So, which feature should be considered for classification is an issue? This issue is resolved by replacing all the features with a standard set of features. That will reproduce the inter-correlations of the features [6]. Then according to the features selected the number of clusters is calculated. This paper uses PCA to identify the number of clusters for FCM.

The remaining paper organised as follows: Section 2 explains FCM and PCA, section 3 includes the proposed algorithm, section 4 gives the experimental results, and section 5 concludes the study.

II. FUZZY C-MEANS

Clustering involves partition of image pixels into different clusters such as the pixels in the same clusters are similar features, whereas pixels of different clusters are as dissimilar as possible. The clusters are identified through similarity measures. FCM clustering allows a single pixel to belong to two or more clusters. It is achieved by assigning membership to each pixel corresponding to each cluster based on the distance between the cluster center and the pixel. Minimum the distance between pixel and the center more the membership of pixel towards that specific cluster center. The sum of membership of each pixel should be one [7]. The membership of each pixel is updated in each iteration until there is no further updation. The membership of each pixel is calculated by the membership function, which is as follows:

$$\mu ij = 1/\sum_{k=1}^{c}(dij/dik)^{\left(\frac{2}{m}-1\right)} \quad (1)$$

Where uij is membership of ith pixel to the jth cluster center, dij represents Euclidean distance between the ith pixel and jth cluster center, m is the fuzziness index, and c represents cluster number. Along with the membership function, the cluster center is also updated in each iteration.

$$vj = \frac{(\sum_{i=1}^{n}(uij)^m xi)}{(\sum_{i=1}^{n}(uij)^m)}, \forall j = 1,2, \ldots c \quad (2)$$

Where vj represents the jth cluster center, n is the number of pixels and xi is the ith pixel.

Fuzzy clustering is carried out through an iterative updation of membership function and cluster center and optimization of the objective function.

$$J(U,V) = \sum_{i=1}^{n}\sum_{j=1}^{c}(uij)^{m}\ ||xi - vj||^{2} \qquad (3)$$

'||xi – vj||' is the Euclidean distance between ith data and jth cluster center.

'β' is the termination criterion between [0, 1].

➢ *Algorithm:*

- Initialize U=[uij] matrix.
- At k- step: for given cluster numbers c, calculate cluster centers using equation 2.
- Update U using equation 1.
- If ||U(k+1) - U(k)|| < β, then stop; otherwise go to step 2.
- The lower the value of B, the good the quality of result but at the cost of computational complexity. It gives better results than k-means but has some disadvantages such as cluster number should be known prior [8].
- This paper focuses on the cluster number problem and to solve this; the principal component analysis is used.

III. PRINCIPAL COMPONENT ANALYSIS

The principal component analysis is a process of reducing the broad set of variables or features to small set that still contains most of the information. It transforms a large number of correlated variables into a smaller number of uncorrelated variables. These smaller number of variables are called principal components. The first principal component consists of most of the variability in data and succeeding components consists of remaining. This paper uses the covariance matrix instead of the correlation matrix as the variances of a pixel are not much [9].

IV. PROPOSED METHOD

There are many extended versions of FCM available in the literature [10-12] but, the basic FCM is one of the most useful general-purpose clustering schemes [13]. The basic FCM requires determination of several parameters such as cluster number, fuzzy index, stopping criterion, cluster centers etc. the initial cluster centers are randomly chosen. In contrast, the fuzzy index value is hypothetically considered good between 1.1 and 5; there is no theoretical basis for it. The main problem is to decide the number of cluster in FCM [14]. The proposed method tries to solve this problem by using the reduction of n dimensions variables to k dimensions variables of PCA. The proposed algorithm has the following steps:

- The first step in image segmentation using the FCM is to read the input image and convert it into a grey image. The FCM is used on the grey images in the proposed method.

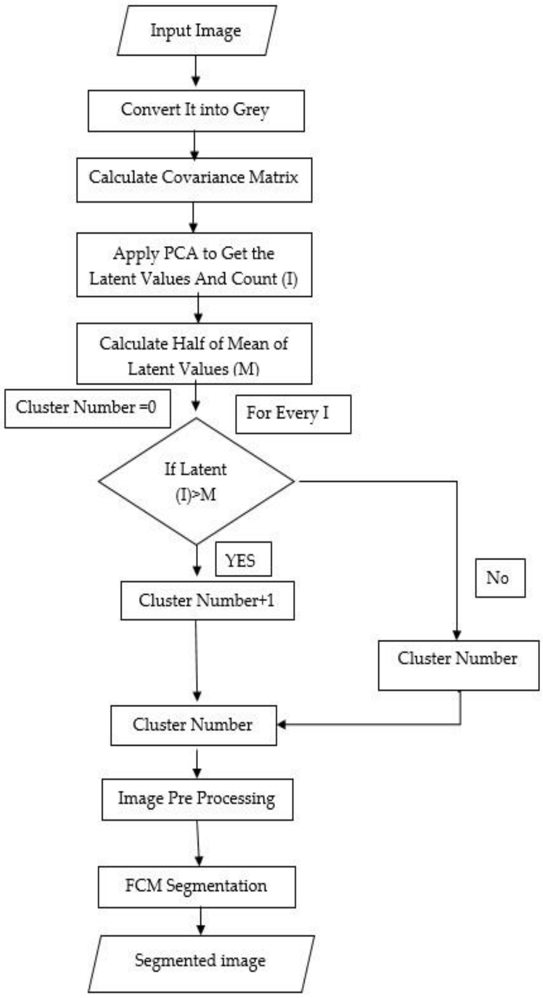

Fig1 Flowchart of the Proposed Algorithm

- The covariance matrix of regions is calculated to obtain the relationship between the pixels. A covariance matrix is a matrix whose x and y elements is in covariance position between the xth and yth elements of a random vector. A grey level covariance matrix is generated.
- Then the principal component analysis is applied on the grey level covariance matrix. The latent variables are extracted from the results. The latent variables are also called hidden variables. The latent variables are the true variables which do not considers all the features of an image but considers a few of them from which all others can be derived [15]. These few latent values can be used instead of other features of an image.
- Half of the mean of the latent variable values are calculated. This value is used to decide the number of clusters present in an image.
- The number of latent variables having a value higher than the value calculated in the above step is counted. This number is taken as cluster number.
- Now, FCM is applied to the pre-processed image having cluster number calculated from the above steps.
- These are the steps of the proposed algorithm. The proposed algorithm is applied to the number of test images of Berkeley's database.

ISSN No:-2456-2165

V. EXPERIMENTAL RESULTS

To assess the performance and efficiency of the proposed algorithm, several test images are used from Berkeley's database. The test images are converted to grey images, and a covariance matrix is generated. After that, PCA is applied to get the number of clusters, which is used as an input parameter to FCM, and the results are obtained. The results show the efficiency of the proposed method.

Table 1 Shows the (a) Original Image, (b) Segmented Image and (c) Cluster Number Calculated from the Proposed Method

Test images				
Segmented images				
Cluster numbers	**4**	**4**	**3**	**2**

The table 1 shows that the test images and their segmented images through Fuzzy c-means whose cluster numbers are calculated from the proposed algorithm. The figure also shows the calculated cluster number.
To quantitatively analyse the results of the proposed method, the size of a compressed or segmented image and the original image is given in Table 2

Table 2 Result Analysis Shows the Compressed Size of Images after Segmentation.

Result analysis		**Image 1**	**Image 2**	**Image 3**	**Image 4**
Original size (kb)		52.9	73.2	68.8	68.3
Segmented image size (kb)	Calculated Cluster-1	22.5	44.7	36.3	33.0
	Calculated Cluster	**20.5**	**25.6**	**23.2**	**23.0**
	Calculated Cluster+1	50.8	52.2	55.2	46.3
Compression percentage	Calculated Cluster-1	57.5	38.9	47.2	51.7
	Calculated Cluster	**61.2**	**65.0**	**66.3**	**66.4**
	Calculated Cluster+1	3.96	28.6	31.4	32.2

The table 2 shows that the proposed method gives good results in terms of segmented image as well as a compressed image. The comparative results are shown in the table which includes the size of segmented image using calculated cluster number (represented in bold characters), one less than the calculated cluster number and one greater than the calculated cluster number. The results show that the size of the segmented image using cluster number obtained from proposed algorithm is less than the other cluster numbers. It is also found that the compression percentage of all the segmented test images using the proposed method is above 50 percent, which is considered to be useful to reduce the size of the image. Hence, it is found that the proposed method gives good segmented image and has less size than the original image.

From the results, it is evident that the proposed method is capable of finding good cluster number values for various test images.

VI. CONCLUSIONS

The cluster number of an image is calculated using PCA applied to the covariance matrix. PCA is a dimension reducing algorithm which done the cluster number calculation task very efficiently. The image obtained after segmentation using the calculated cluster number is more compressed and small in size than the other segmented images. Hence, the results show that the cluster number obtained from the proposed method gives more accurate results than the other cluster numbers. The cluster number computation is a significant problem in the automation of the FCM algorithm, which is fulfilled by the proposed algorithm. There are other problems in FCM which can be considered for further research.

REFERENCES

[1]. R. M. Haralick, "Image segmentation survey," *Fundamentals in computer vision, OD Faugeras, ed., Cambridge Univ. Press, Cambridge,* pp. 209-224, 1983.
[2]. R. C. Tryon, *Cluster analysis: Correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality*. Edwards brother, Incorporated, litho printers and publishers, 1939.
[3]. J. C. Bezdek, "Cluster validity with fuzzy sets," 1973.
[4]. J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J.Cybern.*, vol. 3, pp. 32-57, 1973.
[5]. J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. New York, Plenum, 1981.
[6]. R. C. Tryon and D. E. Bailey, "The BC TRY computer system of cluster and factor analysis," *Multivariate Behavioral Research,* vol. 1, no. 1, pp. 95-111, 1966.
[7]. W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern recognition,* vol. 40, no. 3, pp. 825-838, 2007.
[8]. J. A. Hartigan, *Clustering algorithms*. 1975.
[9]. I. Jolliffe, "Principal component analysis," in *International encyclopedia of statistical science*: Springer, 2011, pp. 1094-1096.
[10]. M.-C. Hung and D.-L. Yang, "An efficient fuzzy c-means clustering algorithm," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 2001, pp. 225-232: IEEE.
[11]. N. Forghani, M. Forouzanfar, and E. Forouzanfar, "MRI fuzzy segmentation of brain tissue using IFCM algorithm with particle swarm optimization," in *Computer and information sciences, 2007. iscis 2007. 22nd international symposium on*, 2007, pp. 1-4: IEEE.
[12]. C. I. Gonzalez, P. Melin, J. R. Castro, O. Castillo, and O. Mendoza, "Optimization of interval type-2 fuzzy systems for image edge detection," *Applied Soft Computing,* vol. 47, pp. 631-643, 2016.
[13]. J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, no. 2-3, pp. 191-203, January 1984.
[14]. R. L. Cannon, J. V. Dave, and J. C. Bezdek, "Efficient implementation of the fuzzy c-means clustering algorithms," *IEEE transactions on pattern analysis and machine intelligence,* no. 2, pp. 248-255, March 1986.
[15]. S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems,* vol. 2, no. 1-3, pp. 37-52, 1987.