

# A Novel Image Segmentation of Galaxy Images through Optimized Fuzzy K-Means Clustering

J. Saranya<sup>1</sup>, Velmurugan<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, S.N.R Sons College, Coimbatore-6, India

<sup>2</sup>Associate Professor, Department of Computer Applications, S.N.R Sons College, Coimbatore-6, India

**Abstract:** *The image segmentation is the process of partition an image into meaningful regions with respect to a particular application. The segmentation is based on measurements taken from the image and might be grey level, colour, texture, depth or motion. Usually image segmentation is an initial and vital step in a series of processes aimed at overall image understanding. The applications of image segmentation include Identifying objects in a scene for object-based measurements such as size and shape. The objects are identified in the moving scene for the application of video compression. The identification of objects which are at different distances from a sensor using depth measurements, from a laser range finder that enabling path planning for a mobile robot. The traditional hyper connection algorithm provide the applications of image segmentation. It will segment and produce the result for normal images but it can't segment the overlapping image which present in the galaxy images. The fuzzy c-means algorithm is applied to produce the segmented results object in the form of clustering and it will take more computation time for each iterations of an image. The fuzzy c-means algorithm required a number of iterations to produce proper segmentation. In this paper the Fuzzy k-means algorithm is proposed to provide the effective image clustering and segmentation with less computation time and the number of iterations will not affect the image segmentation. To provide a effective segmentation initially the noises will be removed from the image using median filter. The median filter will reduce the noise content and it helps to bring the higher efficient image segmentation. The experimental results shown that the proposed system achieves high precision and outperforms the existing methods.*

**Keywords:** Image Segmentation, Fuzzy k-means, Median filter, hyper connections

## 1. Introduction

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. The image segmentation is the process of dividing an image into multiple parts. This is typically used to identify objects or other relevant information in digital images. The image segmentation technique is used to provide the required information from the image. In the galaxy image segmentation the varies objects like mass, stone, rock, stars and milky way paths will be segmented and each objects having different intensity values and based on this the area is calculated and it can be segmented. Initially when the images are captured from the satellite it contains some noises. This noise will be reduced by the effective segmentation. To remove this kind of noises in galaxy image median filter has been introduced [1] and [2].

The median filtering is a nonlinear method used to remove noises from images [4]. It is widely used as it is very effective at removing noises while preserving edges. It is particularly effective at removing 'salt and pepper' type noises. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighbouring pixels. The pattern of neighbours is called the "window", which slides, pixel by pixel over the entire image. The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value. The filtered images are used in the k-means clustering process as [9] and [10]. The clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. This is done such that patterns in the same cluster

are alike and patterns belonging to two different clusters are different. The rest of the paper is organized as follows: Section II describes the related work of state of art methods about image clustering. The section III describes the overall framework with methods and solution to achieve the image segmentation. The section IV describes the experimental results of our proposed method and performance measures with state-of-the-art methods. Finally section V concludes the paper and outlines possible future work.

## 2. Related Works

Median filters can be connected to morphological operations considering a lexicographic order. The detailed information about the mathematical theory on this connection may be found in [4]. H. Proenc, a and L.A. Alexandre [3] proposed the iris segmentation algorithm using the fuzzy- c means clustering method but it will took higher computational time for clustering. To overcome this drawbacks in previous methods the fuzzy k-means algorithm has been introduced. It provides the effective clustering methodology with the less computation time than the fuzzy c-means clustering method. In [5] the author has been explained a k means algorithm for image clustering in varies aspects. He also compared the other techniques with k-means algorithm. In [6] Mahamed G.H. Omran, Andries P Engelbrecht, and Ayed Salman has explained the unsupervised classification, this basic formation was used to implement the k-means algorithm. In [7] B. Zhang has presented k-means algorithm and explained the performance efficiency for varies iterations. In [8], [9] and [10] the authors gave the methodology to build the quality base assessment for clustering. In [10], [11] used to find the performance calculation for k-means clustering algorithm. In [15] Tapas Kanungo, David M. Mount, have presented an efficient implementation of Lloyd's k-means

clustering algorithm, called the filtering algorithm. The algorithm is easy to implement and only requires that a kd-tree be built once for the given data points. The efficiency is achieved because the data points do not vary throughout the computation and, hence, the data structure does not need to be recomputed at each stage. The experimental results also suggest that inter cluster separation plays a more important role in cluster validation than cluster diameter. The image segmentation is a key step for image understanding, which is a natural manner to obtain high-level semantic [19]. In JSEG approach [20] proposes a notion of “J-image” to measure the confidence of pixels to be boundaries or interiors of colour-texture regions, and uses a region growing method to segment the image based on the J-images. One major problem in JSEG is caused by the varying shades due to the illumination.

### 3. Proposed Methodology

The main objective of the proposed method is to predict the Galaxy specific image segmentation state through an optimized clustering. The linear clustering Suffix tree [16] and [17] separates the data, but it maximizes the distance between the given data point to the nearest data point of each class. The training data set is given by

$$D = \{(x_1, y_1)\Lambda, (x_i, y_i)\Lambda, (x_l, y_l)\}, \quad (1)$$

$$x \in R^n, y \in \{-1, 1\}$$

Where, l – number of training data,

X<sub>i</sub>–training data,

Y<sub>i</sub>–class label as 1 or -1 for x<sub>i</sub> for large data with drifting

A nonlinear function is adopted to map the original input space R<sup>n</sup> into N-dimensional feature space of the large dataset.

$$\psi(x) = \varphi_1(x), \varphi_2(x), \dots, \varphi_N(x) \quad (2)$$

The separating hyper plane is developed in this N-dimensional feature space. Then the clustering function represented as,

$$y(x) = \text{sgn}(\omega\psi(x) + b) \quad (3)$$

Where  $\omega$  - weight vector and b- scalar.

In order to obtain the optimal clustering through ensemble classifier  $\|\omega\|$  should be minimized subject to the following constraints

$$y_i[\varphi(x_i).\omega + b] \geq 1 - \xi_i, \quad i = 1, 2, \dots, l. \quad (4)$$

The variable  $\xi_i$  is the positive slack variables, necessary for misclassification of data in different cluster.

The Suffix tree clustering is known to be efficient in clustering large datasets. This clustering is one in all the best and also the best far-famed unsupervised learning algorithms that solve the well-known clustering problem in terms large data through the steps of big data community.

The objective function is given in,

$$\text{Min } J(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \quad (5)$$

we have,

$$y_i[\varphi(x_i) * \omega + b] \geq 1 - \xi_i \quad (6)$$

Where, C – margin parameter,  $\omega$  - weight vector,  $x_i$  - training data,  $y_i$  - class label (1 or -1)

For  $x_i, \xi_i$  - positive slack variables;  $\xi_i \geq 0, i = 1, \dots, l$ , b – scalar, l – number of training data.

The objective function obeys the principle of structural risk minimization in order to obtain the optimal solution with less false positive rate for the Image clustered. The objective function in Eqn (5) can be re-modified by following Lagrangian principle for the data segmentation and prediction as,

$$L(\omega, b, \xi, a, \gamma) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l a_i (y_i [\varphi(x_i) * \omega + b] - 1 + \xi_i) - \sum_{i=1}^l \gamma_i \xi_i \quad (7)$$

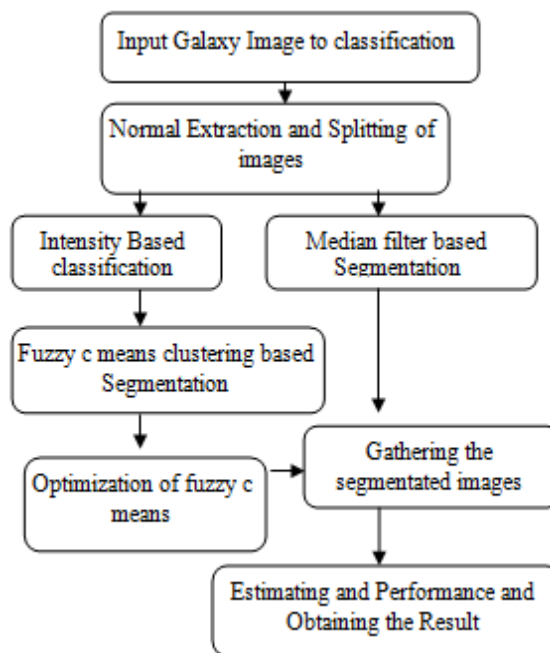


Figure 1: Block Diagram of Proposed Image Segmentation method

where,  $a_i \geq 0, \gamma_i \geq 0 (i = 1, 2, \dots, l), a_i, \gamma_i$  -

$$\max W(a) = -\frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j (\varphi(x_i), \varphi(x_j)) + \sum_{i=1}^l a_i \quad (9)$$

$$\max W(a) = -\frac{1}{2} \sum_{i,j=1}^l a_i a_j y_i y_j k(x_i, x_j) + \sum_{i=1}^l a_i \quad (10)$$

The suffix algorithm aims to partition a group of objects supported their attributes/features, into no. of feature clusters, wherever x may be a predefined or user-defined constant into x clusters. Temporal probability is carried out the density based clustering technique and its cluster employed through the ranking of the document, temporal

pattern relevance is also estimated from the cluster in terms of entropy and Euclidean calculation.

**The ranking Based on the Integration Values through following process:**

**1. Similarity estimation.**

Pair wise alignment is carried through the ranking based on the analysis and pair wise alignment is carried out through the algorithm is based on the observation that the data values belonging to the same attribute usually have the same data type and may contain similar strings, especially since results records of the query for the user query.

**2. Holistic alignment based prediction methods.**

Vertices from the same record are not allowed to be included in the same connected component as they are considered to come from two different attributes of the record. If two vertices from the same record breach this constraint, a path must exist between the two, which we call a breach path.

**3. Nested structure Alignment through user Specific clustering.**

Holistic data value alignment constrains a data value in a Result set to be aligned to at most one data value from another Result set. If a Result set contains a nested structure such that an attribute has multiple values, then some of the values may not be aligned to any other values. Therefore, nested structure processing identifies the data values of a Result set that are generated by nested structures.

**4. Experimental Results**

In this section, Experimental Results for query based prediction from big data with data evolution and feature evolution were carried out using web data and results were performed with performance system configurations to perform the data scaling and extracting into the proper clusters through suffix tree clustering. Initially extracting the framework has been utilized by training, validation and testing data for classification of results using historical prediction models identify the results set estimation efficiently and effectively in large dataset. The performances of the clustering and classification are experimented and presented in terms of relative speed, computational time as properties measure of performance using the large data set.

**(a)Query frequency estimation and temporal probability estimation**

The temporal prediction states observed from the large data set are as follows: supervised data, unsupervised data and semi supervised data.

**(b)Feature extraction through user query modeling**

Feature Extraction is employed in large dataset with data drifting and information retrieval with estimating various factors in the query analysis to the large dataset

Feature extraction:

- (1) The data in the big data is evolved with several feature classification with novel features estimation in each sample such as,  $y_1, y_2, y_3, y_4$  and  $y_5$ , are extracted by the equation as follows:

$$y_k = \frac{c^k}{5} \max_{i=1}^5(c^i) \quad (11)$$

Where  $k=1, 2, \dots, 5$ ,

$C_k$  – Absolute feature data per one sample.

- (2) The absolute information is calculated for different samples given by,

$$Y_6 = \log_{10} \left( \max_{m=1}^5 c^m \right) \quad (12)$$

**Table 1:** Parameters of classification and Prediction of data classification

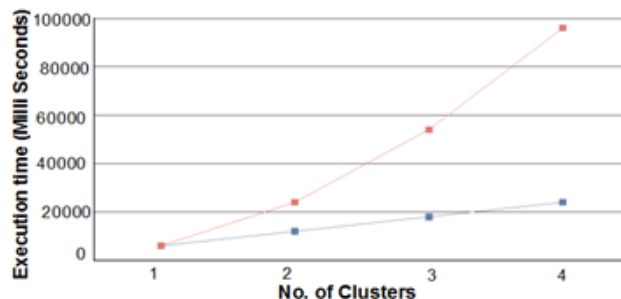
Parameters	Notations used	Values
Learning rate	$\Lambda$	0.01
Scaling factor	$\Sigma$	1

**Table 2:** Performance Parameters to compute Data Extraction mechanism

Parameters	Notations used	Values
Number of iteration	I	15000
Order of the polynomial	Order	3
Scaling factor	$\Sigma$	1

**(c)Result Analysis**

The proposed framework is implemented and tested using different types of Image datasets. An extensive experimental study was conducted to evaluate the efficiency and effectiveness of the proposed methodology on various parameters of benchmark instances and the prediction states are obtained in the graph



The proposed has been done for our project using the database like DIBCO, Galaxy PGC35538 and SDSS-I. The PSNR and F-measures are calculated as shown in Table 3:

**Table 3:** Calculation of F measures and PSNR values

Methods	F measures	PSNR
Building a component tree in quasi linear time method	91.25	18.23
Mask based second generation connectivity and attribute filters method	91.26	18.30
Hyper connection and hierarchical representation using sobel method	87.85	17.11
K-means segmentation	85.21	16.53

The proposed method gives the F-measure and PSNR values as 85.21 and 16.53 for the DIBCO dataset, 85.11 and 16.78 for the Galaxy PGC35538 and 85.20 and 16.60 for SDSS-I(2005-2008). Compare to the previous methods, the proposed method gives the better PSNR and F –measures.



## 5. Dataset

### DIBCO Image



Figure 2: Input noisy image



Figure 3: Red intensity image



Figure 4: Median filter output

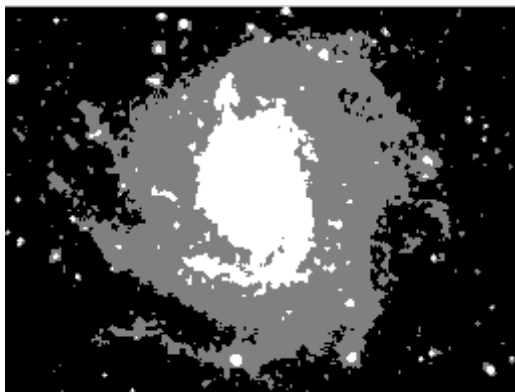


Figure 5: Red intensity clustered image

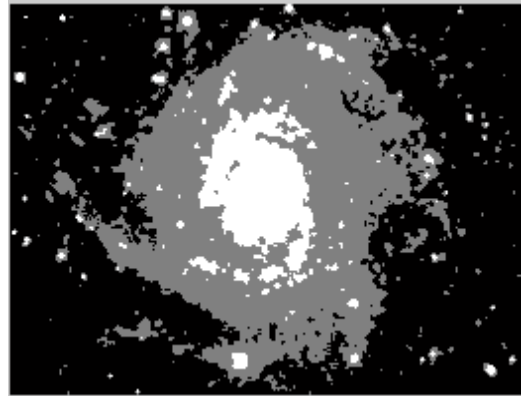


Figure 6: Green intensity clustered image

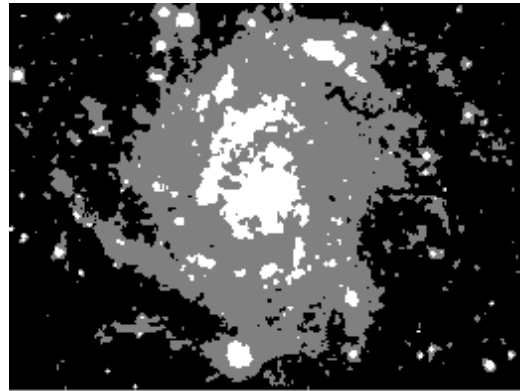


Figure 7: Blue intensity clustered image

## 6. Conclusion

The proposed system has been presented the effective segmentation algorithm for galaxy images using fuzzy k-means. The system is modeled based on its intensity and algorithm provides better the results with number of clusters for segmentation. This cluster provides the better segmentation results with less computation time. Experimental results shown that proposed system achieves high precision and outperforms existing state-of-the-art data clustering methods.

## References

- [1] Valavanis, K.P., Zheng, J., Gauch, J.M.: On Impulse Noise Removal in Colour Images. In: Proc. Int. Conf. on Robotics and Automation, Sacramento, Ca., (1991) 144-149.
- [2] Zheng, J., Valavanis, K.P., Gauch, J.M.: Noise Removal from Color Images. J. Intelligent and Robotic Systems 7 (1993), 257-285.
- [3] H. Proença and L.A. Alexandre "Iris segmentation methodology for non-cooperative recognition" in IEE proceedings vision, image and signal processing, april, 2006, volume 153, issue 2.
- [4] Caselles, V., Sapiro, G., Chung, D.H.: Vector Median Filters, Inf-Sup Operations, and Coupled PDE's: Theoretical Connections. J. of Math. Imaging and Vision 12 (2000)109-120.
- [5] G.Hamerly, C. Elkan, "Alternatives to the k-means algorithm that find better clusterings", Proc. of the ACM Conference on Information and Knowledge Management, CIKM-2002, 2002, 600-607.

- [6] G.H. Omran, A.Salman, A.P. Engelbrecht, "Dynamic clustering using particle swarm optimization with application in image segmentation", *Pattern Anal Applic.*, 8, 2006, 332-344.
- [7] B. Zhang, "Generalized k-harmonic means - Boosting in unsupervised learning. Technical Report HLP-2000-137", Hewlett-Packard Labs, 2000.
- [8] M. Halkidi, M. Vazirgiannis, I. Batistakis, "Quality scheme assessment in the clustering process". In Proc. of the 4th European Conf. on Principles of Data Mining and Knowledge Discovery, LNCS 1910, 2000, 265 -267.
- [9] M.Halkidi et al., "Clustering validity checking methods: Part II", *SIGMOD Rec.*, 31, No. 3, 2002, 19-27 .
- [10] R.H. Turi, "Clustering-based color image segmentation", PhD Thesis, Monash University, Australia 2001.
- [11] J.C. Bezdek, N.R. Pal, "Some new indexes of cluster validity", *IEEE Trans. Sys. Man. Cyb.*, 28, 1998, 301-315.
- MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proc. 5th Symp. Math Statist Prob (pp. 281–297).
- [12] Arthur Dempster, Nan Laird, and Donald Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977
- [13] R. C. Dubes and A. K. Jain. *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [14] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine Piatko, Ruth Silverman, Angela Y. Wu. "An Efficient k-Means Clustering Algorithm: Analysis and Implementation". 2000
- [15] T. Pappas, "An adaptive clustering algorithm for image segmentation, "in *IEEE Trans. Signal Processing*, 1992, vol. 40, pp. 901–914".
- [16] Giegerich, R.; Kurtz, S. (1997), "From Ukkonen to McCreight and Weiner: A Unifying View of Linear-Time Suffix Tree Construction".
- [17] Farach-Colton, Martin; Ferragina, Paolo; Muthukrishnan, S. (2000), "On the sorting-complexity of suffix tree construction".
- [18] M. Betke, N. C. Makris, "Fast Object Recognition in Noisy Images Using Simulated Annealing", *Intl. Conf. on Computer Vision*, pp. 523, Jun. 20-23, 1995.
- [19] Y. Deng, B. Manjunath, "Unsupervised segmentation of color-texture regions in images and video", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800-810, August 2001.