

Image Segmentation by Clustering Methods: Performance Analysis

B.Sathya

Department of Applied Science
Vivekanandha Institute of Engineering and
Technology for Women
Thiruchengode, Tamilnadu, India

R.Manavalan

Department of Computer Science (PG)
K.S.R College of Arts and Science
Thiruchengode, Tamilnadu, India

ABSTRACT

Image segmentation plays a significant role in computer vision. It aims at extracting meaningful objects lying in the image. Generally there is no unique method or approach for image segmentation. Clustering is a powerful technique that has been reached in image segmentation. The cluster analysis is to partition an image data set into a number of disjoint groups or clusters. The clustering methods such as k means, improved k mean, fuzzy c mean (FCM) and improved fuzzy c mean algorithm (IFCM) have been proposed. K means clustering is one of the popular method because of its simplicity and computational efficiency. The number of iterations will be reduced in improved K compare to conventional K means. FCM algorithm has additional flexibility for the pixels to belong to multiple classes with varying degrees of membership. Demerit of conventional FCM is time consuming which is overcome by improved FCM. The experimental results exemplify that the proposed algorithms yields segmented gray scale image of perfect accuracy and the required computer time reasonable and also reveal the improved fuzzy c mean achieve better segmentation compare to others. The quality of segmented image is measured by statistical parameters: rand index (RI), global consistency error (GCE), variations of information (VOI) and boundary displacement error (BDE).

Keywords

K means, improved k means, fuzzy c means, improved c means, rand index, global consistency error, variations of information

1. INTRODUCTION

Image segmentation can be defined as the classification of all the picture elements or pixels in an image into different clusters that exhibit similar features. Segmentation involves partitioning an image into groups of pixels which are homogeneous with respect to some criterion [1]. Different groups must not intersect each other and adjacent groups must be heterogeneous. The groups are called segments. Image segmentation is considered as an important basic operation for meaningful analysis and interpretation of image acquired. It is a critical and essential component of an image analysis and or pattern recognition system, and is one of the most difficult tasks in image processing, which determines the quality of the final segmentation. Researchers have extensively worked over this fundamental problem and proposed various methods for image segmentation. These methods can be broadly classified into seven groups: (1) Histogram thresholding, (2) Clustering (Fuzzy and Hard), (3) Region growing, region splitting and merging, (4)

Edge-based, (5) Physical model- based, (6) Fuzzy approaches, and (7) Neural network and GA (Genetic algorithm) based approaches.[1],[2]

2. PROBLEM DESCRIPTION

An image may be defined as a two dimensional function $f(x, y)$, where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of co-ordinates (x, y) is called the intensity or gray level of the image at that point. The region of interest in the image can be degraded by the impact of imperfect instrument, the problem with data acquisition process and interfering natural phenomena. Therefore the original image may not be suitable for analysis. Thus image segmentation technique is often necessary and should be taken as significant step during image is processed and analyzed. Repeatable experiments with published benchmarks are required for this research field to progress. The problem addressed in this thesis is that the image is spilt into number of segmentation [1]. Choosing an appropriate model for segmentation is difficult task that the modal has the better segmentation with reduced computational time. The problem is reformed with minimized computational time and high quality of the results.

3. IMAGE SEGMENTATION BY CLUSTERING METHODS

Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A definition of clustering could be “the process of organizing objects into groups whose members are similar in some way”. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. [3], [4], [5], [6]

3.1 K Means Clustering

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem [8], [14]. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori [15], [16]. This algorithm aims at minimizing an objective function, in this case a squared error function. The objective function,

$$J = \sum_{j=1}^k \sum_{i=1}^x \left\| x^{(j)} - c_j \right\|^2$$

Where $\|x^{(j)} - c_j\|^2$ is a chosen distance measure between a data point $x^{(j)}$ and the cluster centre c_j , is an indicator of the distance of the n data points from their respective cluster centers.

The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

3.2 Improved K means Clustering

Let $D = \{d^{(j)} / j = 1, \dots, n\}$ be a data set having K -Clusters, $C = \{c_i / i = 1, \dots, k\}$ be a set of K centers and $S_j = \{d^{(j)} \mid d^{(j)} \text{ is member of cluster } K\}$ be the set of samples that belong to the j th cluster [12],[17]. The following function which is defined as an objective function,

$$\text{Cost}(D, C) = \sum_{j=1}^n \text{dist}(d^{(j)}, c_k)$$

Where $\text{dist}(d^{(j)}, c_k)$ measures the Euclidean distance between a points $d^{(j)}$ and its cluster center c_k .

The algorithm has the following steps.

1. Dividing D into K parts according to data patterns;

$$D = \bigcup_{k=1}^K S_k, S_{k_1} \cap S_{k_2} = \emptyset, k_1 \neq k_2$$

2. Let $x^{(0)}$, $k = 1, \dots, K$ be initial clustering centers calculated by,

$$x^{(0)} = \sum_{d^{(j)} \in S_k} d^{(j)} / |S_k|, k = 1, \dots, K.$$

3. Decide membership of the patterns in each one of the K -clusters according to the minimum distance from cluster center criteria.

4. Calculate new centers by the following iterative formula,

$$x_i^{(k+1)} = \sum_{d^{(j)} \in S_k} \frac{d_i^{(j)}}{q_j} / \sum_{j=1}^n \frac{1}{q_j}$$

Where $q_j^k = \|x^{(k)} - d^{(j)}\|$.

5. Repeat steps 3 and 4 till there is no change in cluster centers.

3.3 Fuzzy C-Means Clustering

Fuzzy c -means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [12],[17]. That is it allows the pixels belong to multiple classes with varying degrees of membership. It is based on minimization of the following objective function:

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2$$

Where, m is any real number greater than 1. u_{ij} is the degree of membership of x_i in the cluster j

x_i is the i^{th} of d -dimensional measured data,

c_j is the d -dimension center of the cluster.

The algorithm is composed of the following steps:

1. Initialize $U = [u_{ij}]$ matrix, $U(0)$
2. At k -step: calculate the centers vectors $c(k) = [c_j]$ with $U(k)$

$$c_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m}$$

3. Update $U^{(k)}, U^{k+1}$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

4. If $\|U^{(k+1)} - U^{(k)}\| < \epsilon$ then STOP; otherwise return to step 2.

3.4 Improved Fuzzy C-Means

The improved FCM algorithm is based on the concept of data compression where the dimensionality of the input is highly reduced [20]. The data compression includes two steps: quantization and aggregation.

The quantization of the feature space is performed by masking the lower 'm' bits of the feature value. The quantized output will result in the common intensity values for more than one feature vector. In the process of aggregation, feature vectors which share common intensity values are grouped together. A representative feature vector is chosen from each group and they

are given as input for the conventional FCM algorithm. Once the clustering is complete, the representative feature vector membership values are distributed identically to all members of the quantization level. Since the modified FCM algorithm uses a reduced dataset, the convergence rate is highly improved when compared with the conventional FCM.

The improved FCM algorithm uses the same steps of conventional FCM except for the change in the cluster updating and membership value updating criterions. The modified criterions are showed below,

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m y_j}{\sum_{j=1}^n u_{ij}^m}, \quad u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}}$$

Where $d_{ij} = y_j - c_i$

y = Reduced Dataset

4. EXPERIMENTAL RESULTS ANALYSIS AND DISCUSSION

The proposed algorithms have been implemented using MATLAB. The performance of various image segmentation approaches are analyzed and discussed. The measurement of image segmentation is difficult to measure. There is no common algorithm for the image segmentation. The statistical measurements could be used to measure the quality of the image segmentation [21], [22]. The rand index (RI), global consistency error (GCE), boundary displacement error (BDE), and variations of information (VOI) are used to evaluate the performance. The detailed description with formulae of RI, GCE, BDE, VOI parameters are explained in detail as follows,

4.1 Rand Index (RI)

The Rand index (RI) counts the fraction of pairs of pixels whose labeling are consistent between the computed segmentation and the ground truth averaging across multiple ground truth segmentations[21]. The Rand index or Rand measure is a measure of the similarity between two data clusters. Given a set of n elements and two partitions of S to compare, and, we define the following:

a, the number of pairs of elements in S that are in the same set in X and in the same set in Y

b, the number of pairs of elements in S that are in different sets in X and in different sets in Y

c, the number of pairs of elements in S that are in the same set in X and in different sets in Y

d, the number of pairs of elements in S that are in different sets in X and in the same set in Y

The Rand index (R) is,

$$R = \frac{a + b}{a + b + c + d} = \frac{a + b}{\binom{n}{2}}$$

Where, a + b as the number of agreements between X and Y and c + d as the number of disagreements between X and Y. The Rand index has a value between 0 and 1, with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same.

4.2 Variation of Information (VOI)

The Variation of Information (VOI) metric defines the distance between two segmentations as the average conditional entropy of one segmentation given the other, and thus measures the amount of randomness in one segmentation which cannot be explained by the other [21]. Suppose we have two clustering (a division of a set into several subsets) X and Y where $X = \{X_1, X_2, \dots, X_k\}$, $p_i = |X_i| / n$, $n = \sum_k |X_k|$. Then the variation of information between two clustering is:

$$VI(X; Y) = H(X) + H(Y) - 2I(X, Y)$$

Where, H(X) is entropy of X and

I(X, Y) is mutual information between X and Y. The mutual information of two clustering is the loss of uncertainty of one clustering if the other is given. Thus, mutual information is positive and bounded by $\{H(X), H(Y)\}_{\log_2(n)}$

4.3 Global Consistency Error (GCE)

The Global Consistency Error (GCE) measures the extent to which one segmentation can be viewed as a refinement of the other [21]. Segmentations which are related are considered to be consistent, since they could represent the same image segmented at different scales. Segmentation is simply a division of the pixels of an image into sets. The segments are sets of pixels. If one segment is a proper subset of the other, then the pixel lies in an area of refinement, and the error should be zero. If there is no subset relationship, then the two regions overlap in an inconsistent manner.

The formula for GCE is as follows,

$$GCE = \frac{1}{n} \min \left\{ \sum_i E(s1, s2, p_i), \sum_i E(s2, s1, p_i) \right\}$$

Where, segmentation error measure takes two segmentations S1 and S2 as input, and produces a real valued output in the range [0:1] where zero signifies no error. For a given pixel p_i consider the segments in S1 and S2 that contain that pixel.

4.4. Boundary Displacement Error (BDE)

The Boundary Displacement Error (BDE) measures the average displacement error of one boundary pixels and the closest boundary pixels in the other segmentation[21].

$$\mu_{LA}(u, v) = \begin{cases} \frac{u - v}{L - 1} & 0 < u - v \leq L - 1 \\ 0 & 0 \leq u - v < 0 \end{cases}$$

Let $\mu_{LA}(u, v)$ denotes the membership function that describes the fuzzy relation.

The experiment is conducted over the five images using the algorithms k means, improved k means, FCM, improved fuzzy c means and their results shown in Fig.1 with required statistical parameters and their results are presented in Table 1. The average of results is shown in Table 2. If the value of RI is

higher and GCE, BDE, VOI are lower then the segmentation approach is better.

The Fig.2 average performance analysis chart reveals that the rand index of improved fuzzy c-mean is higher than others and also the global consistency error, variation of information, and boundary displacement error are lower than others. The detailed analysis of the individual statistical measure is given in Fig.3. Table 3 show that the time evaluation of proposed algorithms. The average time of each method projected in the chart Fig.4. It shows that the K means take minimum time compare to others, however it provides poor results. The output provided by the improved k is also pitiable even though the number of iterations taken by improved k means. Comparatively the FCM algorithm provides good result but it acquire more time than K means and improved K means. The average time taken for five images by IFCM is comparatively lower to traditional FCM method. So it was observed that the proposed method IFCM performs better compare to others approaches.

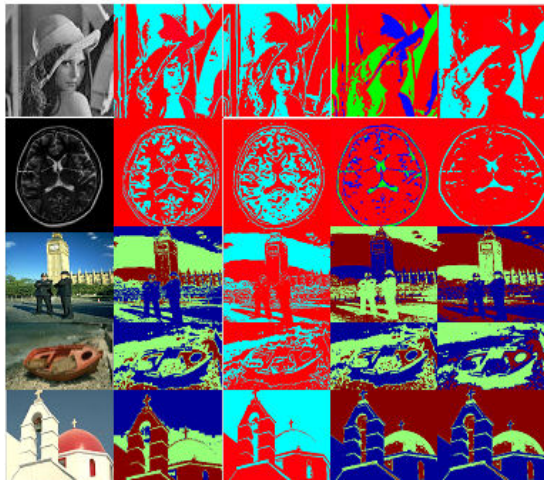


Fig 1: segmentation results using clustering

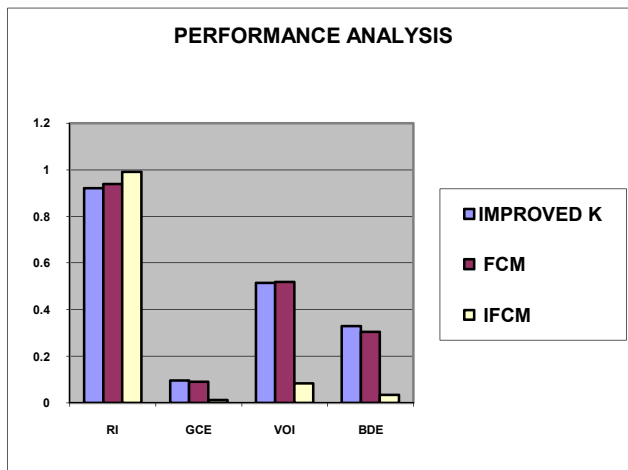


Fig.2 Performance Analysis Chart

Table 1.Performance Evaluation

IMAGES	METHODS	RI	GCE	VOI	BDE
IMG 1	Improved K	0.9874	0.0195	0.1437	0.1167
	FCM	0.9897	0.0164	0.1260	0.1120
	IFCM	1	0	0	0
IMG 2	Improved K	0.9307	0.0924	0.5029	0.1229
	FCM	0.9227	0.1013	0.5401	0.1419
	IFCM	0.9907	0.0137	0.0976	0.0171
IMG 3	Improved K	0.858	0.1598	0.8287	0.2855
	FCM	0.8509	0.1678	0.8676	0.2109
	IFCM	0.9852	0.0227	0.16	0.1534
IMG 4	Improved K	0.8580	0.1598	0.8287	0.1360
	FCM	0.9509	0.1378	0.8676	0.0852
	IFCM	0.9852	0.0227	0.1600	0
IMG 5	Improved K	0.9755	0.0453	0.2677	0.9871
	FCM	0.9843	0.0299	0.1887	0.9748
	IFCM	1	0	0	0

Table 2.Average Calculation of Performance Analysis

METHOD S	RI	GCE	VOI	BDE
IMPROVE D K	0.92192	0.09536	0.51434	0.32964
FCM	0.9397	0.09064	0.5180	0.30496
IFCM	0.99222	0.01182	0.08352	0.0341

Table 3. Time Evaluation

IMAGES	K MEANS	IMPROVED K	FCM	IFCM
ONE	1.3454	1.2874	2.3355	1.6147
TWO	2.5121	1.6592	2.9285	2.1164
THREE	0.4923	1.5493	3.7213	1.7776
FOUR	1.4986	2.9096	2.5963	2.2608
FIVE	2.4846	1.8098	3.3851	2.8517

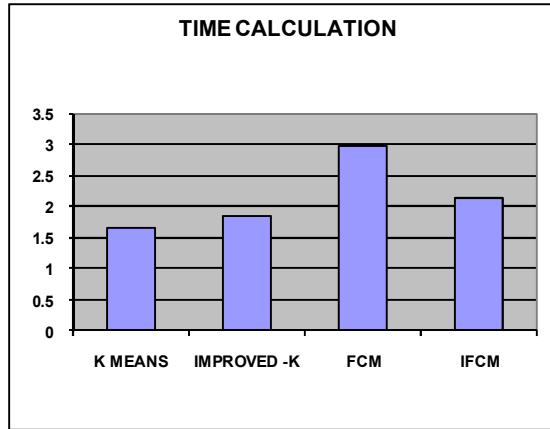


Fig.4 Average Time Calculation

5. CONCLUSION

In this paper, the unsupervised method i.e. cluster based algorithms were proposed for image segmentation. The clustering techniques such as k means, improved k means, fuzzy c mean, improved fuzzy c means were tested in different images. The performance of proposed algorithms is measured using segmentation parameters RI, GCE, VOI, and BDE. The computational results showed that the K means image segmentation consumes less time but it provide poor result. The modified k means algorithm takes minimum numbers of iterations compare to k means. The conventional FCM consume more time and provide good result where as the improved FCM algorithm consume less time compare to traditional FCM and provide good result. Therefore from the computational results conclude that the proposed algorithms the improved FCM algorithm performed better than others in terms of performance accuracy and better convergence rate

6. REFERENCES

[1] Krishna Kant Singh , Akansha Singh, A Study Of Image Segmentation Algorithms For Different Types Of Images IJCSI International Journal of Computer Science Issues 2010

[2] F. MarquCs B. Marcotenui, F. Zanoguera, partion based image representation as basis for user assisted image segmentation,2000

[3] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press (1996)

[4] M.N. Murty, A.K. Jain, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31(3) (1999) 264-323

[5] A.K. Jain, R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, Englewood Cliffs, NJ(1988)

[6] R.T. Ng, J. Han, Efficient and effective clustering methods for spatial data mining, in: Proceedings of the Twentieth International Conference on Very Large Databases, Santiago, Chile(1994) 144-155620

[7] Su, M.C., Chou, C.H., A modified version of the K-means algorithm with a distance based on cluster symmetry. IEEE Trans. Pattern Anal. Machine Intel, 23(6)(2001) 674-680

[8] Xiuyun Li, Jie Yang, Qing Wang, Jinjin Fan, Peng Liu, 2010 Research and Application of Improved K-means Algorithm Based on Fuzzy Feature Selection

[9] Kuo-Liang Chung, Keng-Sheng Lin, An efficient line symmetry-based K-means algorithm, Pattern Recognition Lett, 27(2006) 765-772

[10] Sugar, S.A., James, G.M., Finding the number of clusters in a dataset: an information-theoretic approach. J. Amer. Stat. Assoc. 98(2003) 750-763

[11] Kanugo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., An efficient K-means algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. 24(2002) 881-892

[12] Mario G.C.A. Cimino, Beatrice Lazzerini and Francesco Marcelloni, A novel approach to fuzzy clustering based on a dissimilarity relation extracted from data using a TS system, Pattern Recognition, 39(11)(2006) 2077-2091

[13] Meila, M., Heckerman, D., An experimental comparison of several clustering methods, Microsoft Research Report MSR-TR-98-06, Redmond, WA.(1998)

[14] Shehroz S. Khan, Amir Ahmad, Cluster center initialization algorithm for K-means clustering, Pattern Recognition Letters 25 (2004) 1293-1302

[15] Bradley, P.S., Fayyad, U.M., Refining initial points for Kmeans clustering. In: Sharlik, J. (Ed.), Proc. 15th Internat. Conf. on Machine Learning (ICML98). Morgan Kaufmann, San Francisco, CA,(1998) 91-99 621

[16] K.S.Ravichandran and B. Ananthi, "Color Skin Segmentation Using K-Means" Cluster International Journal of Computational and Applied Mathematics ISSN 1819-4966 Volume 4 Number 2 (2009), pp. 153–157

[17] Jude hemanth.D, D.Selvathi and J.Anitha, "Effective Fuzzy Clustering Algorithm for Abnormal MR Brain Image Segmentation", International/Advance Computing Conference (IACC 2009), IEEE, 2009.

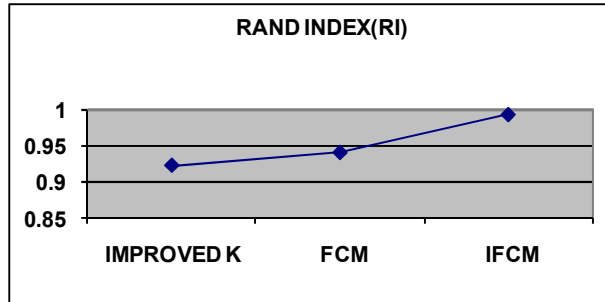
[18] Sorin Istrail, "An Overview of Clustering Methods", With Applications to Bioinformatics.

[19] Shahram Rahimi, M. Zarghamy A. Thakrez D. Chhillar, "A Parallel Fuzzy C Mean algorithm for Image Segmentation".

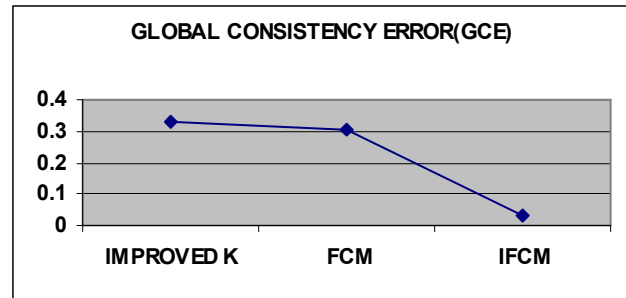
[20] P. Vasuda et. al. / (IJCSE) International Journal on Computer Science and Engineering 2010, Improved Fuzzy C-Means Algorithm for MR Brain Image Segmentation

[21] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 929–944, Jun. 2007.

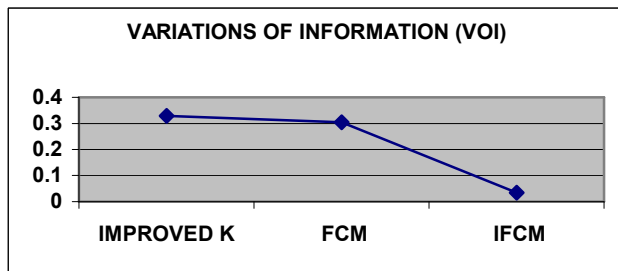
[22] F. Ge, S.Wang, and T. Liu, "New benchmark for image segmentation evaluation," J. Elect. Imag., vol. 16, no. 3, Jul.–Sep. 2007.



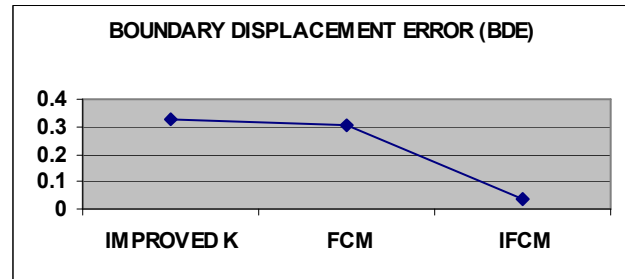
(a)



(b)



(c)



(d)

Fig.3 Statistical Measures (a) RI (b) GCE (c) VOI, (d) BDE