Fuzzy Clustering For Color Image Segmentation Using Gravitational Search Algorithm

Dissertation

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF TECHNOLOGY

IN

SOFTWARE TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SUBMITTED BY

ARVIND KUMAR

ROLL NO:

2K13/SWT/05

MAJOR PROJECT REPORT II

(Paper Code: CO 821)

Under the guidance of

Prof. O.P. Verma

SHAHBAD DAULATPUR, MAIN BAWANA ROAD, NEW DELHI, DELHI

110042 INDIA

DELHI TECHNOLOGICAL UNIVERSITY

NEW DELHI

STUDENT DECLARATION

I hereby undertake and declare that this submission is my original work and to the best of my knowledge and believe, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of any Institute or other University of higher learning, except where due acknowledgement has been made in the text. Project work and published paper associated to the chapters are well discussed and improved under the guide supervision.

DATE:

SIGNATURE:

ARVIND KUMAR

ROLL NO: 2K13/SWT/05

DELHI TECHNOLOGICAL UNIVERSITY

NEW DELHI

CERTIFICATE

This is to certify that the thesis entitled "Fuzzy Clustering For Color Image Segmentation Using Gravitational Search Algorithm", is a bonafide work done by Mr. ARVIND KUMAR in partial fulfillment of requirements for the award of Master of Technology Degree in software technology at Delhi Technological University (New Delhi) is an authentic work carried out by him under my supervision and guidance. The matter embodied in the thesis has not been submitted to any other University / Institute for the award of any Degree or Diploma to the best of my knowledge.

DATE:

SIGNATURE:

Dr. O.P Verma

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGMENT

I am presenting my work on “Fuzzy Clustering For Color Image Segmentation Using Gravitational Search Algorithm” with a lot of pleasure and satisfaction. I take this opportunity to thank my supervisor, Prof. O.P. Verma, for guiding me and providing me with all the facilities, which paved way to the successful completion of this work. His scholarly guidance and invaluable suggestions motivated me to complete my thesis work successfully. I am thankful to my friends and colleagues who have been a source of encouragement and inspiration throughout the duration of this thesis. I am also thankful to the SAMSUNG India Electronic Pvt Ltd. who has provided me opportunity to enroll in the M.Tech Program and to gain knowledge through this program. This curriculum provided me knowledge and opportunity to grow in various domains of computer science. Last but not least, I am thankful to all the faculty members who visited the Samsung premises to guide and teach. Their knowledge and efforts helped me to grow and learn in the field of computer science. This project has provided me knowledge in the area of Segmentation techniques used for color image using gravitational search algorithm, selecting the Data set using fuzzy technique. I have given ample time and guidance to complete my project under timeline defined by the university.

ARVIND KUMAR

ROLLNO: 2K13/SWT/05

ABSTRACT

Clustering is a key activity in numerous data mining applications such as information retrieval, text mining, image segmentation, etc. This research work proposes a clustering approach, Fuzzy-GSA, based on gravitational search algorithm (GSA). In the proposed Fuzzy-GSA approach, a fuzzy inference system is developed to effectively control the parameters of GSA.

The performance of the Fuzzy-GSA algorithm is evaluated against four benchmark datasets from the UC Irvine repository. The results illustrate that the Fuzzy-GSA approach attains the highest quality clustering over the selected datasets when compared with several other clustering algorithms namely, k-means, particle swarm optimization (PSO), gravitational search algorithm (GSA) and, combined gravitational search algorithm and k-means approach (GSA-KM)

In this paper, we propose a new hybrid approach for image segmentation. The proposed approach exploits fuzzy GSA for clustering image pixels into homogeneous regions. In order to improve the performance of fuzzy clustering to cope with segmentation problems, we employ gravitational search algorithm which is inspired by Newton's rule of gravity.

Gravitational search algorithm is incorporated into fuzzy GSA to take advantage of its ability to find an optimum cluster center which minimizes the fitness function of fuzzy GSA. Experimental results show effectiveness of the proposed method in segmentation different types of images as compared to classical fuzzy Algorithm.

Table of Contents

Chapter One: 7
INTRODUCTION 7

SCOPE & MOTIVATION: 7

DIGITAL IMAGE SEGMENTATION 8

THE STUDY OF COLOR IMAGE SEGMENTATION 8

METHOD FOR COLOR IMAGE SEGMENTATION 9
- Region Based Techniques 9
- Split and Merge Technique 9
- Normalized Cuts 10
- Region Growing 10
- Thresholding 11
- Edge Based Techniques 11

GRAVITATIONAL SEARCH ALGORITHM 14
Chapter Two: 17
LITERATURE REVIEW 17

THE DEVELOPED FUZZY INFERENCE SYSTEM 18
Chapter Three: 21
PROPOSED WORK 21
Chapter Four: 24
RESULTS 24
Chapter Five: 30
CONCLUSION & FUTURE WORK 30
Chapter Six: 31
REFERENCES 31

Chapter One:

INTRODUCTION

Scope & motivation:

People are only interested in certain parts of the image in the research and application of the image. These parts are frequently referred as a target or foreground (other part is called background), they generally correspond to the image in a specific and unique nature of the area. It needs to extract and separate them in order to identify and analyze object, on this basis it will be possible to further use for the target. To illustrate the level of the image segmentation in image processing, we have introduced "image engineering" concept ", it bring the involved theory, methods, algorithms, tools, equipment of image segmentation into an overall framework [1.]. Image Engineering is a new subject for research and application of image field, its content is very abundant. According to the different of the abstract degree and research methods, it can be divided into three levels: Image processing, image analysis and image understanding. As shown in Figure 1 Image processing is emphasis on the transformation between the images and improves the visual effects of image. Image analysis is mainly monitor and measure the interested targets in the image in order to get its objective information as a result build up a description of the image, the key point of the image understanding is further study on the nature of each target and the linkage of each other as well obtain an explanation of objective scenario for original image as result guide and plan to action.
Image processing, image analysis and image understanding have different operational, refer to Figure1.

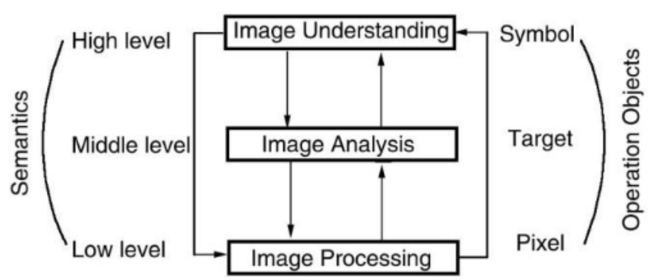


Fig 1: image engineering

Image processing is relatively low-level operations; it is mainly operated on the pixel-level. Then image analysis enters the middle-level, it focuses on measuring, expression and description of target. Image Understanding is mainly high-level operation, essentially it focus on the operation and illation of data symbol which abstracts from the description [2.]. Image segmentation is a key step from the image processing to image analysis, it occupy an important place. On the one hand, it is the basis of target expression and has important effect on the feature measurement. On the other hand, as the image segmentation, the target expression based on segmentation, the feature extraction and parameter measurement that converts the original image to more abstract and more compact form, it is possible to make high-level image analysis and understanding. For example, satellite image processing in the application of remote sensing; the brain MR image analysis in the applications of medicine; the plates of illegal vehicle region segmentation in the traffic image analysis; the image region of interest extraction in the object-oriented image compression and content-based image retrieval.

Digital Image Segmentation

Digital image processing is important domain for many reasons. Actually Digital image processing is a recent subject in computer history. In 1960s; Bell Labs and University of Maryland, and a few other places started to develop several techniques for digital image processing[3.] With application to satellite imagery, wire photo standards conversion, medical imaging, videophone, character recognition, and photo enhancement. But the cost of processing was fairly high with the computing equipment of that era. In the 1970s, image processing proliferated, when cheaper computers and dedicated hardware became available. Images could then be processed in real time, for some dedicated problems such as television standards conversion. As general-purpose computers became faster, they started to take over the role of dedicated hardware for all but the most specialized and compute-intensive operations. In digital image processing, we use computer algorithms to perform image processing. Actually digital image processing has several advantages over the analog image processing; first it gives a high number of algorithms to be used with the input data, second we can avoid some processing problems such as creating noise and signal distortion during signal processing. In 2000s, fast computers became available for signal processing and digital image processing has become the popular form of image processing. Because of that, signal image processing became versatile method, and also cheapest. Image segmentation is important part in many signal processing technique and its applications.

The study of color image segmentation

Image segmentation is the process of separating or grouping an image into different parts. These parts normally correspond to something that humans can easily separate and view as individual objects. Computers have no means of intelligently recognizing objects, and so many different methods have been developed in order to segment images. The segmentation process in based on various features found in the image. This might be color information that is used to create histograms, or information about the pixels that indicate edges or boundaries or texture information. The color image segmentation is also widely used in many multimedia applications, for example; in order to effectively scan large numbers of images and video data in digital libraries, they all need to be compiled directory, sorting and storage, the color and texture are two most important features of information retrieval based on its content in the images and video. Therefore, the color and texture segmentation often used for indexing and management of data; another example of multimedia applications is the dissemination of information in the network. Today, a large number of multimedia data streams sent on the Internet, However, due to the bandwidth limitations; we need to compress the data, and therefore it calls for image and video segmentation.

Method for color Image Segmentation

Image segmentation methods are categorized on the basis of two properties discontinuity and similarity. Methods based on discontinuities are called as boundary based methods and methods based on similarity are called Region based methods Segmentation is a process that divides an image into its regions or objects that have similar features or characteristics. Mathematically complete segmentation of an image R is a finite set of regions R1…Rs,

$R = \bigcup_{i=1}^{S} R_i \qquad R_i \cap R_j = \emptyset \qquad i \neq j$

Image Segmentation Methods can be categorized as below:

1. Region Based Method
2. Edge Based Methods
3. Hybrid Techniques.

Region Based Techniques

Region based methods are based continuity. These techniques divide the entire image into sub regions depending on some rules like all the pixels in one region must have the same gray level. Region-based techniques rely on common patterns in intensity values within a cluster of neighboring pixels. The cluster is referred to as the region, and the goal of the segmentation algorithm is to group the regions according to their anatomical or functional roles.

Split and Merge Technique

There are two parts to this technique first the image is split depending on some criterion and then it is merged. The whole image is initially taken as a single region then some measure of internal similarity is computed using standard deviation. If too much variety occurs then the image is split into regions using thresholding. This is repeated until no more splits are further possible. Quad tree is a common data structure used for splitting. Then comes the merging phase, where two regions are merged if they are adjacent and similar. Merging is repeated until no more further merging is possible. The major advantage of this technique is guaranteed connected regions. Quad trees are widely used in Geographic information system. Kelkar D. and Gupta, S[3] have introduced an improved Quad tree method (IQM) for split and merge .In this improved method they have used three steps first splitting the image, second initializing neighbors list and the third step is merging splitted regions. They have divided the third step into two phases, in-house and final merge and have shown that this decomposition reduces problems involved in handling lengthy neighbor list during merging phase. The drawbacks of the split and merge technique are, the results depend on the position and orientation of the image, leads to blocky final segmentation and regular division leads to over segmentation (more regions) by splitting. This drawback can be overcome by reducing number of regions by using Normalized cuts method.

Normalized Cuts

This technique is proposed by Jianbo Shi and Jitendra Malik is mostly used in segmentation of medical images. This method is based on graph theory. Normalized cuts aim at splitting so that the division is optimal. Each pixel is a vertex in a graph, edges link adjacent pixels. Weights on the edge are assigned according to similarity between two corresponding pixels. The criterion for similarity is different in different applications. Similarity can be defined the distance, color, gray level, textures and so on. The advantage of this technique is that it removes the need to merge regions after splitting. It gives better definition around the edges Shi and Jitendra Malik [4], in their paper Normalized cuts and image segmentation shows how normalized cut is an unbiased measure of disassociation between subgroups of a graph and it has the nice property that minimizing normalized cut leads directly to maximizing the normalized association, which is an unbiased measure for total association within the subgroups. Wenchao Cai, JueWu, Albert C. S. Chung [5] improved the performance of the normalized cut by introducing the shape information. This method can correctly segment the object, even though a part of the boundary is missing or many noisy regions accompany the object. Thus there are various advantages of this method like it presents a new optimality criterion for partitioning a graph into clusters, different image features like intensity, color texture, contour continuity are treated in one uniform network. But there are certain disadvantages like lot of computational complexity involved especially for full-scale images .The performance and stability of the partitioning highly depends on the choice of the parameters.

Region Growing

Of the many proposed image segmentation methods, region growing has been one of the most popular methods. This method starts with a pixel and will go on adding the pixels based on similarity, to the region. When the growth of a region stops another seed pixel which does not belong to any other region is chosen, and again the process is started. The whole process is repeated until all pixels belong to some region. The advantage of this technique is, connected regions are guaranteed. Matei Mancas, Bernard Gosselin and Benoit Macq [4.] have used in their research a method which only needs one seed inside the region of interest (ROI). They have applied it for spinal cord segmentation but have found that it also shows results for parotid glands or even tumors. There are various applications where region growing techniques is mostly used like, to segment the parts of human body during treatment planning process e.g. segmentation of prostrate, bladder and rectum from contrast CT data. There are certain advantages of this technique like multiple criterions can be selected at the same time, gives very good results with less noisy images. But the various disadvantages of this technique are, if seeded region growing method is used then noise in the image can cause the seeds to be poorly placed, over segmentation may take place when the image is noisy or has intensity variations, cannot distinguish the shading of the real images, this method is power and time consuming.

Thresholding

This is the simplest way of segmentation. Using thresholding technique regions can be classified on the basis range values, which is applied to the intensity values of the image pixels. Thresholding is computationally inexpensive and fast, it is the oldest segmentation method and is still widely used in simple applications. Using range values or threshold values, pixels are classified using either of the thresholding techniques like global and local thresholding. Global thresholding method selects only one threshold value for the entire image. Local thresholding selects different threshold values for different regions. To segment complex images multilevel thresholding is required.

Edge Based Techniques

Segmentation Methods based on Discontinuity find for abrupt changes in the intensity value. These methods are called as Edge or Boundary based methods. Edge detection is the problem of fundamental importance in image analysis. Edge detection techniques are generally used for finding discontinuities in gray level images. Edge detection is the most common approach for detecting meaningful discontinuities in the gray level. Image segmentation methods for detecting discontinuities are boundary based methods Edge detection can be done using either of the following methods Edges are local changes in the image intensity. Edges typically occur on the boundary between two regions. Important features can be extracted from the edges of an image (e.g., corners, lines, curves).Edge detection is an important feature for image analysis. These features are used by higher-level computer vision algorithms (e.g., recognition). Edge detection is used for object detection which serves various applications like medical image processing, biometrics etc. Edge detection is an active area of research as it facilitates higher level image analysis. There are three different types of discontinuities in the grey level like point, line and edges. Spatial masks can be used to detect all the three types of discontinuities in an image.

There have been many image segmentation methods created and being created using many distinct approaches and algorithms but still it is very difficult to assess and compare the performance of these segmentation techniques. Researchers would evaluate their image segmentation techniques by using one or more of the following evaluation methods in Fig.2

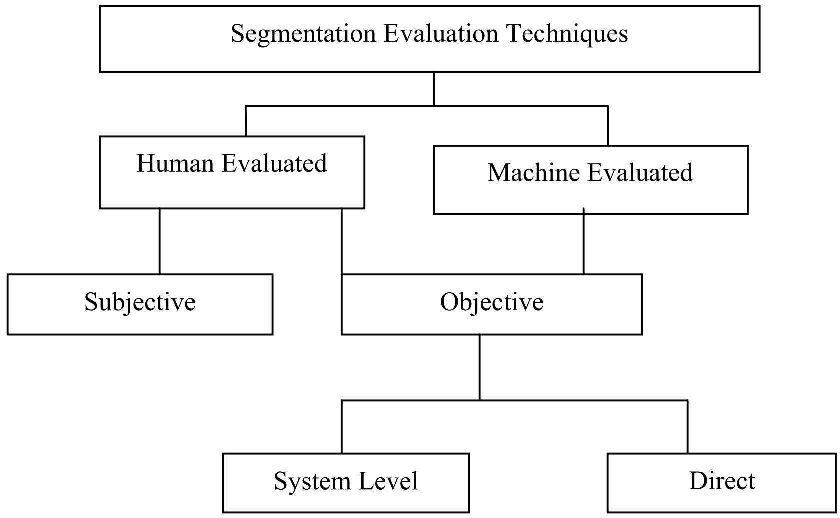


Fig. 2 Evaluation Methods

Clustering Technique

Given an image this methods splits them into K groups or clusters. The mean of each cluster is taken and then each point p is added to the cluster where the difference between the point and the mean is smallest. Since clustering works on hue estimates it is usually used in dividing a scene into different objects. The performance of clustering algorithm for image segmentation is highly sensitive to features used and types of objects in the image and hence generalization of this technique is difficult. Ali, Karmarkar and Dooley[2] presented a new shape-based image segmentation algorithm called fuzzy clustering for image segmentation using generic shape information (FCGS) which integrates generic shape information into the Gustafson-Kessel (GK) clustering framework. Hence using the algorithm presented in[2] can be used for many different object shapes and hence one framework can be used for different applications like medical imaging, security systems and any image processing application where arbitrary shaped object segmentation is required. But some clustering algorithms like K-means clustering doesn't guarantee continuous areas in the image, even if it does edges of these areas tend to be uneven, this is the major drawback which is overcome by split and merge technique

Clustering is an unsupervised learning task since it groups data objects into clusters without any prior information such as class labels. The clustering techniques, thus, should be able to deduce the structure embedded in data without any extra information. Clustering algorithms have been successfully applied in several fields such as information retrieval [1, 2], medicine [3], biology [4], customer analysis [5], image segmentation [6] and many others. Clustering has been an area of active research and many clustering algorithms have been proposed in the literature. The most widely used and the most popular algorithm for clustering is the k-means algorithm, proposed by J. MacQueen in 1967 [16]. K-means algorithm is fairly straightforward, simple to implement and has been employed by several researchers [18-20]. However, it may be easily trapped in a local optimum and fail to achieve a global optimum in several cases since the algorithm's performance is highly dependent on the initial centroids chosen.

To overcome this problem, several heuristic based approaches have been proposed for clustering. Selim and Alsultan [10] provided a simulated annealing (SA) algorithm for clustering in 1991. They have demonstrated that the simulated annealing algorithm converges to a global optimum for the clustering problem. Maulik and Bandyopadhyay [9] presented a clustering technique based on genetic algorithm, known as GA-clustering, in 2000. The centers of a pre-defined number of clusters were encoded using chromosomes and the improved performance of GA-clustering over k-means algorithm was demonstrated with the help of three real datasets. A tabu search based method was presented for solving the clustering problem in [21,22].

Shelokar et al. presented an Ant Colony Optimization (ACO) based technique for optimally assigning objects to a pre-defined number of clusters, in 2004 [12]. The ACO based technique provided very promising results when compared with other heuristic methods such as genetic algorithm simulated annealing and tabu search. Fathian et al. proposed an algorithm for clustering based on honeybee mating optimization (HBMO), in 2007 [13]. The performance of HBMO based approach was better compared to SA, GA, tabu search and ACO when evaluated over several well-known datasets. Ching-Yi et al. provided a Particle Swarm Optimization (PSO) based approach for clustering, in 2004 [11]. They compared the performance of PSO-based approach with traditional clustering algorithms and demonstrated that the PSO-based approach performed better using four artificial datasets.

Hatamlou et al. applied the Gravitational Search Algorithm (GSA) to data clustering, in 2011 [14]. The results over four well-known datasets depicted that GSA based approach performed better than several other clustering algorithms namely PSO, HBMO, ACO, GA, SA and k-means. In 2012, Hatamlou et al., presented a technique combining the benefits of k-means algorithm with GSA, called GSA-KM, in clustering [17]. In GSA-KM approach, the initial population for GSA was generated with the help of k-means algorithm which allowed GSA to converge faster. When compared with other well known algorithms, such as k-means, GA, SA, ACO, HBMO, PSO and the conventional GSA approach, GSA-KM approach provided better results over several real datasets.

Gravitational Search Algorithm

The Gravitational Search Algorithm (GSA) uses a constant value of parameter α for the calculation of gravitational constant. In the beginning, smaller value of α allows for a greater exploration of the search space. Furthermore, higher value of α during the last little iteration enhances the search space exploitation. Therefore, the approach based on GSA can be improved by adapting and controlling the value of parameter α as the algorithm proceeds.
Gravitational Search Algorithm (GSA) is an optimization algorithm proposed by Rashedi [8] in 2009. It is based on the Newton's laws of gravity and motion. The law of gravity states that "Every particle in the universe attracts every other particle with a force that is directly proportional to the product of the masses of the particles and inversely proportional to the square of the distance between them". By this definition, the gravitational force is determined using the following equation [8]:

$$F = G\,\frac{M_1 M_2}{R^2} \qquad (1)$$

where, F is the gravitational force acting between two masses $M_1$and M_2, G is the gravitational constant with a value of 6.67259 × 10-11 N m2/kg2, and R is the distance between the two masses.

Newton's second law of motion states that when a force acts on a mass, acceleration is produced. The magnitude of acceleration produced is obtained using the equation below [8]:

$$a = \frac{F}{M} \qquad (2)$$

Where, F and M denote the net force acting on a given particle and its mass, respectively.

The Gravitational Search Algorithm (GSA) employs this physical phenomenon for solving optimization problems. Consider a system with N masses or agents. The position of ith mass is defined as:

$X_i = (\, x_i^1, \ldots, x_i^d, \ldots, x_i^n)$, for i = 1, 2, . . . , N,

3)

Where, x_i^d is the position of ith agent in dth dimension and n is the total number of dimensions in the search space. The positions of agents correspond to the solutions of the problem. The mass of each agent is computed, after evaluating the present population's fitness, using the following equations:

$$m_i(t) = \frac{fit_i(t) - worst(t)}{best(t) - worst(t)} \quad 4)$$

$$M_i(t) = \frac{m_i(t)}{\sum_{j=1}^{N} m_j(t)} \quad 5)$$

Where, $fit_i(t)$, denotes the fitness value of ith agent at time t, and best(t) and worst(t) are computed as follows (for minimization problems):

$$best(t) = \min \; fit_j(t), \; j = 1, 2, \ldots, N \quad 6)$$

$$worst(t) = \max \; fit_j(t), \; j = 1, 2, \ldots, N \quad 7)$$

Similarly, for maximization problems best (t) and worst (t) are computed by taking the maximum and minimum fitness values respectively.

The acceleration of an agent is computed next, by considering the total forces from a set of heavier masses using the laws of gravity and motion using Equations 8 and 9. The new velocity of an agent is computed next by adding a fraction of its current velocity to its acceleration (Equation 10), followed by the calculation of its new position (Equation 11).

$$F_i^d(t) = \sum_{j \in kbest, j \neq i} rand_j \, G(t) \frac{M_j(t) M_i(t)}{R_{ij}(t) + \varepsilon} \left(x_j^d(t) - x_i^d(t)\right) \quad 8)$$

$$a_i^d(t) = \frac{F_i^d(t)}{M_i(t)} = \sum_{j \in kbest, j \neq i} rand_j \, G(t) \frac{M_j(t)}{R_{ij}(t) + \varepsilon} \left(x_j^d(t) - x_i^d(t)\right) \quad 9)$$

$$v_i^d(t+1) = rand_i \times v_i^d(t) + a_i^d(t) \quad 10)$$

$$x_i^d(t+1) = x_i^d(t) + v_i^d(t+1)$$

11)

where, $rand_i$ and $rand_j$ are two random numbers uniformly distributed in the range of [0, 1], ε is a small value to prevent division by zero, $R_{ij}(t)$ is the Euclidean distance between agent i and agent j. Kbest is the set of first K agents with best fitness values and thus, largest mass. Kbest is dependent on time, initialized to Ko at the start and decreases as time progresses. The gravitational constant, $G(t)$, decreases with time to control the search accuracy. The value of $G(t)$ is calculated using the following equation:

$$G(t) = G_o e^{\frac{-\alpha t}{T}} \tag{12}$$

Where, Go is the initial value of gravitational constant, α is a parameter which governs the degree of exploration versus exploitation of the search and T is the maximum number of iterations.

Chapter Two:

LITERATURE REVIEW

Image segmentation, which is an advancing research field in computer vision, refers to clustering an image into set of pixels possessing similar characteristics in terms of intensity values and geometrical features. Much effort has been done to achieve an efficient segmentation algorithm to be able to work in a wide range of images. Efficient algorithms have been used for image segmentation, can be categorized as follows. Thresh holding based techniques [5.] exploit certain values called threshold to discriminate between intensity values of objects in images. Active contours [5.] employ a set of primary contours which move towards the object boundaries according to the information derived from energy function. Besides, active contours have demonstrated good performance in segmentation of organs in medical images. Graph-based techniques define segmentation problem as a graph G = (V; E) in which each node virefers to a pixel in the image, and the edges in E connect certain pairs of neighboring pixels. Felzenszwalb proposed an efficient graph based image segmentation in [5.]. Boykov and Funkalea [5.] suggested graph cut algorithm for N-dimensional image segmentation.

One alternative solution for segmentation problems is to relate the concept of clustering to segmentation. Generally, clustering is assigning a set of n data points into a smaller number of c clusters. Besides, clustering techniques have been widely applied to the image segmentation in which among them K-means and fuzzy c-means attracted more popularity due to their high performances. Fuzzy c-means has proved to be the most effective clustering technique for image segmentation purposes. The algorithm employs an iterative procedure to minimize the objective function depending on the Euclidean distance of the pixels to the cluster centers. Since correlation of pixels in the immediate neighborhood is high, incorporation of spatial correlation into fuzzy c-means algorithm can improve the results of conventional fuzzy c-means. Hence, spatial fuzzy c-means [67] can be used as an alternative in such applications. Despite high performance of fuzzy c-means, it happens that the algorithm is unable to segment precisely. One solution is to incorporate evolutionary algorithms into the problem. Particle swarm optimization [6.]has been applied to the problem and generated promising results in terms of removing spurious segments.

Gravitational Search Algorithm (GSA) is a heuristic optimization method which was recently developed by Rashedi et al. [78] It is inspired by the Newton's rule of gravity and laws of motion. The algorithm exploits a group of agents which interact with each other by the gravity force, and laws of motion. Global movement is directed towards agents possessing heavier masses. They represent good solutions and on the other hand, are equivalent with slower motions to satisfy exploitation part of algorithm. Algorithm is converged to an optimum value in an iterative process. Moreover, being computationally effective motivates further application of this algorithm in new research fields.

Clustering or cluster analysis refers to the process of grouping a set of data objects such that objects belonging to the same group are similar, whereas those belonging to different groups are distinct. The final groups are called clusters or classes. It is a major data mining task and is used as a common technique for analysis of statistical data in many fields such as pattern recognition, machine learning, information retrieval etc. In the process of clustering, it is important to define an appropriate similarity or dissimilarity measure over which the data objects are to be clustered. The same set of objects may be partitioned into different groups depending on the choice of similarity or dissimilarity measure. The number of clusters in the final partition may be pre-assigned or may be considered as an internal parameter of the clustering algorithm to be deduced based on the input data.
In this paper, we incorporate the Fuzzy clustering into GSA algorithm means algorithm to find better data sets for image segmentation.

The Developed Fuzzy Inference System

The FIS is developed with two input variables and one output variable. The input variables are as follows:

IT: The current iteration number.
Fbest: The best value of fitness achieved till the current iteration.

IT enables us to consider how far we have reached in the search process. During the initial iterations, i.e. when *IT* is low, a lower value of α is desired since lower the value of α, higher the value of gravitational constant, $G(t)$, will be (Equation 12) and thus, higher the force, F, (Equation 8) resulting in a higher acceleration, a, (Equation 9) and velocity, $v(t)$ (Equation 10). This allows for higher exploration at the beginning of search. Similarly, towards the final few iterations, i.e. when *IT* is high, a higher value of α is desired to promote higher exploitation. Figure 2 depicts the membership function for *IT*. The iterations are represented as a fraction of the maximum number of iterations allowed, such that 0.5 means half of the total iterations and 1 represents the maximum iterations.

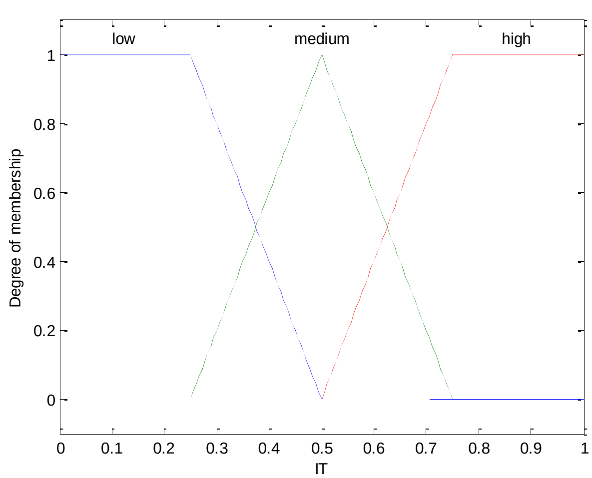


Fig 3: Membership Function for *IT*

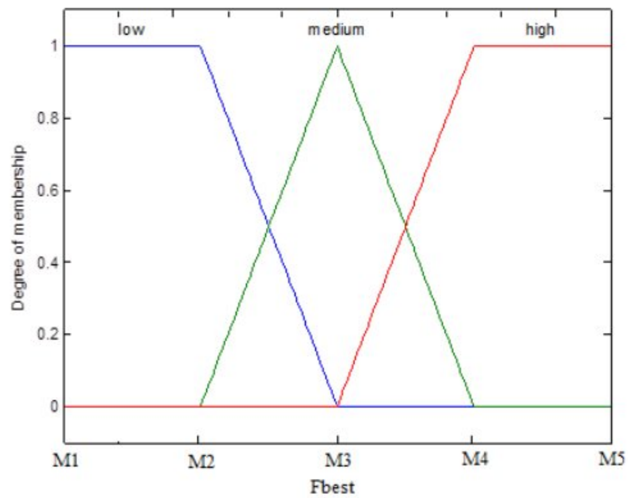


Fig 4: Membership Function for *Fbest*

Fbest represents the lowest value of fitness, since clustering is a minimization problem with the fitness function as mean square error, achieved till the current iteration. If the value of *Fbest* is high, then we need to reduce α to promote a greater exploration, since higher values for *Fbest* mean we are still far from the solution. However, if *Fbest* is low, we should increase α to allow for a higher exploitation as we are near the solution. Figure 3 shows the membership function for *Fbest*. Note that the membership function for *Fbest* needs to be tuned as per the input dataset being considered, since the acceptable values of fitness function will vary for different datasets.

To obtain the marked value M5 in Figure 3, we executed five independent runs of GSA for a single iteration and equated M5 to the maximum value of *Fbest* obtained, after adding hundred and then rounding it off to the nearest hundred. For the value of M3, we considered the integer part of the best fitness value obtained using GSA, taken from [14], for that dataset. M4 was computed by adding one to the value of M3, and M2 was calculated by rounding off M3 to the nearest ten smaller than M3. Finally, M1 was obtained by subtracting ten from M2.

Note that the fitness function, representing the total mean square error or the sum of intra-cluster distances, is computed using the following equation [23]:

$$f(O,C) = \sum_{l=1}^{k} \sum_{O_i \in C_l} d(O_i, CC_l)^2 \tag{13}$$

where, CC_l represents the centroids of the cluster C_l , $d(O_i, CC_l)$ denotes the distance or dissimilarity between object O_i and cluster centroid CC_l. The most popular and widely used distance metric is the Euclidean distance, which we have used in this work. Euclidean distance between two objects X_i and X_j with *d* dimensions is calculated as:

$$d(X_i, X_j) = \sqrt{\sum_{p=1}^{d} (x_i^p - x_j^p)^2} \tag{14}$$

where, x_i^p denotes the value of p^{th} dimension for the object X_i and x_j^p denotes the value of p^{th} dimension for the object X_j.

The developed FIS consists of one output variable, i.e. *alpha(t)*, which denotes the value of parameter α in Equation 12. Figure 4 shows the membership function for *alpha(t)*. The range of parameter α is taken as [0, 50] to provide a wide range of search on the value of *alpha(t)*.

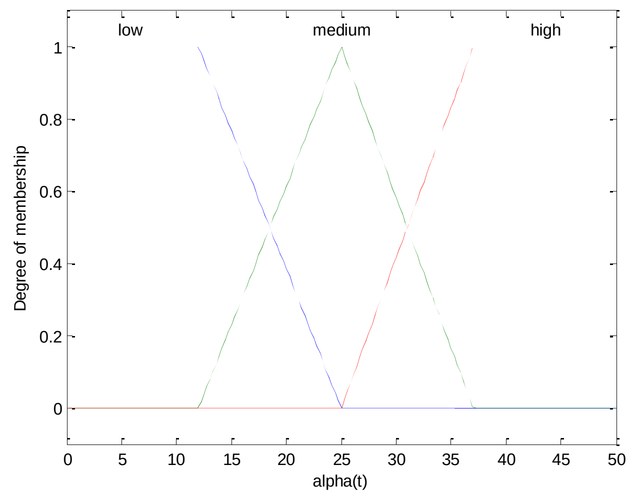


Fig 5: Membership Function for *alpha(t)*

The following eight fuzzy rules were formulated to control the parameter α in the calculation of the gravitational constant (Equation 12):

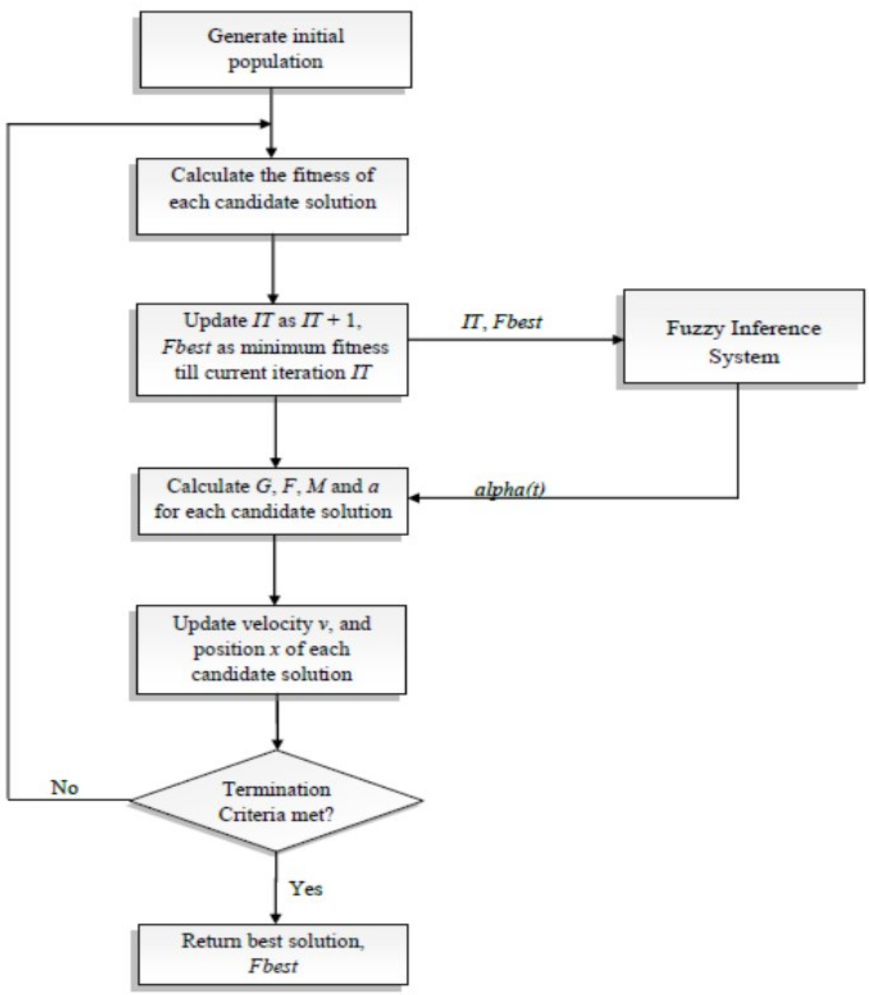
i. If (*IT* is low) and (*Fbest* is low) then (*alpha(t)* is high)
ii. If (*IT* is low) and (*Fbest* is medium) then (*alpha(t)* is medium)
iii. If (*IT* is low) and (*Fbest* is high) then (*alpha(t)* is low)
iv. If (*IT* is medium) and (*Fbest* is high) then (*alpha(t)* is low)
v. If (*IT* is medium) and (*Fbest* is medium) then (*alpha(t)* is medium)
vi. If (*IT* is high) and (*Fbest* is high) then (*alpha(t)* is medium)
vii. If (*IT* is high) and (*Fbest* is medium) then (*alpha(t)* is medium)
viii. If (*IT* is high) and (*Fbest* is low) then (*alpha(t)* is high)

The method used in the developed fuzzy inference system for "And" was min and for "Or" was max. The implication method was min, aggregation method was max and defuzzification method was centroid.

Chapter Three:

PROPOSED WORK

The proposed algorithm, Fuzzy-GSA, comprises of two main steps. The first step is to generate an initial population for GSA. We have generated the initial population by considering three agents (or candidate solutions) corresponding to the maximum, minimum and median values for all features in a given dataset, respectively. This provides a better initial population which would allow for a higher exploration since a wide range of values, including maximum, minimum and median, are present while searching the solution space. The rest of the agents are generated randomly by considering the range of features in the given dataset.
The second step involves application of GSA, described in Chapter 1, to the given dataset and using the fuzzy inference system developed to control the parameter α while searching for the solution. The flow diagram for the proposed Fuzzy-GSA algorithm is depicted by Figure 6.



The step by step algorithm for the proposed approach is stated next. Let N denote the population size, C_i be the i^{th} candidate solution or agent, k be the number of clusters, d be the number of features in a given dataset.

Step 1: Generate initial population, $P = \{C_1, C_2, \ldots, C_N\}$.

- Generate C_1 consisting of maximum values of all the features.
- Generate C_2 consisting of minimum values of all the features.
- Generate C_3 consisting of median values of all the features.
- Generate the remaining N-3 candidates randomly within the range of minimum to maximum values for all features.

Step 2: Apply GSA and use the developed FIS, for parameter adaptation.

- Calculate the fitness function, as per Equation 13, for all the candidate solutions.
- Feed the values of *IT*, current iteration number, and *Fbest*, best fitness achieved, as inputs to the developed FIS, and obtain the value of parameter α.
- Calculate G, F, M and a for all the candidate solutions using Equations 5, 8, 9 and 12 as described in the Gravitational Search Algorithm (GSA).
- Update the velocity and position of each candidate solution as per Equation 10 and 11 respectively.
- Check if termination criteria, i.e. maximum number of iterations allowed is reached or fitness function is not exhibiting a minimum improvement, are met. If yes, then return the best value of fitness function achieved as the final solution, else reiterate through step 2.

The final solution consists of the best value of fitness function, i.e. the minimum mean square error, achieved by running the proposed Fuzzy-GSA algorithm.

To verify the effectiveness (qualities and robustness) of the proposed technique, we conduct several experiments with this procedure on several images. There are some steps of our proposed technique are given below:

Phase 1: Firstly we develop a particular GUI for this implementation. After that we develop code for the loading dataset folder in MATLAB database. After that we calculate the feature of the dataset.

Phase 2: Develop a code for the loading the Query image in the MATLAB workspace. After this apply the feature extraction on this loaded query image.

Phase 3: Develop a code for the retrieving the images from the loaded dataset. For the retrieving purpose we use two techniques based on Texture and Colour.

Phase 4: After that we implement Fuzzy GSA Search Algorithm in our concept. In our proposed method Fuzzy Based Gravitational Search Algorithm is used for the optimization purpose.

Phase 5: After that we develop the code for the result in the MATLAB figure window. Our results are retrieved images.

Chapter Four:

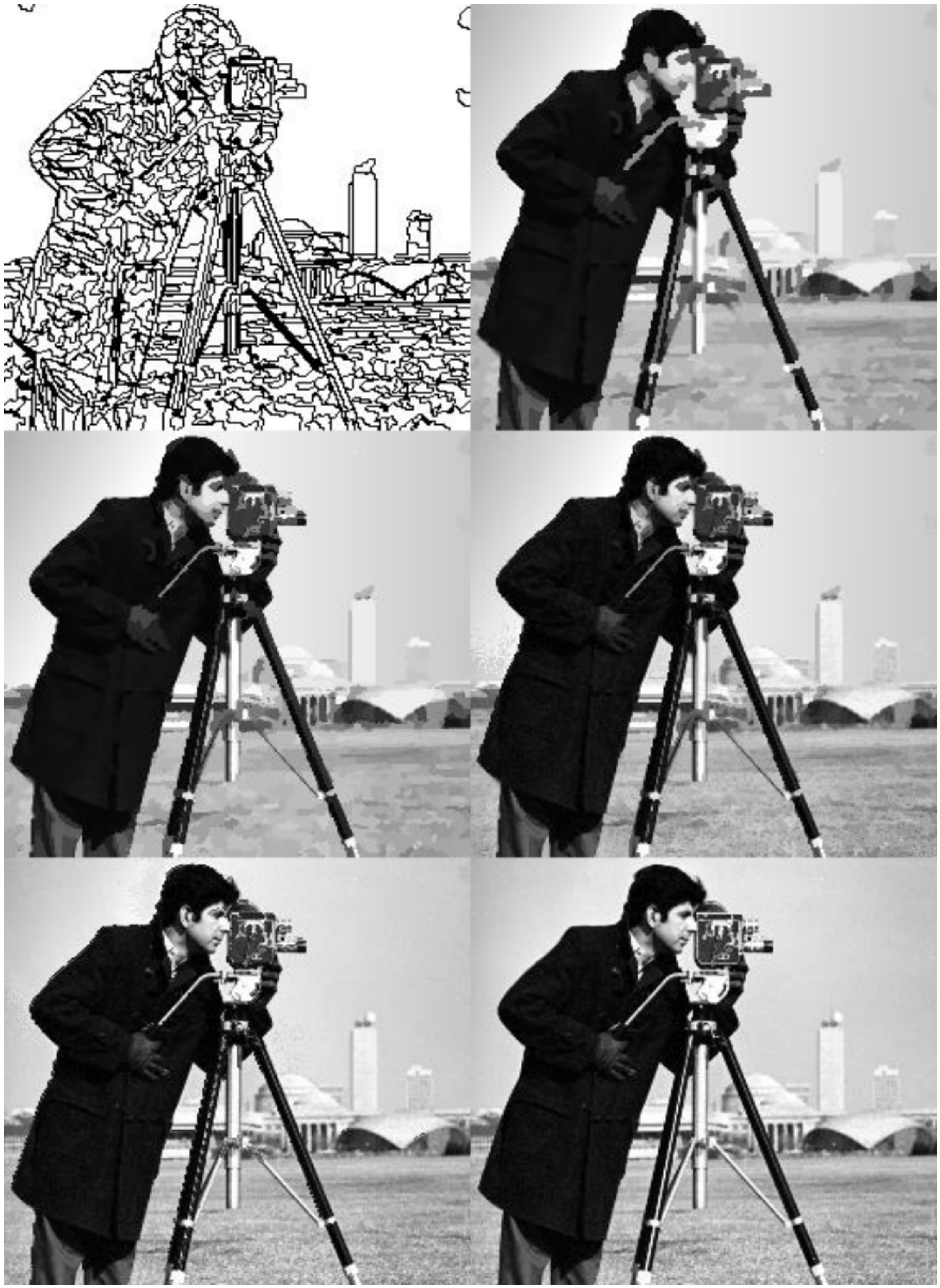
RESULTS

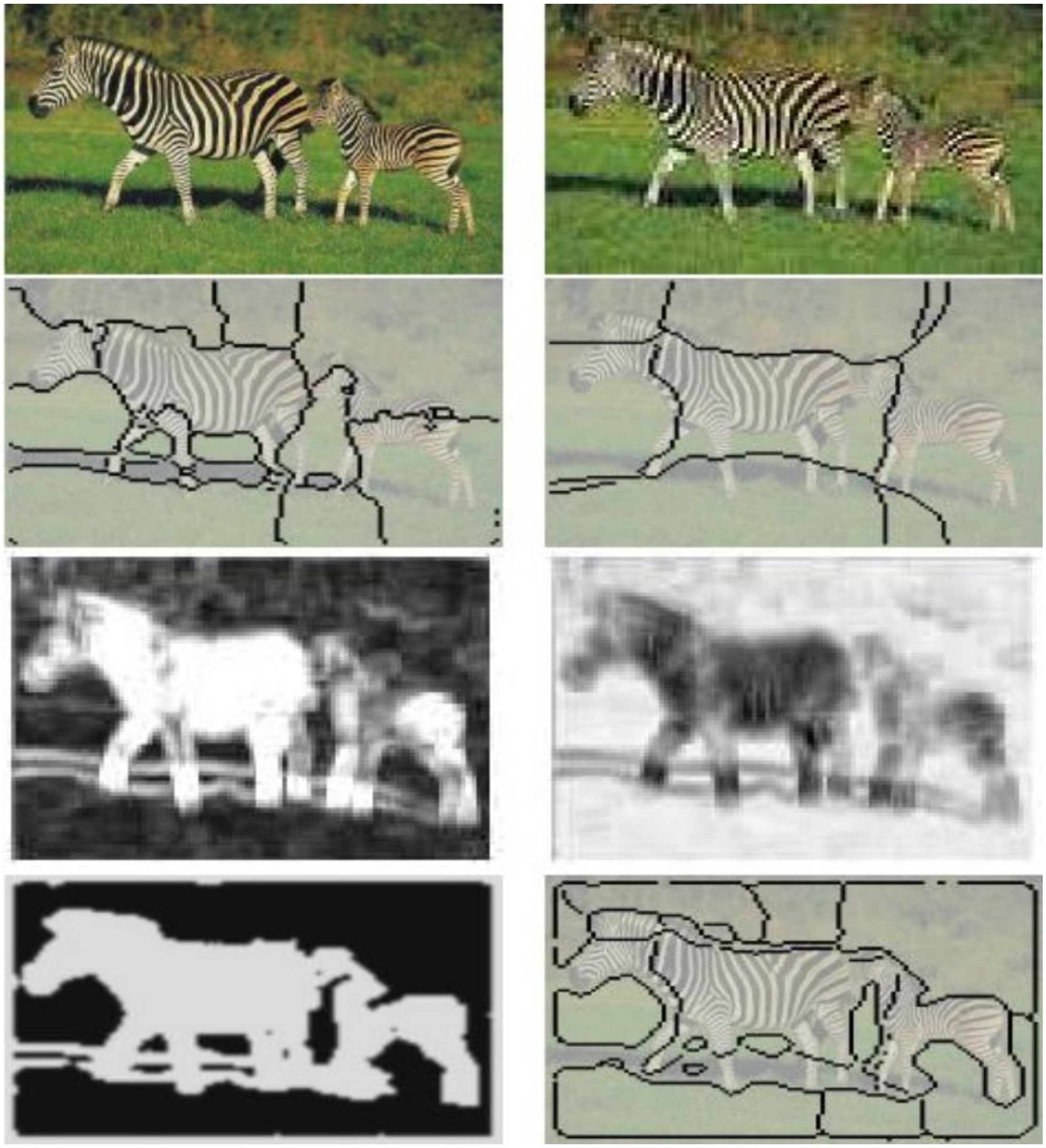
This Section provides information regarding results, interpretation of results and comparison of proposed content based image retrieval using Fuzzy GSA. We conduct several experiments with this procedure on several images.

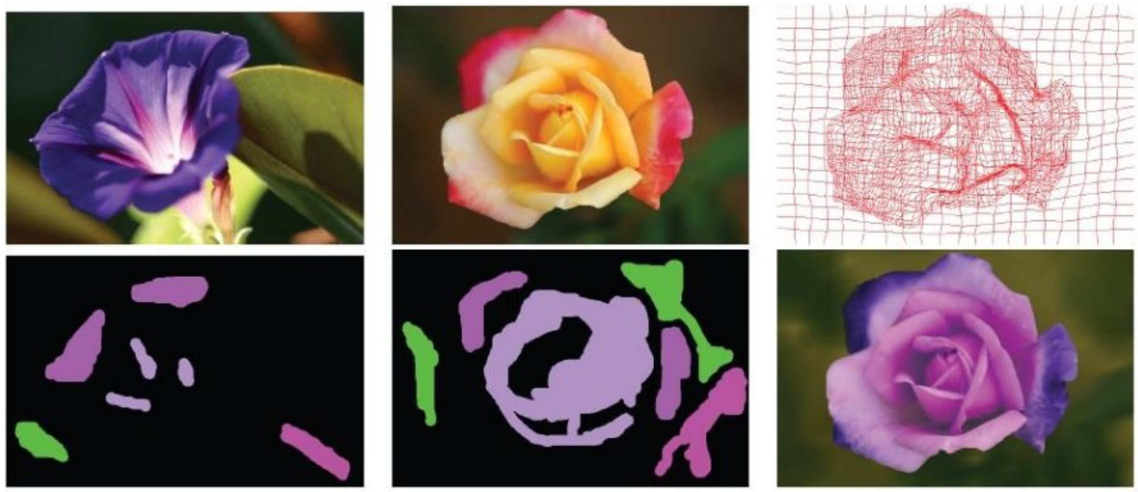
Fuzzy-GSA algorithm generates a better quality initial population for GSA by considering the nature of dataset being considered. It generates three candidate solutions consisting of maximum, minimum and median values, respectively in a dataset thereby building an initial population which covers a wider range. This helps in achieving a higher exploration.

The performance of Fuzzy-GSA is evaluated by comparing its best, average and worst solutions with several other clustering algorithms over four selected benchmark datasets namely, Iris, Wine, Breast Cancer Wisconsin and CMC, considering 20 independent runs.

The results show that Fuzzy-GSA achieves the highest quality clustering with very small standard deviation, when compared with several other clustering algorithms







Full Tone palette.
Colors slider set to maximum accuracy of 100, which results in 6463 colors for this tracing.

Limited palette.
Colors slider set to only eight colors in this tracing.

Document Library palette.
The color group is the "Brights" group of six colors shown in the color group above.

Grayscale Mode.
Maximum number of grays used (100 on slider).

(a) u_0 (b) $\sum_{i=0}^{1} u_i$ (c) $\sum_{i=0}^{2} u_i$

(f) v_0 (g) $\sum_{i=0}^{1} v_i$ (h) $\sum_{i=0}^{2} v_i$


JPG
80% quality, 14.7 KB
Truecolor PNG-24
16,777,216 colors, 53.6 KB
Indexed PNG-8
256 colors, 16.4 KB
Indexed PNG-8
128 colors, 12.9 KB
Indexed PNG-8
64 colors, 9.8 KB
Indexed PNG-8
32 colors, 7.6 KB
Indexed PNG-8
16 colors, 6 KB
Indexed PNG-8
8 colors, 4.6 KB
Indexed PNG-8
4 colors, 3.6 KB
Indexed PNG-8
2 colors, 1.9 KB

Chapter Five:

CONCLUSION & FUTURE WORK

This paper proposes an algorithm, Fuzzy-GSA, for clustering. Fuzzy-GSA algorithm is based on the conventional Gravitational Search Algorithm (GSA) with a provision for adapting the value of parameter α used in the calculation of the gravitational constant.
In the beginning, a smaller value of α is desired to achieve a higher exploration, whereas towards the end of search, a relatively higher value of α helps in achieving a higher exploitation. Fuzzy-GSA algorithm incorporates Fuzzy Inference System (FIS) into the conventional GSA to allow for parameter adaptation. The parameter α is controlled by using eight formulated fuzzy inference rules, in Fuzzy-GSA.

Also, Fuzzy-GSA algorithm generates a better quality initial population for GSA by considering the nature of dataset being considered. It generates three candidate solutions consisting of maximum, minimum and median values, respectively in a dataset thereby building an initial population which covers a wider range. This helps in achieving a higher exploration.
The performance of Fuzzy-GSA is evaluated by comparing its best, average and worst solutions with several other clustering algorithms over four selected benchmark datasets namely, Iris, Wine, Breast Cancer Wisconsin and CMC, considering 20 independent runs.

The results show that Fuzzy-GSA achieves the highest quality clustering with very small standard deviation, when compared with several other clustering algorithms.

Chapter Six:

REFERENCES

[1] N. Jardine, C.J. van Rijsbergen, “The use of hierarchic clustering in information retrieval”, Information Storage and Retrieval, vol. 7, issue 5, pp. 217–240, 1971.
[2] A. Tombros, R. Villa, C.J. van Rijsbergen, “The effectiveness of query-specific hierarchic clustering in information retrieval”, Information Processing & Management, vol. 38, issue 4, pp. 559–582, 2002.
[3] L. Liao, T. Lin, B. Li, “MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach”, Pattern Recognition Letters, vol. 29, issue 10, pp. 1580–1588, 2008.
[4] G. Kerr, H.J. Ruskin, M. Crane, P. Doolan, “Techniques for clustering gene expression data”, Computers in Biology and Medicine, vol. 38, issue 3, pp. 283–293, 2008.
[5] B. Saglam, M. Türkay, F.S. Salman, S. Sayın, F. Karaesmen, E.L. Örmeci, “A mixed-integer programming approach to the clustering problem with an application in customer segmentation”, European Journal of Operational Research, vol. 173, issue 3, pp. 866–879, 2006.
[6] Y. Xia, D. Fenga, T. Wangd, R. Zhaob, Y. Zhangb “Image segmentation by clustering of spatial patterns”, Pattern Recognition Letters, vol. 28, issue 12, pp. 1548–1555, 2007.
[7] John Yen and Reza Langari, Fuzzy Logic – Intelligence, Control and Information, Prentice Hall, 1st edition.
[8] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, “GSA: a gravitational search algorithm”, Information Sciences, vol. 179, issue 13, pp. 2232–2248, 2009.
[9] U. Maulik, S. Bandyopadhyay, “Genetic algorithm-based clustering technique”, Pattern Recognition, vol. 33, issue 9, pp. 1455–1465, 2000.
[10] S.Z. Selim, K. Alsultan, “A simulated annealing algorithm for the clustering problem”, Pattern Recognition, vol. 24, issue 10, pp. 1003–1008, 1991.
[11] C. Ching-Yi, Y. Fun, “Particle swarm optimization algorithm and its application to clustering analysis”, in IEEE International Conference on Networking, Sensing and Control, 2004.
[12] P.S. Shelokar, V.K. Jayaraman, B.D. Kulkarni, “An ant colony approach for clustering”, Analytica Chimica Acta, vol. 509, issue 2, pp. 187–195, 2004.
[13] M. Fathian, B. Amiri, A. Maroosi, “Application of honey-bee mating optimization algorithm on clustering”, Applied Mathematics and Computation, vol. 190, issue 2, pp. 1502–1513, 2007.

[14] A. Hatamlou, S. Abdullah, H. Nezamabadi-pour, “Application of gravitational search algorithm on data clustering”, Rough Sets and Knowledge Technology, Springer, pp. 337–346, 2011.
[15] C.L. Blake, C.J. Merz, UCI repository of machine learning databases. Available from: http://www.ics.uci.edu/-mlearn/MLRepository.html.
[16] J. MacQueen, “Some methods for classification and analysis of multivariate observations”, In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281-297, 1967.
[17] A. Hatamlou, S. Abdullah, H. Nezamabadi-pour, “A combined approach for clustering based on K-means and gravitational search algorithms”, Swarm and Evolutionary Computation, vol. 6, pp. 47–52, 2012.
[18] A.K. Jain, “Data clustering: 50 years beyond K-means”, Pattern Recognition Letters, vol. 31, issue 8, pp. 651–666, 2010.
[19] E.W. Forgy, “Cluster analysis of multivariate data: efficiency versus interpretability of classifications”, Biometrics, vol. 21, issue 2, 1965.
[20] L. Kaufman, P.J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis”, John Wiley & Sons, New York, 1990.
[21] K.S. Al-Sultan, “A tabu search approach to the clustering problem”, Pattern Recognition, vol. 28, issue 9, pp. 1443–1451, 1995.
[22] C.S. Sung, H.W. Jin, “A tabu-search-based heuristic for clustering”, Pattern Recognition, vol. 33, issue 5, pp. 849–858, 2000.
[23] S. Yang, R. Wu, M. Wang, L. Jiao, “Evolutionary clustering based vector quantization and SPIHT coding for image compression”, Pattern Recognition Letters, vol. 31, issue 13, pp. 1773–1780, 2010.
[24]https://www.researchgate.net/publication/222827594_Multiregion_competition_A_level_set_extension_of_region_competition_to_multiple_region_image_partitioning?el=1_x_8&enrichId=rgreq-3936a1df-6b2c-4f6c-99cf-0b7006027261&enrichSource=Y292ZXJQYWdlOzI1ODcxNDgz