HARMONY SEARCH-BASED FUZZY CLUSTERING ALGORITHMS FOR IMAGE SEGMENTATION

OSAMA MOH’D RADI ALIA

UNIVERSITI SAINS MALAYSIA

2011

HARMONY SEARCH-BASED FUZZY CLUSTERING ALGORITHMS FOR IMAGE SEGMENTATION

by

OSAMA MOH'D RADI ALIA

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

February 2011

ACKNOWLEDGEMENTS

IN THE NAME OF ALLAH THE ALL-COMPASSIONATE, ALL-MERCIFIFUL

All thanks are due to ALLAH, the most merciful, the most compassionate. For the countless blessings He has bestowed upon me.

Then, I would also like to express my deepest gratitude towards my thesis advisor, Associate Professor Dr. Mandava Rajeswari, for her continuous wisdom, encouragement, friendship, guidance and support over this course of my study. I am forever grateful for the chance to work in her research group.

Thank you also to Dr. Dhanesh Ramachandram for all the useful ideas he has given. His assistance and useful advice in publications-related issues are highly appreciated.

I would like to thank my co-supervisors, Associate Professor Dr. Mohd Ezane Aziz and Dr. Nor Azman Mat Zain of USM Hospital for spending much of their valuable time explaining the medical terminologies and issues related to malignant bone tumour “osteosarcoma”. These have been great sources of knowledge for guiding this work, especially in the automatic segmentation and labeling of different types of tumour cells. I also highly appreciate the comments and the many suggestions regarding the quality of this work, as well as how to make the developed tools more applicable in the setting of medical research and clinical use.

I would like to thank Prof. Ibrahim Lutifi Shuaib for his help and useful discussions, without which this thesis would not have been able to reach its completion. Not forgetting friends and labmates at the Computer Vision Research Group at Universiti Sains Malaysia, who have been a tremendous source of support and a lot of fun to work with. I will forever remember moments together. Thank you very-very much Mahmoud, Ahmad, Alfian, Mozaher, Ehsan, Idayu, and Anusha. Special thanks also to Ali Alghafri, Khalid Jaber, Mohammad Said Abu Al-Rub, Asef Al-khatib and Mohammad Al-Betar for their support.

Special thanks to School of Computer Sciences USM, for providing a conducive environment during the course of my research. The researcher is also deeply grateful to the Research Assistance Scheme at Universiti Sains Malaysia, USM Fellowship Scheme provided by the Institute of Graduate Studies (IPS), USM, and “Universiti Sains Malaysia Research University Grant” titled “Delineation and visualization of Tumour and Risk Structures - DVTRS” under grant number $1001/PKOMP/817001$, which supported my living cost during my study and research at the campus. I would like to thank the USM members for giving me the chance to gain my degree, and for providing me and the other students excellent facilities to conduct research.

Last but not least, may Allah have great compassion towards my parents’ souls. Gratitude beyond words to my beloved wife Noor; lovely sons Abdalrahman and Zain Alabideen; beautiful daughter Maria; brothers, Radi, Khaldon, and Adnan; and sister, Ameera, who have given constant love, support, encouragement, and patience.

Thank you all

LIST OF PUBLICATIONS

Journal Publications:

1. Alia O. M. and Mandava R. and Aziz M. E. (2011). A Hybrid Harmony Search Algorithm for MRI Brain Segmentation, DOI: 10.1007/s12065-011-0048-1, *Evolutionary Intelligence, Springer.*
2. Alia O. M. and Mandava R. (2011). The Variants of the Harmony Search Algorithm - An Overview , DOI: 10.1007/s10462-010-9201-y, *Artificial Intelligence Review, Springer.*

Conference Publications:

1. Alia O. M. and Mandava R. (2011). Data Clustering using Harmony Search Algorithm, *The 15th Online Conference on Soft Computing in Industrial Applications, (WSC15), 15-27 November 2010, Advances in Intelligent and Soft Computing, Springer Berlin/ Heidelberg.*
2. Mandava R. and Alia O. M. and Wei B. C. and Ramachandram D. and Aziz M. E. and Shuaib I. L. (2010). Osteosarcoma Segmentation in MRI using Dynamic Harmony Search Based Clustering, *International Conference of Soft Computing and Pattern Recognition, SOCPAR10, 7-10/12/2010, Cergy-Pontoise, France.*
3. Alia O. M. and Mandava R. and Aziz M. E. (2010). A Hybrid Harmony Search Algorithm to MRI Brain Segmentation, *The 9th IEEE International Conference on Cognitive Informatics, ICCI2010, pp. 712-719, Beijing, China.*
4. Alia O. M. and Mandava R. and Ramachandram D. and Aziz M. E. (2009). Dynamic fuzzy clustering using Harmony Search with application to image segmentation, *IEEE International Symposium on Signal Processing and Information Tech-*

nology, ISSPIT09, pp. 538-543, Ajman, UAE.

5. Alia O. M. and Mandava R. and Ramachandram D. and Aziz M. E. (2009). A Novel Image Segmentation Algorithm Based on Harmony Fuzzy Search Algorithm, *International Conference of Soft Computing and Pattern Recognition, SOCPAR09, pp. 335-340, Malacca, Malaysia.*

6. Alia O. M. and Mandava R. and Ramachandram D. and Aziz M. E. (2009). Harmony search-based cluster initialization for fuzzy c-means segmentation of MR images, *International Technical Conference of IEEE Region 10, TENCON, pp. 1-6, Singapore.*

TABLE OF CONTENTS

CHAPTER 1 – INTRODUCTION

CHAPTER 2 – THEORETICAL BACKGROUND

CHAPTER 3 – LITERATURE REVIEW

CHAPTER 4 – METHODOLOGY

CHAPTER 5 – A HARMONY SEARCH-BASED FUZZY CLUSTERING ALGORITHMS

CHAPTER 6 – DYNAMIC CLUSTERING BASED ON HARMONY SEARCH ALGORITHM

CHAPTER 7 – APPLICATIONS

CHAPTER 8 – CONCLUSION

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
ABC	Artificial Bee Colony
ANN	Artificial Neural Network
ACO	Ant Colony Optimization
bw	bandwidth parameter
CSF	CerebroSpinal Fluid
CT	Computed Tomography
DCHS	Dynamic Fuzzy Clustering using the Harmony Search
DE	Differential Evolution
DICOM	Digital Imaging and Communications in Medicine
EA	Evolutionary Algorithm
EM	Expectation Maximization Algorithm
EOR	Empty Operator Rate
EP	Evolutionary Programming
ES	Evolutionary Strategies
FLAIR	Fluid Attenuation Inversion Recovery
FCM	Fuzzy C-Means Algorithm

FCMR	Fuzzy C-Means Rate
GA	Genetic Algorithm
GM	Gray Matter
HFCM	Harmony Fuzzy C-Means Algorithm
HFISA	Harmony Fuzzy Image Segmentation Algorithm
HM	Harmony Memory
HMCR	Harmony Memory Considering Rate
HMS	Harmony Memory Size
HS	Harmony Search Algorithm
MRI	Magnetic Resonance Imaging
MSLs	Multiple Sclerosis Lesions
NI	Number of Iterations
NP	Non-deterministic Polynomial-time
PAR	Pitch Adjusting Rate
PC	Partition Coefficient Index
PD	Proton Density Image
PE	Partition Entropy Index
PSO	Particle Swarm Optimization
ROI	Region Of Interest
SA	Simulated Annealing

T1WI	T1-Weighted Image
T2WI	T2-Weighted Image
SOM	Self Organizing Maps
STIR	Short Tau Inversion Recovery
TS	Tabu Search
WM	White Matter
XB	Xie-Beni Index

ALGORITMA-ALGORITMA PENGKELOMPOKAN KABUR BERASASKAN CARIAN-HARMONI UNTUK SEGMENTASI IMEJ

ABSTRAK

Algoritma-algoritma pengkelompokan kabur, yang tergolong di dalam kategori pembelajaran mesin tanpa selia, adalah di antara kaedah segmentasi imej yang paling berjaya. Namun demikian, terdapat dua isu utama yang membataskan keberkesanan kaedah ini: kepekaan terhadap pemilihan pusat kelompok permulaan dan ketidakpastian terhadap bilangan kelompok sebenar di dalam set data. Tesis ini bermatlamat untuk menyelesaikan masalah-masalah ini dengan mengunakan algoritma metaheuristik eficien yang dikenali sebagai algoritma Carian Harmoni (HS). Pertama, dua kaedah alternatif pengkelompokan kabur berasaskan HS dicadangkan. Tujuan kedua kaedah ini adalah untuk mengatasi kelemahan algoritma pengkelompokan kabur kovensional yang boleh menghasilkan kelompok suboptimum bergantung kepada pemilihan kelompok permulaan. Kedua, algoritma pengkelompokan kabur berasaskan HS dinamik (DCHS) baharu dicadangkan untuk menganggar bilangan kelompok secara automatik serta memperoleh pembahagian kabur yang baik bagi set data yang digunakan. Kesemua algoritma baharu ini telah diaplikasikan kepada permasalahan segmentasi imej. Pelbagai imej dari domain aplikasi berbeza, termasuk imej dunia sebenar dan sintetik, telah digunakan di dalam tesis ini untuk menunjukkan kebolehgunaan algoritma-algoritma yang dicadangkan. Akhir sekali, DCHS diaplikasikan ke atas dua permasalahan imej perubatan, iaitu segmentasi tumour tulang malignan (osteosarkoma) dan MRI Otak. Hasil kajian menunjukkan kemajuan memberangsangkan berbanding pendekatan lain di dalam domain yang sama.

HARMONY SEARCH-BASED FUZZY CLUSTERING ALGORITHMS FOR IMAGE SEGMENTATION

ABSTRACT

Fuzzy clustering algorithms, which fall under unsupervised machine learning, are among the most successful methods for image segmentation. However, two main issues plague these clustering algorithms: initialization sensitivity of cluster centers and unknown number of actual clusters in the given dataset. This thesis aims to solve these problems using an efficient metaheuristic algorithm, known as the Harmony Search (HS) algorithm. First, two alternative HS-based fuzzy clustering methods are proposed. The aim of these methods is to overcome the limitation faced by conventional fuzzy clustering algorithms, which are known to provide sub-optimal clustering depending on the choice of the initial clusters. Second, a new dynamic HS-based fuzzy clustering algorithm (DCHS) is proposed to automatically estimate the appropriate number of clusters as well as a good fuzzy partitioning of the given dataset. These algorithms have been applied to the problem of image segmentation. Various images from different application domains, including synthetic and real-world images, have been used in this thesis to show the applicability of the proposed algorithms. Finally, the proposed DCHS algorithm is applied to two real-world medical image problems, namely, malignant bone tumour (osteosarcoma) and magnetic resonance imaging brain segmentation. The experimental results are very promising showing significant improvements compared to other approaches in the same domain.

CHAPTER 1

INTRODUCTION

1.1 Data clustering

Data clustering is a technique used to discover patterns and associations within data. It is a multivariate statistical procedure that allows the user to handle large volumes of data and attempts to reorganize these data into relatively homogeneous groups. These groups should allow the user to deal with and utilize the original volume of data more effectively. Accuracy of clustering is essential because it would be counter-productive if the compact form of the data does not accurately represent the original data (Shihab, 2000).

Here, data are a set of points or patterns usually represented as vectors of measurements, features, or points in a multidimensional space. These collected features are combined into a list, which then acts as the input to a chosen computational clustering algorithm. This algorithm then provides a description of the grouping structure that it has discovered within the patterns. The grouping structure, clustering process, is based on some similarity or distance measurements that allocate the given patterns into predefined clusters (classes). Intuitively, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster.

Data clustering is known as an unsupervised classification technique different from the other classification technique known as supervised classification. The main difference between the two is that in the latter, a collection of labeled (pre-classified) patterns is provided, therefore, the problem is to label a newly encountered, yet unlabeled, pattern. In the case of clustering, the problem is to group a given collection of unlabeled patterns into meaningful clusters. In other words, no labeled patterns are provided, therefore, the labels associated with clusters are data driven; that is, they are obtained solely from the data (Jain et al., 1999).

Clustering has many useful characteristics, which has made it one of the most popular machine learning algorithms in many domains such as machine learning, artificial intelligence, pattern recognition, web mining, data mining, biology, remote sensing, marketing, image segmentation, etc. (Jain et al., 1999; Brian et al., 2009). During the last several decades, clustering algorithms have proved reliable, especially in categorization tasks that call for semi or full automation (Hore et al., 2008; Zhou and Schaefer, 2009).

1.2 Clustering Approaches

Generally, clustering is a typical unsupervised learning technique used to group similar data points according to some measurement of similarity. This measurement will seek to minimize inter-cluster similarity while maximizing intra-cluster similarity (Jain et al., 1999).

Clustering algorithms can generally be categorized into two groups: hierarchical, and partitional (Jain et al., 1999). The former produces a nested series of partitions, whereas the latter does clustering with one partitioning result. According to Jain et al. (2000) partitional clustering is more popular in pattern recognition applications because it does not suffer from drawbacks such as static-behavior (i.e., data points assigned to a cluster cannot move to another cluster), and the probability of failing to separate overlapping clusters, problems that are prevalent in hierarchical clustering. Partitional clustering can further be divided into two: 1) crisp (or hard) clustering, where each data point belongs to only one cluster, and 2) fuzzy (or soft) clustering, where data points can simultaneously belong to more than one cluster at the same time, based on some fuzzy membership grade. Fuzzy is considered more appropriate than crisp clustering for datasets that exhibit unclear boundaries between clusters or regions (Hore et al., 2008; Kang et al., 2009; Zhou and Schaefer, 2009). The fuzzy clustering approach and its characteristics are discussed later in this thesis.

1.3 Clustering-based Image Segmentation Approach

Image segmentation is the task of subdividing the image into constituent regions, in which each region shares similar feature properties. Image segmentation is thus considered a core process and is one of the most challenging tasks in any computer vision system (Gonzalez and Woods, 2008). Image segmentation plays a major role in many different domains. In medical imaging for instance, image segmentation techniques can assist doctors and radiologists locate tumours and other pathologies, measure tissue volume, diagnose illnesses, aid in computer-guided surgery, treatment planning, surgical simulation, therapy evaluation, and also study anatomical structure. In the pattern recognition domain, image segmentation is used to isolate regions of interest (ROIs) from images containing characters, fingerprints, signatures, faces, and gestures. In remote sensing, regions such as roads, buildings, rivers, etc. can be identified using image segmentation. In the manufacturing industry, those inspecting and assembling manufacturing products also benefit from image segmentation. The foregoing are but few examples from a much larger number of possible applications of image segmentation (Sergios and Konstantinos, 2008).

Such diversities in applications have thus spurred interest from digital image processing experts to develop advanced algorithms to improve segmentation results within a given domain. This is necessary, as different domains not only deal with different types of images, but also demonstrate different image properties. Furthermore, there is the issue of image complexity, which pertains to the amount of subjective information contained in images. Using one algorithm in different areas simply cannot get the job done (Zhang et al., 2008). Thus, many algorithms have been proposed over the last several decades, each of which uses different induction principles (Pal and Pal, 1993; Pham et al., 2000). These algorithms can be categorized into various groups such as thresholding-based, deformable models-based, clustering-based, histograms-based, classification-based, etc. (Pal and Pal, 1993; Pham et al., 2000).

Among these algorithms, fuzzy clustering-based segmentation methods are of considerable benefit because most images exhibit unclear boundaries between their regions. In this context, fuzzy clustering has shown tremendous potential, as it can naturally cope with such data characteristics. It is therefore not surprising that the fuzzy clustering algorithms represented by the fuzzy c-means (FCM) algorithm (Bezdek, 1981), is the most widely used algorithm in numerous image applications (Hore et al., 2008; Kang et al., 2009). Both image segmentation and clustering share the same goal of finding accurate classification of their input. Clustering algorithms consider the image pixels as patterns and each pixel is assigned to a cluster (image region) based on some feature similarity (Rosenfeld and Kak, 1982). In this thesis, the application of the proposed fuzzy clustering algorithms is invested in solving the problem of image segmentation as can be seen in Chapters 5 and 6. In Chapter 7, we describe a specific application of the proposed fuzzy clustering algorithms to two difficult real-world medical image problems.

1.4 Metaheuristic-based Clustering

Metaheuristic algorithms are well-known approximate algorithms that can solve optimization problems with satisfactory results (Blum and Roli, 2003, 2008). They can be defined as " …high level strategies for exploring search spaces by using different methods " (Blum and Roli, 2003). Metaheuristics thus came forth to overcome the major drawback of well-known approximate algorithms, local search algorithms, that may stop at a very poor quality local optima. Metaheuristic is a general heuristic method applicable to a wide range of different optimization problems. They can be categorized into two classes: local search-based metaheuristic and population-based metaheuristic, where the former is based on evolving a single solution, while the latter is based on a population of solutions. The population-based metaheuristic approach has some advantages over the local search-based metaheuristic approach as seen in Appendix A. The main advantages of population-based metaheuristic algorithms are their abilities to cope with local optima and explore large solution spaces effectively by maintaining, recombining, and comparing several candidate solutions simultaneously. Many are inspired by natural phenomena as in the case of Particle Swarm Optimization (PSO), Simulated Annealing (SA), Genetic Algorithm (GA), and Harmony Search (HS) algorithm. These algorithms are intelligently inspired by natural phenomena to provide efficient solution techniques to yield high-quality solutions in a reasonable time.

HS is a relatively new metaheuristic algorithm developed by Geem et al. (2001) to solve optimization problems. Ever since the emergence of this algorithm, it has been able to attract many researchers to develop HS-based applications in many optimization problems (see (Ingram and Zhang, 2009) and references therein). HS imitates the natural phenomenon of musicians' behaviors when they collectively tune the pitches of their instruments together to achieve a fantastic harmony as measured by aesthetic standards. It is a successful metaheuristic algorithm that can explore the search space of the given dataset in a parallel optimization environment, where each solution (harmony) vector is generated by intelligently exploring and exploiting a search space. It is thus considered a population-based algorithm with local search-based aspects (Geem, 2009b). This feature distinguishes HS, along with other features such as (1) the generation of a new vector after considering all existing vectors, rather than considering only two vectors as in GA (parents); (2) the independent consideration for each decision variable in harmony memory vector; (3) the consideration of continuous decision variable values without any loss of precision; (4) no requirement of decimal-binary conversions or a fixed number (2n) of decision variable values as in GA; and (5) no need for any starting values of the decision variables or require complex derivatives as in gradient-based methods. These make it a preferable technique not only as standalone algorithm but also when combined with other metaheuristic algorithms

Due to the Non-deterministic Polynomial-time hard ,known as NP-hard, nature of partitional clustering methods, where the minimization of an objective function (e.g., the sum-of-squared errors) is the principle of these methods, clustering problems can be classified as optimization problems (Falkenauer, 1998). Therefore, metaheuristic algorithms are widely believed to be used as a clustering algorithm. This is based on the ability of such algorithms to solve NP-hard problems with satisfactory near-optimal solutions and significantly less computational time compared with exact algorithms (Hruschka et al., 2009).

1.5 Mapping between Clustering, Segmentation, and Optimization

For an overview of the relationship between these three domains, image segmentation, clustering, and optimization, it can be said that image segmentation can be considered a clustering problem and the clustering problem can regarded as an optimization problem. The mapping between these domains is described as follows.

Image segmentation-clustering mapping: Simply, each pixel in an image can be mapped as a pattern in the clustering domain, while image regions are also mapped to be clusters or classes. Furthermore, the concept of both domains is the same since their goal is to find the accurate classification of their input.

Clustering-optimization mapping: Both approaches share the same goal of selecting best elements from sets of available alternatives. In optimization, these elements are called decision variables while in clustering they could be called cluster centers (centroid-based), cluster labels (label-based), or cluster medoids (medoid-based). This process of selecting best elements from sets of available alternatives can be achieved by minimizing or maximizing the objective function. Both clustering and optimization share the same goal in terms of minimizing or maximizing of an objective function to reach the best elements.

Generally, in each iteration of the optimization process and based on its evolving process, a new solution vector is generated in which its decision variables are values that represent cluster centers (which represent an image's regions), then a reallocation of each pattern (i.e., pixel) to the nearest region with membership degree (i.e., fuzzy membership $\in [0,1]$) is performed. Once this has been done, a calculation to such solution in terms of objective function is performed, and a judgment on its value makes it accepted or rejected. This process is repeated until the stopping criterion is met. At the end, the optimal clusters with their pixel members represent the segmented image.

After this mapping, a conclusion can be drawn that optimization algorithms such as metaheuristic algorithms are suitable for clustering and image segmentation problems. However, one question remains to be answered namely why an optimization approach, especially metaheuristic, needs to be used. To answer this, it must be admitted that the current approach of clustering algorithms has some weaknesses that plague their performance, and the main solution for such weaknesses, as will be seen in the following section, is to use the metaheuristic approach as a clustering approach.

1.6 Motivation of the Research - Problem Statement

Despite their strengths and popularity as algorithms of choice for clustering purposes, partitional clustering algorithms suffer from some drawbacks. Among the main issues are the following:

1. It is sensitive to the cluster centers initialization step, therefore the tendency to be trapped in local optima is very high.
2. Required prior knowledge of number of clusters for a given dataset.
3. Sensitivity to noise and outliers.

This thesis intends to shed light on two weaknesses present in partitional clustering algorithms, namely 1 and 2, as previously mentioned. However, the proposed algorithms are tested against noise and outliers. A description is provided for both problems as follows:

1.6.1 Cluster Centers Initialization Sensitivity - Local Optima Problem

Selecting the initial cluster centers is considered one of the most challenging tasks in clustering algorithms. Generally, clustering algorithms seek to minimize an objective function, although it is unfortunately guaranteed only to yield local minima (Bezdek et al., 1987; Hathaway and Bezdek, 1986; Selim and Ismail, 1984). Improper selection of initial cluster centers will lead the searching process toward an optimal solution that stays in local optima, and therefore produces an undesirable clustering result. The main cause for this local optimal problem is when search algorithms work in a similar fashion to a hill climbing algorithm (Kanade and Hall, 2007). The hill climbing algorithm is a local search-based algorithm that moves in one direction without performing a wider scan of the search space to minimize (or maximize) the objective function. This behavior prevents the algorithm from exploring other regions in the search space that might have a better, or even the desired solution. Since hill climbing-like algorithms are only guaranteed local optimal solutions, consequently, the same initial cluster centers in a dataset will always generate the same cluster results, and better results might as well be obtained if the algorithm is run with different initial cluster centers.

1.6.2 Prior Knowledge of the Number of Clusters

Most existing partitional clustering techniques (including crisp and fuzzy), are manually supplied with the number of classes (clusters), instead of automatically being determined during execution. In many real-world data, the appropriate number of clusters is normally unknown or even difficult to be approximated subjectively (Hruschka et al., 2009; Maulik, 2009; Das, Abraham, Chakraborty and Konar, 2009; Campello et al., 2009). For example, in automatic medical diagnostic systems (e.g., to detect areas of pathological cells in the brain, bone, or breast), such systems should be able to automatically identify the different types of human cells and tissues (such as valuable cells, necrotic cells, edema cells, normal cells, and other body tissues) in each image. However, the number of tissue types is unknown in each image, since these images might be scanned at different positions of the human body, which may or may not have these types of tissues. Furthermore, these systems normally deal with a huge number of images generated from one or more medical modalities such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT). Moreover, different sequences of images are available from these imaging modalities such as T1-Weighted Image (T1WI), T2-Weighted Image (T2WI), Short Tau Inversion Recovery (STIR), Fluid Attenuation Inversion Recovery (FLAIR), etc., where each sequence provides different types of information for the tissues under study. For natural images, the situation may be more difficult since the actual number of regions is normally huge and uncertain. Therefore, depending on the aforementioned, segmentation based on automatic determination of appropriate number of clusters is required and desired. yet it is by no means an easy task to do.

In the spirit of the main cause of the initialization sensitivity problem, the local search behavior, several global-based or improved local-based search algorithms have been proposed in the last few decades to address this problem such as (Kanade and Hall, 2007; Selim and Alsultan, 1991; Al-Sultan and Selim, 1993; Al-Sultan and Fedjki, 1997; Bezdek et al., 1994; Hall et al., 1999; Yong-Guo et al., 2004; Pham et al., 2007; Lili et al., 2007; Mahdavi et al., 2008; Maulik and Saha, 2009). The main advantages of these algorithms are their abilities to cope with local optima and explore large solution spaces effectively by maintaining, recombining, and comparing several candidate solutions simultaneously. These solutions can be categorized into two: 1) using the metaheuristic algorithm to find the appropriate cluster center values, and then using these values as initial values for the clustering algorithms as can be seen in (Hall et al., 1999; Pham et al., 2007; Kanade and Hall, 2007), and 2) using a metaheuristic as a clustering algorithm such as (Mahdavi et al., 2008; Maulik and Saha, 2009). Having thoroughly scrutinized these algorithms and their development, one can say that the competition to improve their performance to solve the clustering problem was the goal behind such development. These advancements, it should be emphasized, were based on improving the natural behavior of the optimization process of these algorithms. For instance, balancing between the exploration and exploitation strategies of metaheuristic algorithms is one way to improve metaheuristic-based clustering algorithms. Using various evolving techniques such as encoding schemes (i.e., real or binary), selection mechanism or crossover and mutation operators is another way of modifying these algorithms to improve their performance and accuracy. Experimenting with new metaheuristics has also contributed to further improvement in the field. However, despite the promising results shown by these algorithms, and since there is no exact solution in such category, the approximation algorithms, it is desirable to develop a new metaheuristic-based algorithm that can improve the performance even further.

For the second clustering problem on prior knowledge of number of clusters, metaheuristic-based clustering algorithms were proposed in the literature under the name of the dynamic clustering approach. Such approach can automatically determine the appropriate number of clusters as well as a good clustering of the given dataset. However, little effort has been made in the direction of this approach during the last several years (Hruschka et al., 2009), as can be seen in the literature review chapter. The same concepts behind such diversity of metaheuristic algorithms are used in this approach as mentioned earlier for the first clustering problem. Despite the promising results shown by these algorithms, it is desirable to develop a new metaheuristic-based dynamic clustering algorithm that can improve the performance even further and explore the ability of the new algorithm to provide satisfactory solutions.

The main motivation for this thesis is to improve the clustering performance and overcome its weaknesses and thus, improve all related clustering-based applications such as image segmentation. Furthermore, exploring the ability of the new metaheuristic, HS algorithm, to solve such problems is investigated. These motivations are summarized in the following section.

1.7 Research Objective

The primary objectives of this thesis can be summarized as follows.

1. To show that HS algorithm can be successfully used to solve difficult problems in image segmentation domain.
2. To develop an HS-based fuzzy partitional clustering algorithm that will overcomes the limitation of the partitional clustering algorithms.
3. To develop an efficient HS-based algorithm that can automatically predict the appropriate number of clusters as well as improve the clustering performance of the fuzzy clustering algorithms.
4. To validate the developed algorithms with benchmarked datasets and then applying them to address real-world medical problems including:

- Automatic MRI brain image segmentation.
- Automatic MRI malignant bone tumour known as osteosarcoma.

1.8 Scope

The scope of the presented work is defined as follows:

1. This study is scoped to fuzzy partitional clustering approach.
2. The application of the proposed clustering algorithms is scoped to the image segmentation problem.
3. The image type used in this thesis is limited to gray scale.

1.9 Overview of Methodology

In this thesis, new fuzzy clustering algorithms based on HS algorithm are proposed. These new proposed algorithms take advantage of the inherited features of HS to avoid local optimal and come up with an appropriate assumption of number of clusters for the test images. The application of these new algorithms to the problem of image segmentation is investigated. Experimental results are then obtained using various synthetic images with well-known characteristics and natural images from different areas such as medical images and remotely sensed satellite images are also used to show the wide applicability of the proposed approaches. The results of state-of-the-art algorithms when applied to the same test images are also reported to show the relative performance of the proposed approaches compared with other well-known approaches. Due to the stochastic nature of the proposed algorithms, all presented results are averages and standard deviations over several simulations.

1.10 Importance of the Study

This study is particularly important as it is closely connected with several major related topics. Such topics, as mentioned earlier, are image analysis and pattern recognition together with their applications. Image segmentation is considered the core of such applications, and the degree of success mainly depends on the segmentation results. Interestingly, in many of these applications, partitional clustering algorithms seem to be the popular choice for image segmentation (Kang et al., 2009; Zhou, 2009; Zhou and Schaefer, 2009; Xiaohe et al., 2008; Zhou and Rajapakse, 2008). However, due to the shortcomings of partitional clustering algorithms, as mentioned in Section 1.6, results are still unsatisfactory.

As a result of these shortcomings, the author has been motivated to investigate ways to improve this particular clustering category for image segmentation. This research thus intends to propose novel clustering algorithms that can overcome the shortcomings of partitional clustering. These algorithms will be beneficial as they can improve the accuracy of segmentation which will,in turn, increase the performance of image analysis and pattern recognition applications.

Medical doctors and radiologists (henceforth referred to as medical experts) can also benefit from this study through the proposed medical systems that can help them automatically delineate ROIs in MRI images. Generally, the quantitative analysis of medical images is challenging, as the segmentation of the structure of interest is the prerequisite to quantification. Manual segmentation of the tissue of interest (e.g., tumour) from each image slice by a trained radiologist, while remaining the most accepted practice, is a laborious and time-consuming process. It is also affected by inter- and intra-observer variations. Automated approaches, on the other hand, are generally considered faster, objective measures and provide accurate tissue quantification and/or tissue classification. In this context, two systems are proposed in this study; one for MRI brain images while the other for osteosarcoma (i.e., malignant bone tumour). In this thesis, the aim is to add a new valuable episode in a chain of improvements directed to this area of research.

1.11 Contributions

The main contributions of this thesis can be summarized as follows.

1. The development of an efficient fuzzy clustering algorithm called Harmony Fuzzy C-Means (HFCM) algorithm consists of two stages. In the first stage, HS explores the search space looking for the near-optimal cluster center values. In the second stage, the output of the first stage is used to initialize the fuzzy clustering algorithm, FCM, where the later performs the clustering. Such algorithm will not suffer from the initialization sensitivity and its local optimal problem, and it is particularly suitable for applications in which there is domain knowledge that can suggest a reasonable value for the number of clusters.

2. The development of another alternative clustering algorithm called Harmony Fuzzy Image Segmentation Algorithm (HFISA) based on using HS algorithm as a clustering algorithm. Each harmony memory vector is actually a fuzzy membership matrix, where its width represents the number of data points in the given dataset and its height represents the number of predefined clusters. Consequently, the evolving process of the HS is directed to find the near-optimal fuzzy membership value for each data point to the predefined number of clusters. This algorithm is an alternative to the HFCM algorithm in which the local optimal problem is avoided and the domain knowledge is available to suggest a reasonable value for the number of clusters.

3. The development of an efficient dynamic clustering algorithm called Dynamic Fuzzy

Clustering using the HS (DCHS). The DCHS algorithm could automatically determine the appropriate number of clusters as well as a good fuzzy partitioning of the given dataset. This algorithm is suitable for applications in which there is no domain knowledge available that can give any indication of what is the proper value of number of clusters.

4. The development of an efficient DCHS-based approach to automatically delineate the ROIs, i.e., white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), of the brain MRI images.

5. The development of an efficient automatic medical image segmentation framework based on the DCHS algorithm to delineate the tumour region of osteosarcoma.

1.12 Thesis Outline

Chapter 2: This chapter provides the basic definitions and concepts of harmony search, image segmentation, and data clustering that will be used in this thesis.

Chapter 3: This chapter provides a review of the state-of-the-art techniques available in the literature for the main topics of this thesis including HS algorithm, image segmentation, and the main weaknesses of the clustering approach.

Chapter 4: This chapter provides an overview of the methodology used to achieve the research objectives of this thesis.

Chapter 5: Presents the first contribution of this thesis, new HS-based fuzzy clustering algorithms. Two alternative methods from different points of view are proposed. The first method consists of two stages. In the first stage, HS explores the search space looking for the near-optimal cluster center values. In the second stage, the output of the first stage is used to initialize the fuzzy clustering algorithm, FCM, where the latter performs the clustering. The second method is based on using HS as a clustering algorithm, in which each harmony memory vector represents a clustering solution. The application of the proposed clustering algorithm to the problem of image segmentation is investigated. To illustrate its wide applicability, it is applied to different types of images such as natural images, synthetic images, medical MR images, and remote sensing images.

Chapter 6: Presents a new dynamic clustering algorithm called DCHS. In this algorithm, the capability of standard HS is modified to automatically evolve the appropriate number of clusters, as well as the locations of cluster centers. This approach is applied to unsupervised image classification and is tested on synthetic, natural, medical, and remote-sensing images. All images were selected to show the wide applicability of the proposed algorithm and to compare its results with state-of-the-art algorithms in the domain of dynamic clustering, such as GA, PSO, and other metaheuristic algorithms.

Chapter 7: This chapter presents two applications of the proposed HS-based algorithms. They are two difficult real-world medical image problems. The first application is the automatic MRI brain image segmentation problem, where the attractive tissues in brain such as WM, GM, and CSF are automatically segmented. The second application is from USM Medical center named as malignant bone tumour (osteosarcoma), in which a new framework based on multi-spectral information from various MRI sequences is proposed. In both applications, a set of experiments was conducted and the results were visually and statistically compared with other state-of-the-art methods.

Chapter 8: Concludes the main components of the proposed work, and presents plans for future research directions.

CHAPTER 2

THEORETICAL BACKGROUND

This chapter presents an overview of three topics namely Harmony search algorithm, image segmentation and clustering. Firstly, an explination of the harmony search algorithm, which is the fundamental element of the work in this thesis, is presented. In the context of the harmony search and clustering, the second part will explain the basic concepts of image segmentation. Thirdly, the theoretical background of data clustering is discussed. Basic definitions of terms will be provided about the topic followed by explanations about clustering techniques with an emphasis on fuzzy clustering within partitional clustering approaches.

2.1 The HS Algorithm

HS (Geem et al., 2001) is a relatively new population-based metaheuristic optimization algorithm that imitates the music improvisation process where the musicians improvise their instruments' pitch by searching for a perfect state of harmony. It has been successfully tailored to various scientific and engineering applications such as music composition (Geem and Choi, 2007), sudoku puzzle solving (Geem, 2007a), tour planning (Geem, Tseng and Park, 2005), web page clustering (Forsati et al., 2008; Mahdavi and Abolhassani, 2009), structural design (Lee and Geem, 2004; Geem, 2009a), water network design (Geem, 2009c), vehicle routing (Geem, Lee and Park, 2005), dam scheduling (Geem, 2007b), ground water modeling (Ayvaz, 2009, 2007), soil stability analysis (Cheng et al., 2008), ecological conservation (Geem and Williams, 2008), energy system dispatch (Vasebi et al., 2007), heat exchanger design (Fesanghary et al., 2009), transportation energy modeling (Ceylan et al., 2008), satellite heat pipe design (Geem and Hwangbo, 2006), medical physics (Panchal, 2009), timetabling (Al-Betar et al., 2008; Al-Betar, Khader and Liao, 2010), RNA structure prediction (Mohsen et al., 2010), etc. For further information on these applications see (Ingram and Zhang, 2009) and references therein. It is a very successful metaheuristic algorithm that can explore the search space of a given data in parallel optimization environment, where each solution (harmony) vector is generated by intelligently exploring and exploiting a search space (Geem, 2009c).

HS as mentioned mimic the improvisation process of musicians' with an intelligent way as can be seen in Figure 2.1. The analogy between improvisation and optimization is likely as follows (Geem, 2010):

1. Each musician corresponds to each decision variable.
2. Musical instrument's pitch range corresponds to the decision variable's value range.
3. Musical harmony at a certain time corresponds to the solution vector at a certain iteration.
4. Audience's aesthetics corresponds to the objective function.

Just like musical harmony is improved time after time, solution vector is improved iteration by iteration. In general, HS has five steps and they are described as in (Geem, Tseng and Park, 2005) as follows:

Step 1. **Initialize the Optimization Problem and HS Parameters**

The optimization problem is defined as follows:

$$\begin{aligned} &\text{minimize/maximize} && f(a), \\ &\text{subject to} && a_i \in \mathbf{A}_i,\ i = 1, 2, \ldots, N \end{aligned} \tag{2.1}$$

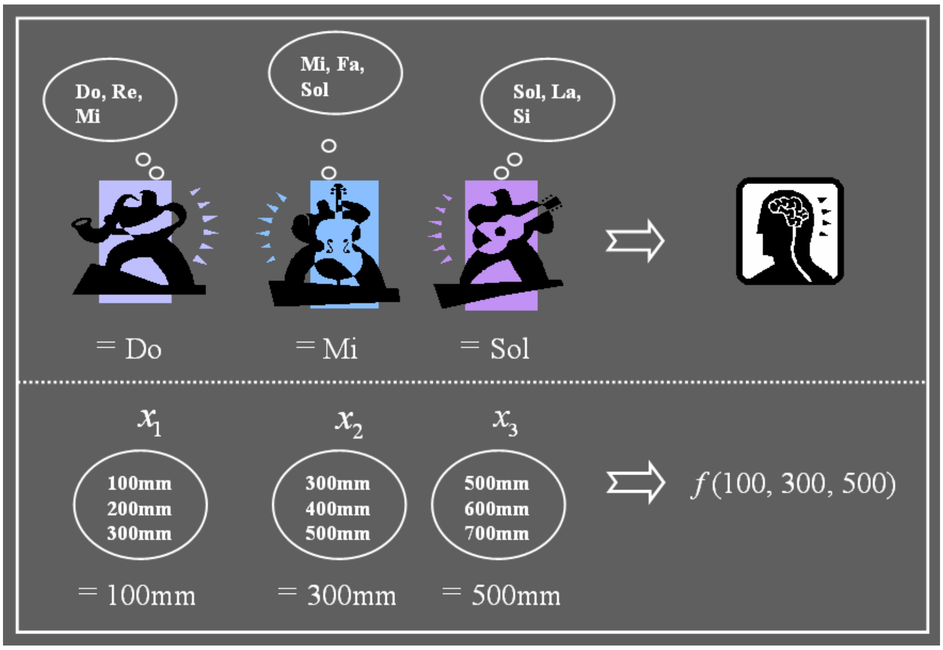


Figure 2.1: Analogy between Improvisation and Optimization, obtained from (Geem, 2010)

where $f(a)$ is an objective function; a is the set of decision variable (a_i); $\mathbf{A}_i$ is the set of possible range of values for each decision variable, ${}_La_i \leq \mathbf{A}_i \leq {}_Ua_i$; ${}_La_i$ and ${}_Ua_i$ are the lower and upper bounds for each decision variable (a_i) respectively. The variable N is the number of decision variables.

Then, the parameters of the HS are initialized. These parameters are:

(a) Harmony Memory Size (HMS) (i.e., number of solution vectors in harmony memory);

(b) Harmony Memory Considering Rate (HMCR), where HMCR $\in [0,1]$;

(c) Pitch Adjusting Rate (PAR), where PAR $\in [0,1]$;

(d) Stopping Criteria (i.e., number of improvisation (NI));

More explanation of these parameters is provided in the next steps.

Step 2. **Initialize Harmony Memory**

The harmony memory (HM) is a matrix of solutions with a size of HMS, where each HM vector represents one solution as can be seen in Eq.2.2. In this step, the solutions are randomly constructed and optionally rearranged in a reversed order to HM, based on their objective function values such as $f(a^1) \leq f(a^2) \cdots \leq f(a^{HMS})$, where $f(a^j)$ is the objective function value for j^{th} harmony memory vector.

$$HM = \left[\begin{array}{cccc|c} a_1^1 & a_2^1 & \ldots & a_N^1 & f(a^1) \\ a_1^2 & a_2^2 & \ldots & a_N^2 & f(a^2) \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_1^{HMS} & a_2^{HMS} & \ldots & a_N^{HMS} & f(a^{HMS}) \end{array}\right] \tag{2.2}$$

Step 3. **Improvise New Harmony**

This step is the essence of the HS algorithm and the cornerstone that has been laid for building this algorithm. In this step, the HS generates (improvises) a new harmony vector, $a^{'} = (a_1^{'}, a_2^{'}, a_3^{'}, \ldots, a_N^{'})$. It is based on three operators: memory consideration; pitch adjustment; or random consideration. In the memory consideration, the values of the new harmony vector are randomly inherited from the historical values stored in HM with a probability of HMCR. Therefore, the value of decision variable $(a_1^{'})$ is chosen from $(a_1^1, a_1^2, a_1^3, \ldots, a_1^{HMS})$ that is stored in HM. The next decision variable $(a_2^{'})$ is chosen from $(a_2^1, a_2^2, a_2^3, \ldots, a_2^{HMS})$, and the other decision variables, $(a_3^{'}, a_4^{'}, a_5^{'}, \ldots, a_N^{'})$, are chosen consecutively in the same manner with the probability of HMCR $\in [0,1]$. The usage of HM is similar to the step where the musician uses his or her memory to generate a good tune. This cumulative step ensures that good harmonies are considered as the elements of new harmony vectors.

Alternativally, where the other decision variable values are not chosen from HM, according to the HMCR probability test, they are randomly chosen according to their possible range, $a_i^{'} \in \mathbf{A}_i$. This case is referred to as random consideration (with a probability of (1-HMCR)), which increases the diversity of the solutions and drives the system further to explore various diverse solutions so that global optimality can be attained.

The following equation summarizes these two steps i.e., memory consideration and random consideration.

$$a_i^{'} \leftarrow \begin{cases} a_i^{'} \in \{a_i^1, a_i^2, a_i^3, \ldots, a_i^{HMS}\} & w.p.\ HMCR \\ a_i^{'} \in A_i & w.p.\ (1 - HMCR) \end{cases} \tag{2.3}$$

For example, a HMCR of 0.95 indicates that the HS algorithm will choose the decision variable value from historically stored values in the HM with a 95% probability or from the entire possible range with a 5% probability. Therefore, if a generated random number $w \in [0,1]$ is equal to 0.78 then, the value of the decision variable is picked out from HM since it is less than the HMCR(0.95).

Furthermore, an additional search for good solutions in the search space is achieved through tuning each decision variable in the new harmony vector, $a^{'} = (a_1^{'}, a_2^{'}, a_3^{'}, \ldots, a_N^{'})$, inherited from HM using PAR operator. These decision variables are examined to be tuned with the probability of PAR $\in [0,1]$ as in Eq.5.3.

$$a_i^{'} \leftarrow \begin{cases} Adjusting\ Pitch & w.p.\ PAR \\ Doing\ Nothing & w.p.\ (1 - PAR) \end{cases} \tag{2.4}$$

If a generated random number $rnd \in [0,1]$ within the probability of PAR then, the new decision variable $(a_i^{'})$ will be adjusted based on the following equation:

$$(a_i^{'}) = (a_i^{'}) \pm rand() * bw \tag{2.5}$$

Here, bw is an arbitrary distance bandwidth, or fret width fw as renamed in the updated version of HS (Geem, 2010), used to improve the performance of HS and $(rand())$ is a function that generates a random number $\in [0,1]$. Actually, *bw* determines the amount of movement or changes that may have occurred to the components of the new vector. The value of *bw* is based on the optimization problem itself i.e., continuous or discrete. In general, the way that the parameter (PAR) modifies the components of the new harmony vector is an analogy to the musicians' behaviors when they slightly change their tone frequencies in order to get better harmonies. Consequently, it explores more solutions in the search space and improves the searching abilities.

All of these operators are well-illustrated using pseudo code as in Fig 2.2.

Step 4. **Update the Harmony Memory**

In order to update HM with the new generated vector $a^{'} = (a_1^{'}, a_2^{'}, a_3^{'}, \ldots, a_N^{'})$, the objective function is calculated for each new harmony vector $f(a^{'})$. If the objective function value for the new vector is better than the worst harmony vector stored in HM, then the worst harmony vector is replaced by the new vector. Otherwise, this new vector is ignored.

$$a^{'} \in HM \land a^{worst} \notin HM \tag{2.6}$$

However, for the diversity of harmonies in HM, other harmonies (in terms of least-similarity) can be considered. Also, the maximum number of identical harmonies in HM can be considered in order to prevent premature HM.

Step 5. **Check the Stopping Criterion**

The iteration process in steps 3 and 4 is terminated when the maximum number of improvisations (NI) is reached. Finally, the best HM vector is selected and is considered to be the best solution to the problem under investigation.

HS Algorithm

```
begin
Define fitness function f(a), $a = (a_1, a_2, \ldots, a_N)^T$
Define (HMCR), (PAR), (HMS), (NI), (bw)
$HM \leftarrow GenerateInitialPoplution()$
min = minimum visible value.
max = maximum visible value.
while ($iter \leq NI$) do
  while ($i \leq N$) do
    if ($rand \in (0,1) \leq HMCR$) then
      $randLoc = 1 + floor(rand \in (0,1) \times N)$
      $a_i^{'} = a_i^{randLoc}$
      if ($rand \in (0,1) \leq PAR$) then
        adjust the value of $a_i^{'}$ by:
        $a_i^{'} = a_i^{'} \pm rand \in (0,1) \times bw$
      end if
    else
      choose a random variable:
      $a_i^{'} = min + rand \in (0,1) \times (max - min)$
    end if
  end while
  if ($f(a^{'}) \leq f(a^{worst})$) then
    $a^{'} \in HM,\ a^{worst} \notin HM$
  end if
end while
best=find the current best solution
end
```

Figure 2.2: Pseudo code of the HS algorithm.

2.1.1 HS Characteristics

The HS algorithm has several characteristics that make it one of the most important metaheuristic algorithms (Geem et al., 2001). These distinguish it from other metaheuristics such as (1) the generation of a new vector after considering all existing vectors, rather than considering only two vectors as in GA (parents); (2) the independent consideration for each decision variable in HM vector; (3) the consideration of continuous decision variable values without any loss of precision; (4) no requirement of decimal-binary conversions or a fixed number (2n) of decision variable values as in GA; and (5) no need for any starting values of the decision variables or require complex derivatives as in gradient-based methods.

The other important strengths of HS are their improvisation operators, memory consideration; pitch adjustment; and random consideration, that play a major rule in achieving the desired balance between the two major extremes for any optimization algorithm, Intensification and diversification (Yang, 2009a). Essentially, both pitch adjustment and random consideration are the key components of achieving the desired diversification in HS. In random consideration, the new vector's components are generated at random mode, have the same level of efficiency as in other algorithms that handle randomization, where this property allows HS to explore new regions that may not have been visited in the search space. While, the pitch adjustment adds a new way for HS to enhance its diversification ability by tuning the new vector's component within a given bandwidth. A small random amount is added to or subtracted from an existing component stored in HM. This operator, pitch adjustment, is a fine-tuning process of local solutions that ensures that good local solutions are retained, while it adds a new room for exploring new solutions. Further to that, the pitch adjustment operator can also be considered as a mechanism to support the intensification of HS through controlling the probability of PAR. The intensification in the HS algorithm is represented by the third HS operator, memory consideration. A high harmony acceptance rate means that good solutions from the history/memory are more likely to be selected or inherited. This is equivalent to a certain degree of elitism. Obviously, if the acceptance rate is too low, solutions will converge more slowly.

Finally, the structure of the HS algorithm is relatively easy. This advantage makes it very flexible to combine HS with other metaheuristic algorithms as can be seen in Section 3.1.

2.2 Overview of Image Segmentation

This section introduces a brief description of an image and its representation. The short form of image representation is also introduced where this form is used to reduce the computational time of the proposed algorithms as will be demonstrated throughout this study.

2.2.1 Definitions

An image is a collection of data that is represented in two-dimensional (2D) or three-dimensional (3D) space. Each location in 2D image space is considered as a pixel while in 3D plane is voxel, where the entirety pixels produce the image. In gray image, the value of each pixel is called intensity or gray level while in color image the value of each pixel is represented by three subpixels (color channels: red, green, blue). The pixel's gray level can be formally set to $f(x,y)$, where x and y are spatial coordinates of this pixel in image space, and the function f here is intensity level (Gonzalez and Woods, 2008). This study will focus only on gray 2D image; therefore, the term pixel will be used instead of voxel.

Classically, image segmentation can be defined as the process of subdividing a digital image into its constituent regions, and extracting these ROIs, which are the objects (Gonzalez and Woods, 2008). Formally this process can be described as follows:

Each image can be represented as a matrix of real numbers $W_{m\times n}$, where m and n represent the number of rows and columns of image respectively. The segmentation process normally produces a set of nonempty disjoint regions $\{R_1, R_1, \cdots, R_c\}$ which obey the following conditions:

1. $W = \bigcup_{i=1}^{c} R_i$;
2. $R_i \cap R_j = \varnothing \;\; i \neq j$;

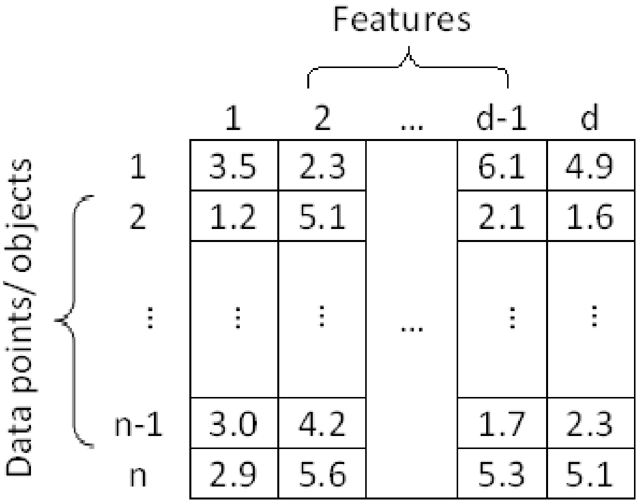
3. $P(R_i) = TRUE \ \ for\ i = 1, 2, \cdots, c;$

4. $P(R_i \cup R_j) = FALSE \ \ for\ any\ adjacent\ regions\ R_i\ and\ R_j$;

The first two conditions guarantee that each pixel in the test image belongs and belongs only to a region. The last two conditions (3 and 4) are that each region's members must satisfy the homogeneity conditions (e.g., intensity, color, or texture), while the members of two different regions fail in this homogeneity test (Jain et al., 1999).

2.2.2 Image Data Reduction Method - the Histogram

In order to reduce the computational demands of the segmentation process, a short form of image representation is used in this thesis. This form is valid only when spatial information of the given image is not considered and the intensity value of each pixel is the sole image information. This is the case in this thesis except the proposed framework for osteosarcoma, where the spatial information obtained from GLCM matrix is used. The simplification process is dependent on finding the frequency of occurrence of each pixel intensity in the given image (Bezdek et al., 1999). In other words, the histogram of the image is calculated and the frequency of each pixel intensity is produced. Therefore, the image is represented in a model such as $X = ((x_1, h_1), \ldots, (x_i, h_i), \ldots, (x_q, h_q))$ where h_i is the frequency of occurrence x_i in the image, and q is the total number of distinct x value in the image with $(\sum_{i=1}^{q} h_i = n)$, where n is the total number of the pixels in the given image. The execution time of the segmentation process will depend on the number of distinct gray levels (e.g., 256 in case of 8-bit gray scale image) in the given image which is usually very small compared to the standard number of image pixels (e.g., 262144 pixels in case of image with 512×512 size). Consequently, the calculations used in this thesis such as Euclidean distance, objective function, and validity indices will be modified and explained in their respective positions.



	1	2	...	d-1	d
1	3.5	2.3		6.1	4.9
2	1.2	5.1		2.1	1.6
⋮	⋮	⋮	...	⋮	⋮
n-1	3.0	4.2		1.7	2.3
n	2.9	5.6		5.3	5.1

Figure 2.3: An example of a dataset.

2.3 Data Clustering

In the following sections, an elaborate discussion of related clustering topics such as clustering terminology, clustering methods, cluster validity measurement, clustering characteristics is provided.

2.3.1 Notation and Terminology

Clustering algorithm classically is performed on a set of n patterns or objects which represent the elements of a dataset X, where $X = \{x_1, x_2, \cdots x_n\}$, each of which, $x_i \in \Re^d$, is a feature vector consisting of d real-valued measurements describing the features of the object represented by x_i. Figure 2.3 illustrates this concept.

Each cluster can be characterized among other clusters by some representative point(s). This charactrerization is utilized by most of the clustering methods such that, for a given cluster i, there exits an ideal point v_i, such that $v_i \in \Re^d$, which best represents cluster i's members. This point is called the center, centroid or prototype of the cluster. Thus, the clustering problem becomes that of finding a set of c centroids, $V = \{v_1, v_2, \cdots, v_c\}$ *where* $v_i \in \Re^d \; \forall \; i \in \{1, 2, \cdots, c\}$

that best represents the clustering structure in X. These clusters v_i contain data points (patterns) that are similar among them (Shihab, 2000). Normally, therefore, the process of measuring the similarity between the patterns in the dataset is considered as an essential process laying the foundation upon which the data points are clustered. The similarity measure will provide an indication of proximity, likeness, affinity, or association. The more two data objects be like one another, the larger the similarity indicator and, on the other hand, the smaller the dissimilarity indicator.

Datasets may not always contain numeric data, but they may contain different types of data such as binary, nominal, non-continuous, missing, heterogeneous data types, etc. Thus, the selection of a proper similarity measure should be done carefully. For a wise choice, and in the case of continuous numeric data, which is the case in this thesis, the Euclidean distance is the most widely used measure (Jain et al., 1999). The Euclidean distance between two data vectors is a dissimilarity index, whereas the correlation is a similarity index. The Euclidean distance is defined as follows:

$$d_2\left(x_i, x_j\right) = \left\|x_i - x_j\right\| = \sqrt{\sum_{k=1}^{d}\left(x_{i,k}, x_{j,k}\right)^2} \tag{2.7}$$

Euclidean distance is a special case (when $\alpha = 2$) of the more generalized Minkowski metric (Jain et al., 1999) defined as

$$d_\alpha\left(x_i, x_j\right) = \left\|x_i - x_j\right\| = \sqrt[\alpha]{\sum_{k=1}^{d}\left(x_{i,k}, x_{j,k}\right)^\alpha} \tag{2.8}$$

When $\alpha = 1$, the above measure is referred as the Manhattan distance (Everitt, 1993).

Nevertheless, there are many distance measure methods available in the literature that can be used as similarity measure such as Manhattan distance, cosine distance, Mahalanobis distance, matching coefficients, etc. Naturally, the selection of this method will be based on the data type of the dataset undergoing. For an introduction to common ways of extracting similarity measures for different data types, refer to (Backer, 1995; Xu and Wunsch, 2005).

2.3.2 Clustering Techniques

Clustering algorithms can generally be categorized into two groups: hierarchical, and partitional (Jain et al., 1999). The former produces a nested series of partitions, whereas the latter does clustering with one partitioning result. According to Jain et al. (2000) partitional clustering is more popular in pattern recognition applications because it does not suffer from drawbacks such as static-behavior (i.e., data points assigned to a cluster cannot move to another cluster), and the probability of failing to separate overlapping clusters, problems that are prevalent in hierarchical clustering. Partitional clustering can further be divided into two: 1) crisp (or hard) clustering, where each data point belongs to only one cluster, and 2) fuzzy (or soft) clustering, where data points can simultaneously belong to more than one cluster at the same time, based on some fuzzy membership grade.

In hard clustering, the goal would be to partition the dataset X into non-overlapping, non-empty partitions $G_1, \ldots, G_c$, where c represents the number of clusters that dataset is partitioned to, such $2 \leq c < n$. It can be defined as follows:

1. $X = \bigcup_{i=1}^{c} G_i$;

   where

2. $G_i \cap G_j = \varnothing \;\; i \neq j,\; i, j \in \{1, 2, \cdots, c\}$;

   and

3. $G_i \neq \varnothing \;\; , \; i \in \{1, 2, \cdots, c\}$;

In fuzzy clustering algorithms, the goal would be to partition the dataset X into partitions that allowed the data object to belong to a particular (possibly null) degree to every fuzzy cluster. The clustering output is in a form of a membership matrix called a fuzzy partition matrix $U = [u_{ij}]_{(c \times n)}$, where $u_{ij} \in [0,1]$ represents the fuzzy membership of the ith object to the jth fuzzy cluster. Fuzzy is considered more appropriate than crisp clustering for datasets that exhibit unclear boundaries between clusters or regions (Hore et al., 2008; Kang et al., 2009; Zhou and Schaefer, 2009). The partitional fuzzy clustering approach and its characteristics are discussed as follows.

2.3.3 Fuzzy Partitional Clustering

Fuzzy clustering is a partitional clustering technique that is based on the elements of a fuzzy set theory. The fuzzy set theory holds the notion that for a given universe of discourse, every element in the universe belongs to a varying degree to all sets defined in the universe (Zadeh, 1965, 2008). In fuzzy clustering, the universe of discourse is all the objects in the given dataset and the sets defined on the universe are the clusters. Objects are not classified as belonging to one and only one cluster, but instead, they all possess a degree of membership with each of the clusters.

The fuzzy sets framework provides a way for dealing with problems in which the source of imprecision in defining the criteria of class membership is the case. Fuzzy clustering fits well with these types of problems. It has been successfully used in, for example, various pattern recognition and image analysis problems whose datasets exhibit unclear boundaries between clusters. For instance, medical images, in particular, often consist of regions with fuzzy and disjointed boundaries. For that, fuzzy clustering has shown tremendous potential as it can naturally cope with such data characteristics.

In this thesis, fuzzy clustering algorithms is given the utmost concern, with more investigation into the most widely used fuzzy clustering algorithm, FCM (Bezdek, 1981). In the following sections, a description of FCM is introduced and fuzzy validation clustering techniques are discussed. Furthermore, the strengths and weaknesses of FCM are highlighted. A discussion of these weaknesses and their related works is provided in the next chapter.

2.3.3(a) The FCM Algorithm

FCM is one of the most used fuzzy clustering algorithms in the data clustering field. The first model of FCM was proposed by Dunn (1973), while the current version of this algorithm was proposed by Bezdek (1981). Iteratively, FCM optimizes the following objective function:

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ij}^{m} \|x_i - v_j\|^2 \tag{2.9}$$

where $\{v_j\}_{j=1}^{c}$ are the centroids of the clusters c and u_{ij} represents the fuzzy membership of ith data point to jth cluster, where the clustering output is the partitioning matrix $U = [u_{ij}]_{(c\times n)}$ which represents the membership of all data points to predefined clusters c. The membership of all data points to the clusters that belong to the following conditions, where $U \in M_{fcn}$ as in Eq. 2.10. $\|.\|$ denotes an inner-product norm (e.g., Euclidean distance as in Eq. 2.7) from the data point x_i to the v_j cluster center, and the parameter $m \in [1,\infty)$, is a weighting exponent on each fuzzy membership that determines the amount of fuzziness of the resulting classification. The value of this parameter as reported in (Pal and Bezdek, 1995) is set to 2, where this number represents a good choice for it.

$$M_{fcn} = \left\{ U \in \Re^{c\times n} \middle| \sum_{j=1}^{c} u_{ij} = 1, 0 < \sum_{i=1}^{n} u_{ij} < n, and\ u_{ij} \in [0,1]; 1 \leq j \leq c; 1 \leq i \leq n \right\} \tag{2.10}$$

FCM's steps can be summarized as follows (Bezdek, 1981):

1. Determine the number of fuzzy clusters, c that the given dataset may have.

2. Select initial cluster centers $v_1, v_2, \ldots, v_c$.

3. Compute the elements of the fuzzy partition matrix:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \tag{2.11}$$

4. Compute the cluster centers:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \tag{2.12}$$

5. Repeat steps 3 and 4 until the number of iterations t exceeds a given limit or a termination criterion is satisfied:

$$\|v_{new} - v_{old}\| < \varepsilon \tag{2.13}$$

   where $\varepsilon < 0.001$

The following section discusses the validation techniques for measuring the quality of the clustering result produced by FCM.

2.3.3(b) Fuzzy Clustering Validation Measurements

The cluster validity index is normally used to evaluate the quality of different solutions provided by different settings of a given clustering algorithm (or even by different algorithms) (Halkidi et al., 2001). In other words, the discovery of the best fuzzy clustering solution among a set of candidates requires an accurate index to quantify the quality of the fuzzy partitions obtained.

These indices are based on two main criteria (Halkidi et al., 2001):

1. Compactness: The measurement of how each cluster members are coherent to each other (e.g., cluster variance).
2. Separation: The measurement of how close the clusters are to each others.

Several indices for fuzzy clustering assessment have been proposed such as: PC (Bezdek, 1981), PE (Bezdek, 1981), KYI (Kim et al., 2004), FS (Fukuyama and Sugeno, 1989), XB (Xie and Beni, 1991), SC (Zahid et al., 1999), FHV (Gath and Geva, 1989), PCAES (Wu and Yang, 2005), PBMF (Pakhira et al., 2004; Maulik and Bandyopadhyay, 2002), PS (Chou et al., 2003, 2004), (for further information see (Wang and Zhang, 2007; El-Melegy et al., 2007; Halkidi et al., 2001) and references therein). In the following, some of them which are most used in the literature and are used in this thesis are discussed.

Bezdek's Partition Coefficient (PC) and Partition Entropy (PE) indices are well-known fuzzy cluster validity measures (Bezdek, 1981). These types of indices only use membership values (compactness) in their calculations; therefore the advantage is being easy to calculate. Both of them are similar in calculating the fuzziness of the cluster partition only. To achieve proper clustering results, PC $\in [1/c, 1]$ index should be maximized while PE $\in [0, log_a c]$ index minimized.

The indices were defined as follows:

$$V_{PC} = \frac{1}{n} \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^2 \tag{2.14}$$

$$V_{PE} = -\frac{1}{n} \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij} \; log_a \; u_{ij} \tag{2.15}$$

where a is the base of the logarithm and u_{ij} is the membership value of the ith data point to the jth cluster. The main disadvantages of these indices are their monotonic decreasing with number of cluster c, sensitivity of the fuzzifier, m and the lack of using the property of the data themselves (Wang and Zhang, 2007). (Dave, 1996) proposed a modification to PC index that could overcome the PC's monotonic problem.

Xie-Beni (XB) index (Xie and Beni, 1991) is considered a representative index of fuzzy clustering indices that uses both membership values and the dataset (Halkidi et al., 2001, 2002). The XB index is defined as a ratio of the total variation (compactness) to the minimum separation of the clusters. The objective is therefore to minimize the XB index for achieving proper clustering.

The index is defined as follows:

$$XB = \frac{\sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^{2} \left\| x_i - v_j \right\|^2}{n \; min_{i,j} \left\| v_i - v_j \right\|^2} \tag{2.16}$$

The PBMF-index is the fuzzy version of PBM-index (Pakhira et al., 2004) (also known as the $\mathscr{I}$-index (Maulik and Bandyopadhyay, 2002)). PBMF is a recently-developed index that exhibits a good trade-off between efficacy and computational concern. The PBMF index is a product of three factors and can be formulated as follows:

$$PBMF(c) = \left(\frac{1}{c} \times \frac{E_1}{E_c} \times D_c \right)^p \tag{2.17}$$

where c is the number of clusters. Here

$$E_c = \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^{m} \left\| x_i - v_j \right\| \tag{2.18}$$

and

$$D_c = max_{i,l} \| v_i - v_l \| \tag{2.19}$$

The variable n is the total number of data points in the given dataset while the power p is used to control the contrast between the different cluster configurations and it is set to be 2. m is the fuzziness weighting exponent. $E1$ is a constant term for a particular dataset which is used to avoid the index value from approaching zero and its value is dataset dependant. D_c measures the maximum separation between two clusters over all possible pairs of clusters, while E_c measures the sum of c within-cluster distances (i.e., compactness). It is also worth mentioning here that according to Pakhira et al. (2005) the maximum value of c cluster centers allowed is $\sqrt{n}$ which is considered as a safe measurement to avoid the monotonic behavior of this index. Esentially, the main goal of PBMF-index is to maximize the inter-cluster distances (separation) while minimizing the intra-cluster distances (compactness). Hence, the maximization of PBMF-index shows accurate clustering results.

2.3.3(c) Strengths and Weaknesses of FCM

FCM has several advantages that make it a preferable clustering algorithm. These advantages are summerised below:

1. The fuzzy natural of FCM provides the algorithm with more information on the given dataset.
2. Simple and straightforward programming implementation.
3. Suitable for very large datasets since its time complexity is $O(n)$.
4. Produces very good results in some conditions (i.e., hyperspherically shaped well-separated clusters).
5. Robust and is proved to converge to local optimal solution.

However, FCM like any other algorithm has some weaknesses. A summary of these weaknesses is provided below:

1. The number of clusters in the given dataset should be known a priori.
2. It is sensitive to the cluster centers initialization step, therefore a tendency to be trapped in local optima is very high.
3. It is sensitive to noise and outliers.

The main thrust of this thesis is to improve the clustering performance and overcome its weaknesses namely (1 and 2); and thus improve all related clustering-based applications such as image segmentation. Furthermore, an exploration of the ability of the new metaheuristic, HS algorithm, to solve such problems is investigated in this thesis. In the upcoming chapter, an overview of the partitional clustering weaknesses namely (1 and 2) related work is introduced. While later, the proposed solutions with application in image segmentation domain and with some more focus on medical image processing are introduced.

2.4 Summary

In this chapter, the basic definitions and concepts of harmony search, image segmentation and data clustering that will be used in this thesis are introduced. A detailed description of the population-based metaheuristic HS which is considered as the main algorithm of this thesis is introduced. The definition of image segmentation concept is also introduced as well as the concept of the short form representation of gray image. Finally, the basic concepts of clustering with its two main types partitional and hierarchal are presented. An emphasis on fuzzy partitional clustering with its representative algorithm, FCM, is also one of the major concerns of this chapter.

In the subsequent chapters, the definitions presented here will be revisited and used in the focus of the thesis - harmony based fuzzy clustering algorithms. Later the proposed algorithms will be presented and their evaluation will be given using the material described in this chapter.

CHAPTER 3

LITERATURE REVIEW

This chapter gives an overview of the related works of the main topics of this thesis including HS algorithm and the main weaknesses of the clustering approach. The HS algorithm which is the basics of the proposed algorithms in this thesis has attracted many researchers from various fields especially those working on solving optimization problems. Consequently, this algorithm guided researchers to improve on its performance to be in line with the requirements of the applications being developed. These improvements primarily cover two aspects: 1) improvements in terms of parameters setting, and 2) improvements in terms of hybridizing HS components with other metaheuristic algorithms. This chapter presents an overview of these aspects, with a goal of providing useful references to fundamental concepts accessible to the sake of improving and adapting the HS to be more suitable for fuzzy clustering and image segmentation approaches. Furthermore, the related work of the two main issues that plague clustering approaches: initialization sensitivity of cluster centers and unknown number of actual clusters in the given dataset, is presented.

3.1 The Variants of the HS Algorithm

HS algorithm attracted the attention of many researchers to solve many optimization problems such as engineering and computer science problems. Consequently, the interest in this algorithm led the researchers to improve and develop its performance in line with the requirements of the problems that to be solved. These improvements primarily cover two aspects: 1) improvement of HS in terms of parameters setting, and 2) improvements in terms of hybridizing of HS components with other metaheuristic algorithms. This section will highlight these developments and improvements to this algorithm in the ten years of this algorithm's age. The first part introduces the improvement of HS in terms of parameters setting, while the second part introduces the development of HS in terms of hybridizing of HS with other metaheuristic algorithms.

3.1.1 Variants based on Parameters Setting

The proper selection of HS parameter values is considered as one of the challenging tasks not only for HS algorithm but also for other metaheuristic algorithms. This difficulty stems from a variety of causes, and the most important one is the absence of general rules governing this aspect. Actually, setting these values is a problem dependent and therefore the experimental trials are the only guide to the best values. However, this matter guides the research into new variants of HS based on adding some extra components or concepts to make part of these parameters dynamically adapted. Table 3.1 presents a summarization of some of these modifications

3.1.2 Variants based on Hybridization of HS with other Metaheuristics

In this section, the hybridization of HS with other Metaheuristics is introduced. This hybridization can be categorized into two approaches; the first is the integration of some components of other metaheuristic algorithms into HS structure, while the second is in the opposite direction, where the integration of some HS components is integrated into other metaheuristic algorithm structure (Ingram and Zhang, 2009). In general, such hybridization process is introduced to improve the search abilities of these optimization algorithms (Blum and Roli, 2008; Grosan and Abraham, 2007). In both cases, the origin of the ability of HS algorithm to be integrated with other metaheuristic return to the relative ease and flexible structure of HS as reported in (Yang, 2009b).

Table 3.1: Variants of HS based on parameters setting improvements.

Algorithm Name	Modified Parameters	Description	References
IHS	PAR, *bw*	Dynamic setting during the improvisation process, where the PAR value is linearly increased and *bw* value is exponentially decreased.	Mahdavi et al. (2007)
GHS	*bw*	The PSO concept, global best particle, is incorporated by replacing the bw parameter altogether and adding randomly selected decision variables from the best harmony vector in HM.	Omran and Mahdavi (2008)
HS-variant	PAR, *bw*, HM initialization	Dynamic selection of bw and PAR parameters. *bw* is totally replaced by maximal and minimal values in HM. The PAR value is linearly decreased. The initialization of HM is performed using low-discrepancy sequences.	Wang and Huang (2010)
HS-variant	*bw*	*bw* will be the standard deviation of the current population when HMCR is close to 1.	Mukhopadhyay et al. (2008)
DHS	PAR	A replacement of the PAR operator with a mutation strategy borrowed from the DE is proposed.	Chakraborty et al. (2009)
HS-variant	HM	Generating two times of HMS initial harmonies but placed only the best HMS of these into the initial HM.	Degertekin (2008)
HS-variant	Stopping criterion	The stopping criterion is replaced by best-to-worst (BtW) harmony ratio in the current harmony memory.	Kattan et al. (2010)
HS-variant	PAR	A Multi-pitch Adjusting Rate strategy is proposed.	Geem, Tseng and Park (2005); Al-Betar, Khader and Liao (2010)
HS-variant	*bw*	*bw* set to a range from 1% to 10% of the total value data range.	Geem (2006)

3.1.2(a) Hybridizing HS with other Metaheuristic Components

In the first approach, where other metaheuristic components or concepts are integrated into HS, different approaches have been proposed during the last few years. A summarization of some of what has been done in hybridizing of HS with other metaheuristic components is described in the following Table 3.2.

3.1.2(b) Hybridizing HS as Components in other Metaheuristics

The second approach of HS hybridization as mentioned earlier is the integration of HS concepts or components into other metaheuristic algorithms to improve their performance. A summarization of some of what has been done in hybridizing of other metaheuristics with some HS components or concepts is described in the following Table 3.3.

3.1.3 Discussion

As an important tool for the optimization domain, the metaheuristic HS algorithm explores the search space of the given data in both intensification and diversification parallel optimization environment and provides a near-optimal solution within a reasonable time. It has many features that makes it stand out as a preferable technique not only as standalone algorithm but also to be combined with other metaheuristic algorithms.

Even though the standard HS has been successfully implemented in various applications, many modifications and improvements to this algorithm have however been reported in the literature by many researchers in various domains. Each of them is tightly related to some aspects of this algorithm such as parameters setting, balancing of intensification and diversification of HS and finally hybridizing it with other metaheuristic components.

This chapter focuses on this algorithm and survey most of the modifications proposed in

Table 3.2: Variants of HS based on hybridizing improvements

Type of Hybridization	Description	References
HS+SA	This combination is used to modify the PAR parameter using the cooling strategy of SA.	Taherinejad (2009)
HS+PSO	The PSO concept, global best particle, is incorporated by replacing the bw parameter altogether and adding randomly selected decision variables from the best harmony vector in HM.	Omran and Mahdavi (2008),
HS+PSO	The PSO concept, global best particle, is used to improve the selection process in harmony memory consideration operator (HMCR).	Geem (2009c)
HS+GA	Roulette-Wheel memory consideration which uses the survival for the fittest principle is used to improve the selection process in HMCR.	Al-Betar, Khader and Nadi (2010)
HS+CSA	The CSA is used to fine-tune all HM vectors and improve the convergence capability of HS.	Wang et al. (2009)
HS+GA+SA+AIS	This combination is used to enhance the solutions stored in HM, to speed up the convergence, and to prevent the HS from getting stuck in the local optimal problem.	Lee and Zomaya (2009)
HS+PSO+GA	It is used to make HS as a global optimization algorithm by adding two operations: position updating and genetic mutation.	Zou et al. (2010)
HS+SQP	SQP is used to support the exploitation mechanism of HS.	Fesanghary et al. (2008)
HS+K-means	k-means is used as a local search component in HS.	Mahdavi et al. (2008); Forsati et al. (2008)
IHS+FCM	FCM is integrated into IHS to improve its local search ability and fine-tune the clustering result as a final step	Malaki et al. (2008)
HS+Solver	Solver is used to support the exploitation mechanism of HS.	Ayvaz et al. (2009)

Table 3.3: Hybridizing of HS components and concepts in other metaheuristics.

Type of Hybridization	Description	References
PSO+HS	The (HM) concept in HS is integrated into PSO algorithm to prevent the pbest concept of PSO to violate the variables' boundary.	Li et al. (2007)
PSOPC+ACO+HS	HM concept is used to control the variable constraints in PSOPC.	Kaveh and Talatahari (2009)
GA+Simplex+TS+HS	The HS concept of searching is used to improve the performance of GA.	Qinghua et al. (2006)
GA+HS	The concept of selecting the decision variables from all vectors stored in the HM is mimicked to improve the GA selection mechanism.	Li et al. (2008)
GA+HS	HS is used to maintain a balance between the exploration and exploitation concepts in GA.	Nadi et al. (2010)
LDA+HS	HS is used as a preprocessing technique to overcome the LDA's problem.	Moeinzadeh et al. (2009)

the literature. Although many examples of successful applications of HS, there still remain many untackled problems due to the existence of many inherent uncertain factors. These problems have already attracted and will continue to attract intensive efforts from a wide range of disciplines.

This thesis also explores the performance of the standard HS in the domain of fuzzy clustering and image segmentation. Furthermore, a modification for HS is also introduced to this algorithm to improve its performance for some special aspects in dynamic clustering for the image segmentation approach. In this thesis, a hybridization step with local search-based algorithm, FCM, is introduced into the standard HS. This step is meant to improve the performance of HS since HS is good at finding a promising area of the search space, while FCM algorithm performs better in localized based search within those areas. This step is added for two reasons: first to increase the convergence speed of HS and secondly to improve the quality of the clustering and image segmentation results. This step is incorporated in HS in a two of the proposed algorithms in this thesis named HFISA in Chapter 5 and DCHS in Chapter 6. Furthermore, the calling procedures for FCM algorithm in DCHS was controlled by a new parameter named the Fuzzy C-Means Rate *'FCMR'*. The value of this new parameter is automatically updated and was inspired by Mahdavi et al. (2007) work, IHS. In addition, a new HS operator called the *'empty operator'* is introduced to support the variable length concept of the HM vectors in the DCHS algorithm as can be seen in Chapter 6.

3.2 Image segmentation Categories

During the last several decades different algorithms have been proposed in the literature (Fu and Mui, 1981; Pal and Pal, 1993; Pham et al., 2000; Lucchese and Mitra, 2001). These algorithms, which normally work with intensity (gray level) images, can be categorized into different groups such as thresholding-based techniques, region-based techniques, edge-based techniques, and deformable-based techniques (for more information on these groups, see the overview in Appendix B). The clustering-based technique is another category of images segmentation techniques that concerned with classifying objects into groups according to certain properties of these objects. This process is analogous to the image segmentation process which is also concerned with classifying pixels of an image into regions of interest according to certain properties of these pixels (Rosenfeld and Kak, 1982). Based on this consensus, many of the clustering algorithms have been used in the image segmentation problems (Kang et al., 2009; Kannan, 2008).

In this context, fuzzy clustering-based segmentation methods are of considerable benefit, because most of the images exhibit unclear boundaries between regions. Fuzzy clustering has shown tremendous potential as it can naturally cope with such data characteristics. It is therefore not surprising that the FCM algorithm is the most widely used algorithm in numerous applications Balafar et al. (2010); Withey and Koles (2008); Hore et al. (2008). During the last three decades, many approaches based on the FCM algorithm were proposed to address the problems of image segmentation in different fields. Some of these research studies were focused on improving the performance of FCM in segmenting brain volumes to mitigate the impact of MRI artifacts such as noise and outliers. For instance, Pham and Prince Pham (1999) modified the standard FCM objective function by incorporating the smooth membership function, and a parameter was set to control the tradeoff between them. Ahmed et al. Ahmed et al. (2002) proposed a similar approach. They modified the FCM objective function to compensate for intensity inhomogeneity and to allow the labeling of a pixel (voxel) to be influenced by the labels in its immediate neighborhood. Moreover, the authors in Zhang and Chen (2004) modified the objective function of the FCM algorithm by replacing the Euclidean distance metric with other metric named the kernel-induced distance.

However, one of the main disadvantages of these algorithms is that they compute the neighborhood term in each iteration step, which is very time-consuming Shen et al. (2005). Due to this, and in order to reduce the computational demands of such algorithms, new approaches based on image histogram representation were proposed in the literature (e.g. Liew and Hong (2003); Szilagyi et al. (2003); Chen and Zhang (2004); Shen et al. (2005); Chuang et al. (2006); Cai et al. (2007); Liao et al. (2008)). These algorithms used the gray scale levels of the given image instead of the standard pixel level representation. The simplification process is based on finding the frequency of occurrence of each pixel in the given image. However, these algorithms still suffered from a major drawback which was their lack of ability to provide a fully automatic segmentation framework and local optima problem. The optimal number of clusters (regions) in each image is assumed to be known and entered by the operator, which makes the system semi-automatic and therefore remains time-consuming and subject to operator variability.

In this context, where the determination of the number of clusters in a given dataset is considered to be a main problem in the partitional clustering domain, few efforts have been reported during the last several years focusing on the development of a clustering algorithm that can automatically determine the appropriate number of clusters without any prior knowledge. A metaheuristic-based clustering approach was considered a preferable choice for such problem. This approach is feasible and practical due to the NP-hard nature of partitional clustering problems Falkenauer (1998). Metaheuristic population-based algorithms (see Chiong (2009); Chiong et al. (2009)) are widely believed to be able to solve NP-hard problems with satisfactory near-optimal solutions and significantly less computational time compared to exact algorithms. Das and Konar Das, Abraham, Chakraborty and Konar (2009) proposed a Differential Evolution algorithm for fuzzy clustering, Saha and Bandyopadhyay proposed fuzzy variable string length genetic point symmetry (Fuzzy-VGAPS) algorithm Saha and Bandyopadhyay (2009a, 2007a), Campello et al. proposed an evolutionary-based algorithm Campello et al. (2009), while the researchers in Pakhira et al. (2005); Maulik and Bandyopadhyay (2003) used a Genetic Algorithm as a clustering algorithm (FVGA). In general, these algorithms apply an optimization algorithm (such as Genetic Algorithm or Particle Swarms Optimization) as a clustering algorithm (either hard or fuzzy) with a cluster validity index as its fitness (objective) function. Next section provide further information regarding metaheuristic-based clustering algorithms.

3.3 Cluster Centers Initialization Sensitivity - Local Optima Problem

As described earlier, selecting the initial cluster center values is considered one of the most challenging tasks in partitional clustering algorithms. Since improper selection of initial cluster centers will lead the searching process towards an optimal solution that stays in local optima, and therefore produce undesirable clustering result (Bezdek et al., 1987; Hathaway and Bezdek, 1986; Selim and Ismail, 1984). The main cause of this problem lies in the way that the clustering algorithm works as its run in a manner similar to the hill climbing algorithm (Kanade and Hall, 2007). The hill climbing algorithm is a local search-based algorithm that moves in one direction without performing a wider scan of the search space to minimize (or maximize) the objective function. This behavior prevents the algorithm to explore other regions in the search space which might have a better, or even the desired solution. Figure 3.1 is a graphical demonstration of the local optimal problem.

To present a practical illustration of such a problem, a segmentation using FCM of an image with various local optima is conducted. Figure 3.2 shows an example of the effect of the local optimal problem in the image segmentation results where Figure 3.2(a)-(b) are the original ball image with its corresponding histogram. It is obvious from this histogram that such image has no clear boundary between its regions; therefore it is difficult to determine such regions. Figure 3.2(c)-(d) are a good segmentation result of the ball image with its corresponding histogram, and Figure 3.2(e)-(f) are the inappropriate segmentation result affected due to the local optimal problem with its corresponding histogram. It is also noticeable from these histograms 3.2(d),(f) that the peaks represent the regions in the ball image, but with different membership of image pixels in each region, which indicates that both segmented images have different initialization cluster centers which in turn led to make the final membership of image pixels different.

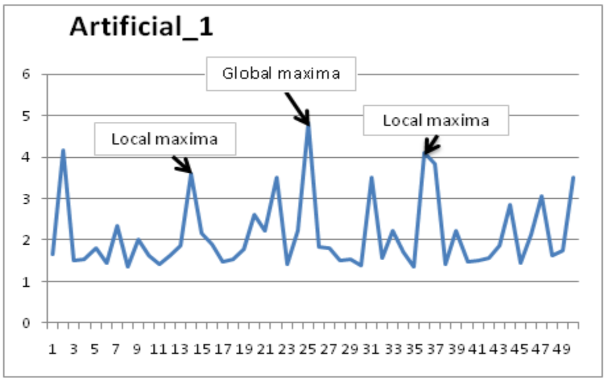
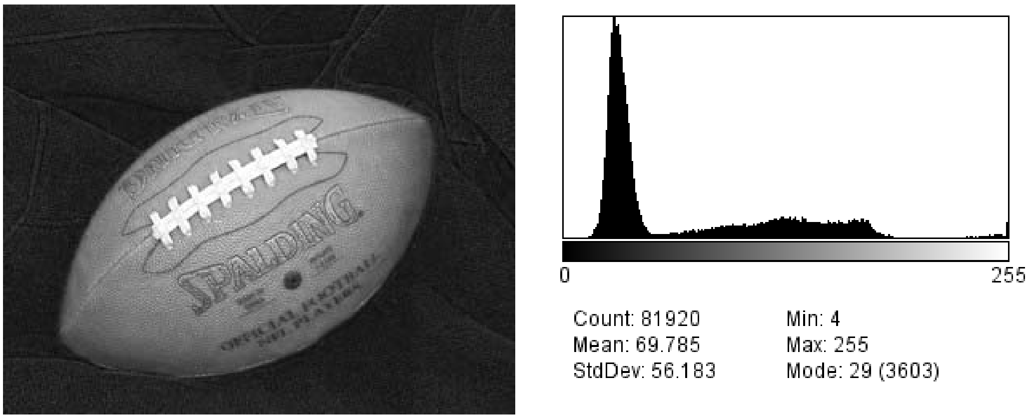


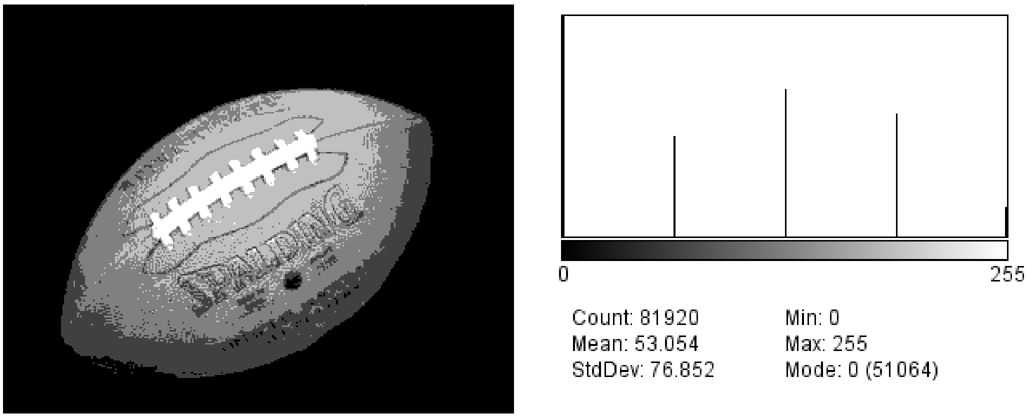
Figure 3.1: Local optima problem (maxima case).

A common and simple approach to alleviate this problem is to re-run the algorithm several times with several cluster initializations. However, this method is inapplicable in many cases such as a large dataset, or complex dataset (i.e., dataset with multiple optima) (Hruschka et al., 2009). Hence, the use of optimization algorithms in solving clustering problem is preferable.

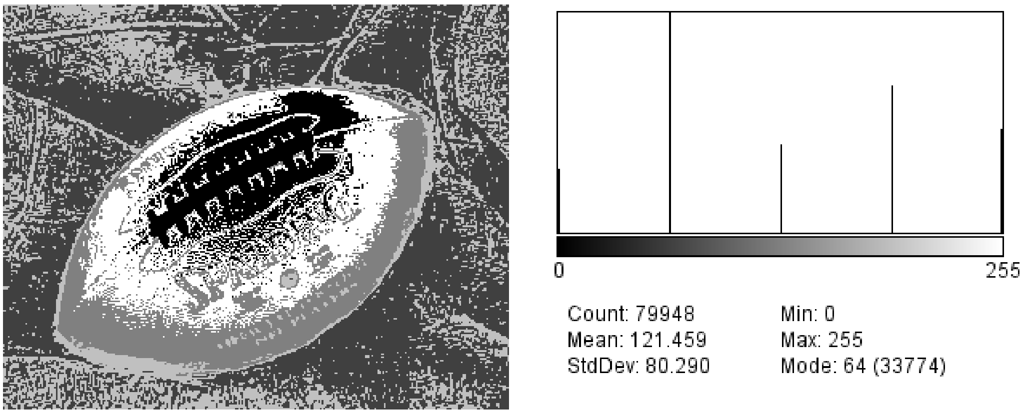
In the spirit of the main cause of this problem, the local search behavior, several global-based or improved local-based search algorithms have been proposed in the last several decades to address this problem (Kanade and Hall, 2007; Selim and Alsultan, 1991; Al-Sultan and Selim, 1993; Al-Sultan and Fedjki, 1997; Bezdek et al., 1994; Hall et al., 1999; Yong-Guo et al., 2004; Pham et al., 2007; Lili et al., 2007; Mahdavi et al., 2008; Maulik and Saha, 2009). Indeed, these algorithms include local search-based metaheuristic algorithms such as SA, TS or population-based metaheuristic algorithms such as EAs (including Genetic Programming, Evolutionary Programming (EP), Evolutionary Strategies (ES), GAs and DE), HS or swarm intelligence algorithms such as PSO, Bees algorithm, Artificial Bee Colony (ABC) and ACO. More details on metaheuristic approaches are given in Appendix A. The main advantages of population-based metaheuristic algorithms are their abilities to cope with local optima and explore large solution spaces effectively by maintaining, recombining and comparing several candidate solutions simultaneously. While the local search-based metaheuristic algorithms have some advantages over classical local search algorithms in coping with the local optima



(a) Ball image (b) Histogram of the ball image



(c) FCM segmentation result (d) Histogram of FCM segmentation result



(e) FCM segmentation result with local optima problem (f) Histogram of the FCM segmentation result affected by local optima problem

Figure 3.2: Effect of local optimal problem on image segmentation results using FCM algorithm.

problem, their results are somehow weak compared with population-based metaheuristic algorithms (Paterlini and Krink, 2006). The following is an overview of these algorithms proposed to solve clustering problem where the number of clusters in known or set up a priori.

3.3.1 Local Search-based Metaheuristic Clustering Algorithms

The local search-based metaheuristic algorithms have been used to solve such clustering problem as in the following:

In (Selim and Alsultan, 1991), the authors discussed the solution of the clustering problem usually solved by the K-means algorithm. The problem is known to have local minimum solutions, which are usually what the K-means algorithm obtains. The SA approach for solving optimization problems is described and proposed for solving the clustering problem. The parameters of the algorithm are discussed in detail and it is shown that the algorithm converges to a global solution of the clustering problem. Al-Sultan (1995) proposed a clustering algorithm with a TS heuristic and showed that the proposed algorithm outperforms both the K-means algorithm and the SA-based clustering algorithm (Selim and Alsultan, 1991). The author also proposed the same algorithm but for the fuzzy clustering problem (Al-Sultan and Fedjki, 1997). An improved TS-based clustering algorithm was proposed by Sung and Jin (2000). Their proposed algorithm is a combination of the TS heuristic with two complementary functional procedures, called packing and releasing procedures. The algorithm was numerically tested for its electiveness in comparison with reference works including the TS algorithm, the K-means algorithm and the SA algorithm. These local search metaheuristics, TS and SA, only refine a single candidate solution and are weak in coping with local optima. These algorithms are particularly parameter sensitive and tuning is highly problem-dependent (Paterlini and Krink, 2006).

3.3.2 Population-based Metaheuristic Clustering Algorithms

The population-based metaheuristic algorithms have been intensively used to solve such problem for both hard and fuzzy partitional clustering. In the following sections, a description of such algorithms based on clustering type, i.e., hard or fuzzy, is presented.

3.3.2(a) Population-based Metaheuristic Algorithms for Hard Clustering

GA-based Hard Clustering:

As reported in (Paterlini and Krink, 2006), the first attempt to use GA as a clustering algorithm was by Raghavan and Birchard (1979). The main idea of their algorithm is to use GA mechanism as an alternative clustering algorithm by using an integer encoding scheme to represent each candidate solution. The length of candidate solution is equal to the number of data points (n) in the given dataset, where each gene represents one data point and the value of this gene is selected from the clusters (k) range $\in [1,k]$. The GA tries to find the optimal partition according to a fitness function which measures the partition goodness. Despite the fact that such an algorithm outperforms k-means in the clustering simulated and real datasets (Hruschka et al., 2009; Campello et al., 2009), it suffers from some shortcomings as a result of using integer encoding where redundant solutions can be generated (i.e., more than one chromosome represents the same solution). Thus, the size of the search space to be explored by GA is much larger than the original space of solutions (Hruschka et al., 2009). Murthy and Chowdhury (1996) considered the same encoding technique used in (Raghavan and Birchard, 1979) for problem representation. Their technique still suffers from the same problem mentioned above for integer encoding. Furthermore, Bandyopadhyay and Maulik (2002a) reported that this type of encoding may cause high computational demands in case of large dataset since the length of each chromosome is linked to the length of dataset. The authors also reported that the GA's operators (e.g., single point crossover) used by Murthy and Chowdhury (1996) to generate a new chromosomes may produce invalid chromosomes which may affect the algorithm performance.

Krishna and Murty (1999) proposed a genetic k-means algorithm (GKA) in order to achieve global search and fast convergence. GKA is based on a hybridization strategy that gathers the strength features of both GA and k-means algorithms. The hybridization takes place when the crossover operator in GA is replaced by 'one step' of k-means algorithm. In addition, a clustering-based mutation operator is proposed to overcome the weakness of other GA-based clustering algorithms. By using finite Markov chain theory, it proved that GKA converges to the best-known optimum. However, the integer encoding that has been used in this algorithm may weaken its performance as mentioned earlier.

Bezdek et al. (1994) used hard partitioning matrice to represent the clustering problem and initialize GA's population. Each solution vector (chromosome) in GA's population is represented by a $(n \times k)$ matrix, where each row corresponds to a cluster and each column represents one data point (object). A '1' in row i and column j means that data point i belongs only to cluster j while others are 0's in column j. This matrix-based binary encoding scheme has the clear disadvantage of requiring $O(k \times n)$ memory space, against $O(n)$ space of the usual string-based binary encoding scheme. However, when the number of features for the given dataset is large, the computational time required to recover the data partition from a given chromosome is less for matrix representation scheme than the string-based binary encoding scheme (Hruschka et al., 2009). Another binary encoding technique, string-based binary encoding, is used by Kuncheva and Bezdek (1997) where each chromosome represents a clustering solution. The length of each chromosome is the same as the length of the number of data points (n). Each position of the binary string corresponds to a particular data point, i.e., the ith position (gene) represents the ith data point. The value of the ith gene is '1' if the ith data point is a prototype (e.g., cluster center or medoid) and '0' otherwise. The evolving process, which included a classical GA's operators, used in this algorithm usually suffers from serious drawbacks (e.g., generating an invalid offspring) in the specific context of evolutionary clustering (Hruschka et al., 2009).

Real encoding or float encoding is another encoding scheme used in the literature to represent the clustering solutions. In this scheme, an attempt to avoid the aforementioned shortcomings of other schemes is considered. Here, cluster centers are the most widely used to represent the partitioning (Hruschka et al., 2009). Each string encodes k clusters in an d dimensional space,$\Re^d$, then its length is $d \cdot k$. Thus, the first d positions represent the d coordinates of the first cluster center, the next d positions represent the coordinates of the second cluster center, and so forth. Maulik and Bandyopadhyay (2000) proposed a GA-based clustering technique, called GA-clustering. They used a real encoding scheme to represent cluster centers. The searching capability of GAs using standered GA's operators was used to search for appropriate cluster centers in the feature space such that a similarity metric (i.e., within cluster spread) of the resulting clusters is optimized. Bandyopadhyay and Maulik (2002a) proposed a GA-based clustering technique called KGA. They used the same real encoding scheme as in (Maulik and Bandyopadhyay, 2000) to represent cluster centers. In their work, they inspired the idea of hybridizing K-means algorithm with GA proposed by Krishna and Murty (1999) to enhance the performance of their proposed algorithm. The hybridization step is introduced in their algorithm as an extra step performed in each iteration of the algorithm to fine tune each gene value (i.e., cluster center) in the chromosome by performing one step k-means algorithm; this is somehow differed from what is found in (Krishna and Murty, 1999). Both GA-clustering and KGA suffer from some shortcomings related to the use of the standard GA's operators, single point crossover and mutation, where such operators can generate illegal solutions as reported in (Hruschka et al., 2009).

Bandyopadhyay and Saha in 2007 proposed a genetic-based hard clustering algorithm named GAPS (Bandyopadhyay and Saha, 2007). In this algorithm, a new point symmetry (PS)-based distance measurement is proposed. This modification was introduced to overcome the expected weakness of using the standard Euclidean distance measurement in clustering a dataset with arbitrary shapes. As the author reported, this algorithm is able to detect any type of clusters, irrespective of their geometrical shape and overlapping nature, as long as they possess the characteristic of symmetry. The experimental results on various numerical artificial and real-life demonstrate that GAPS outperforms other symmetry-based clustering technique SBKM (Su and Chou, 2001), its modified version Mod-SBKM (Chou et al., 2002), and the K-means algorithm. Using of point symmetry as a distance measurement enables the algorithm to detect any type of clusters yet such measurement cannot perform well in case of non-symmetrical properties of the clusters.

Bee-based Hard Clustering:

Pham et al. (2007) proposed a Bees-based hard clustering algorithm. Their proposed clustering method exploits the search capability of the Bees Algorithm to overcome the local optimum problem of the k-means algorithm. Specifically, the task is to search for appropriate cluster centers, which are represented as a real number in each bee, such that the k-means' objective function is minimized. They compared the performance of their proposed algorithm with k-means and GA-based clustering algorithm on some numerical benchmark data. The results showed that Bees-based clustering algorithm outperforms the others. A variant of swarm intelligence class, inspired by bees' behavior, known as Honey Bee Mating Optimization (Abbass, 2001) was applied as a hard clustering algorithm in (Fathian et al., 2007; Fathian and Amiri, 2008). The standard K-means objective function was used as a fitness function to measure the quality of the generated chromosome. The SA was used to select the set of solutions from the search space to make a mating pool for possible information exchange. An intermediate crossover was adopted in their study where the weighted average of parents is used to generate