ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUZZY AND NON-FUZZY APPROACHES FOR DIGITAL IMAGE CLASSIFICATION

MOHAMMED DIYKH, YAN LI
School of Agricultural, Computational and Environmental Sciences
Faculty of Health, Engineering and Sciences
University of Southern Queensland, Australia
{Mohammed.Diykh; Yan.Li}@usq.edu.au

ABSTRACT

This paper classifies different digital images using two types of clustering algorithms. The first type is the fuzzy clustering methods, while the second type considers the non-fuzzy methods. For the performance comparisons, we apply four clustering algorithms with two from the fuzzy type and the other two from the non-fuzzy (partitonal) clustering type. The automatic partitional clustering algorithm and the partitional k-means algorithm are chosen as the two examples of the non-fuzzy clustering techniques, while the automatic fuzzy algorithm and the fuzzy C-means clustering algorithm are taken as the examples of the fuzzy clustering techniques. The evaluation among the four algorithms are done by implementing these algorithms to three different types of image databases, based on the comparison criteria of: dataset size, cluster number, execution time and classification accuracy and k-cross validation. The experimental results demonstrate that the non-fuzzy algorithms have higher accuracies in compared to the fuzzy algorithms, especially when dealing with large data sizes and different types of images. Three types of image databases of human face images, handwritten digits and natural scenes are used for the performance evaluation.

Keywords: *Clustering Algorithms, Fuzzy Clustering, C-Means Clustering, K-Means Clustering, Partitional Clustering.*

1.. INTRODUCTION

In pattern recognition, the purpose of a clustering process is to separate an unlabeled dataset into several groups [13, 15, 20, 26, 27]. Clustering is defined as a process of grouping data items based on a measure of similarity [15, 25]. It is a subjective process as the same set of data items often need to be partitioned differently for different applications. This subjectivity makes the process of clustering difficult. A possible solution lies in reflecting this subjectivity to a certain form of knowledge. This knowledge has been used either implicitly or explicitly in the knowledge based clustering algorithms [14]. Therefore, one single algorithm or approach is not suitable to solve all the clustering problems. For instance, some algorithms are more suitable for clustering documents and texts [6], and they have better performances when document and text types of data are used. In the literature many papers were found for discussing clustering techniques. Some of them are non-fuzzy (also called partitional) algorithms, while others are fuzzy algorithms. Partitional clustering algorithms split data points into k partitions, where each partition represents a cluster. The partitioning is done based on a certain objective function. One such criterion is to minimize the squared error function which is computed as follows:

$$\boldsymbol{E} = \sum_{j=1}^{k}\sum_{i=1}^{n} || \frac{x_i^{(j)}}{c_j} ||^2 \quad (1)$$

where $\boldsymbol{E}$ is the objective function $|| \frac{x_i^{(j)}}{c_j} ||^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, an indicator of the distance of the n data points from their respective cluster centres, k is the number of cluster. The clustering process should exhibit the properties of: each group must contain at least one data point, and each data point belongs to exactly one group.

The main drawback of this type of clustering algorithms is that whenever a data point is close to the center of another cluster, it gives poor results, due to the overlapping of the data points. Corresponding to the fuzzy nature of many practical problems a number of fuzzy clustering methods have been developed following the general fuzzy set theory strategies outlined by Zadeh (1965). The main differences between the traditional hard partitional clustering and fuzzy clustering techniques are: in the hard clustering algorithm, individual points in a dataset belong only to one cluster; while in the fuzzy clustering individual points are allowed to belong to several clusters with a different degree of memberships [3, 12, 14].

This paper makes performance evaluations among four chosen clustering algorithms, with the two of them are fuzzy clustering algorithms, while the other two are non-fuzzy clustering (partitional) algorithms. The evaluation are based on dataset size, the number of clusters, execution time, classification accuracy and k-cross validation. The algorithms are implemented using three same image databases in this study. The databases are: the face images database from Cambridge University Computer laboratory[1], handwritten digits dataset from the United State Post Office Advanced Technology[2] , and the Natural Scene Dataset[3] from the Computational Visual Cognition Laboratory at Massachusetts Institute of Technology, the USA . The details of the databases are introduced in experimental data section.

2. RELATED WORK

Many papers were reported to evaluate clustering algorithms. These evaluations were investigated from the following aspects:

Bataineh et al. (2011) evaluated and tested same type of clustering algorithms (such as fuzzy clustering algorithms): a fuzzy c-means algorithm and a subtractive clustering algorithm were used in their study. The evaluation was made based on the validity measurements of the clustering results. Validity measures are scalar indices that assess the goodness of the partitions obtained. The data sets used in this study were from the MathWorks website (Mathematical Computing Software) which contains 50 distributed points in the three well-defined clusters.

Four clustering algorithms were presented and Evaluated by Abbas (2008). Those four algorithms were: a k-means clustering, a hierarchical clustering, a self-organizing map (SOM), and an expectation maximization (EM) clustering algorithms. The algorithms were applied to several simple random and non-random datasets chosen from several websites from the Internet. The partitional algorithms (k-means and EM) were also applied to huge datasets, while the hierarchical clustering algorithms (hierarchical and SOM) were applied to small datasets. Based their results, the hierarchical and SOM algorithms provided better results compared to k-means and EM algorithms when random datasets were used.

Kaur (2013) made a study to assess the classification qualities of a k-means and a hierarchical algorithms: Both algorithms were tested using a set of student data, consisting of 10 attributes which were total marks, subject marks, etc. The evaluation was based on validation measures, such as entropy, f-measure, coefficient of variance, and execution time. The study showed that the classification accuracies of k-means algorithm were better than those of the hierarchical algorithm, and with a less execution time.

Another evaluation study was made using a single linkage, a complete linkage, a group average, and a ward hierarchic clustering algorithms: A dataset of seven collections of documents, queries and relevance judgments were used for the evaluation. In this study, the accuracy of the data retrieval was used as the criterion for the effectiveness of the algorithms (El-Hamdouchi et al. 1989).

Evaluation of the performances of a fuzzy-means algorithm and an entropy based fuzzy clustering algorithm was carried out by Chattopadhyay et al. (2012). The assessments were done with four different types of datasets which were related to the chemical analysis of different samples of followers, the quality of clustering results, and the computational time.

The performances of a semi-supervised consensus clustering (SSCC) algorithm with three other clustering algorithms: These three algorithms are a k-means, a consensus clustering algorithm and a semi supervised clustering algorithm were assessed and tested. The four algorithms were implemented for analyzing gene expression data. This study investigated the roles of prior knowledge and consensus clustering for improving the clustering process. Eight cancer gene expression datasets were used in this study. The study showed that the SSCC algorithm was effective algorithm among the four. In addition, it was reported that the integration between the semi-supervised clustering and consensus clustering would improve the clustering process, especially for complex datasets (Wang et al. 2014).

[1] http://www.cl.cam.ac.uk/research/dtg/attarchive
[2] http://www.cedar.buffalo.edu/Databases/CDROM1/
[3] http://cvcl.mit.edu/database.htm

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In this paper, the evaluations are done from the following aspects:

- The evaluations are done with the fuzzy and non-fuzzy (partitional) clustering algorithms, using different types of clustering techniques.
- The evaluations are conducted using complex image databases. The databases are the face images datasets from the Cambridge University Computer Laboratory (formerly the ORL database of faces), the United States Post Office dataset for handwritten, and the natural scene dataset from Computational Visual Cognition Laboratory.
- The performance evaluations are evaluated using various criteria of data size, cluster number, execution time, k-cross validation, and clustering accuracy.

3. PARTITIONAL (NON FUZZY) CLUSTERING ALGORITHMS

Partitional clustering algorithms decompose a dataset into a set of disjointed clusters [11, 30]. Assume that a dataset of $\boldsymbol{X}$ points, a partitioning clustering approach constructs $\boldsymbol{k}$ ($\boldsymbol{X}\geq\boldsymbol{k}$) partitions of the data, with each partition representing a cluster. It classifies the data into $\boldsymbol{k}$ groups based on the conditions of:

- Each group contains at least one data point,
- Each data point belongs to exactly one group.

Note that for fuzzy clustering algorithms, a data point can belong to more than one group. Two partitional clustering algorithms are applied in this paper. The first one is an automatic partitional clustering algorithm, while the other one is a non-automatic partitional clustering algorithm.

3.1. Automatic partitional clustering algorithm

This algorithm was reported by Sarsoh et al. (2012). It was developed based on graph theory. The number of the resulted clusters was not given a priori, rather it was automatically determined through the implementation process of the algorithm. The key idea of the automatic partitional clustering algorithm is discussed below.

Firstly, the following terms/symbols are used in the automatic partitional algorithm.

- $d(x,y)$ is the Euclidean distance between individual points, x_i, y.
- y is the set of the neighborhood of x_i.
- $den^*(x_i)$ is the adaptive density of x_i.

Algorithm steps

1) Preprocessing: Given a set of data points $\boldsymbol{X}=\{x_1, x_2, x_3, \dots\dots x_n\}$,
   - Determine the adaptive neighbors $[V^*(x_i)]$ for each data point.
   - Compute the adaptive density $[den^*(x_i)]$ for each data point.
2) Constructing a tree.
   - Find the first point x_i that has a density of more than 1.
   - $\forall\, y \in V^*(x_i)$, $y \neq x_i$,where y is a neighbor of x_i compute the following :

$$\gamma_{xy} = (den^*(x_i) - den^*(y))/d(x_i, y)$$

$$\gamma_x = \min \gamma_{xy} \quad , \quad y \in V^*(x_i)$$

   - Test the value of γ_{xy} and γ_x to determine whether x_i is the root or a leaf of the root.
3) Repeat step (2) until finish all the data points in $\boldsymbol{X}$.

3.2. Partitional K-means clustering algorithm

The k-means clustering algorithm is a popular method using the partitional clustering technique. In this algorithm the number of clusters must be given a priori. The algorithm was staged as follows [14]:

Given a set of initial clusters (k clusters).

1) Assign each data point in the dataset to one of the k clusters.
2) Then each cluster center is replaced by the mean point for the relevant cluster.
3) Repeat steps (1) and (2) until the convergence is reached.

4. FUZZY CLUSTERING ALGORITHMS

Fuzzy clustering techniques are beneficial to multi-dimensional data sets, where the datasets have partial or fuzzy relations among the elements/data points. This means that each member in a dataset can belong to one or several clusters with different degrees [11]. Assume a set of n objects: $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$, where x_i is a d-dimensional point. A fuzzy clustering method attempts to partition the finite collection set of object $\boldsymbol{X}$ into a collection of $\boldsymbol{k}$ clusters, $k_1, k_2, \ldots, k_k$. Partition matrix $\boldsymbol{W} = w_{i,j} \in [0, 1]$, for $i = 1 \ldots n$ and $j = 1 \ldots k$, where each element $w_{i,j}$ is a weight that represents the degrees of memberships of object x_i in cluster k_j [26]. A lot of clustering algorithms have been developed, such as the fuzzy c-means clustering algorithm and the maximum tree clustering algorithm.

In this paper, two fuzzy clustering algorithms, an automatic-fuzzy algorithm and a non-automatic fuzzy algorithm are chosen for comparing this type of clustering techniques.

4.1. Automatic fuzzy algorithm

Sarsoh et al. (2007) proposed an effective automatic fuzzy clustering algorithm. This algorithm uses the neighborhood concept and the number of clusters is automatically determined during the implementation process of the algorithm. The key concept of this algorithm is summarized as follows:

1) Let $\boldsymbol{X}=\{x_1, x_2, \ldots, x_n\}$ be a vector containing the dataset. Determine the adaptive neighbors for each individual data point (x_i) for a given threshold, δ.
2) Compute the density of each x_i as follows:
   Density (x_i)=*Cardinal (adaptive _neighbors*(x_i)*)*
3) Sort the elements of the vector density (*Density*(x_i)) in descending order, and swap the corresponding image in $\boldsymbol{X}$, according to the sorted results, the adaptive neighbors will be also swapped.
4) The first element in $\boldsymbol{X}$ creates the first cluster. All its adaptive neighbors are also assigned to that cluster.
5) Consider the second data point in $\boldsymbol{X}$ for clustering:

If it has been assigned to any existing cluster, then

all its adaptive neighbors are also assigned to that cluster.

Else

The data point creates a new cluster and all its adaptive neighbors are assigned to this new cluster.

6) The process continues until the last element of $\boldsymbol{X}$ is clustered.

4.2. Fuzzy C-means clustering (FCM) algorithm

The FCM is a non-automatic fuzzy clustering algorithm. It is one of the most popular clustering algorithms which allow one piece of data to belong to more than one cluster. The number of clusters is predefined. The steps of this algorithm are shown in the following [3, 50]:

(1) Initialize $U = [u_{ij}]$, where u_{ij} is a degree of membership of x_i in cluster *j;* x_i is the *i*th element of d-dimensional measured data.
(2) Calculate the center of vectors $C^k= [c_j]$ with $U^{(k)}$, where k is the iteration step.

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

where c_j is the d-dimension center of the cluster.

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

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

(4) If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$, then stop; otherwise return to step (2), Where ε is a termination criterion threshold value between 0 and 1.

5. EVALUATION CONDITIONS

The experiments are conducted based on the following conditions:

1) Using the same programming language Matlab 2013b.
2) Using the same computer (Intel (R) core™ i7, CPU 3.40GHz, 8.00 GB RAM, 64-bit Microsoft windows).
3) Using the same datasets: the human face images data sets, the USPS handwritten images and the natural scene images dataset, for all the four chosen algorithms
4) Using the same comparison criteria, which are dataset size, execution time, cluster numbers, k-cross validation and clustering accuracy.

As mentioned before, the automatic partitional clustering algorithm and partitional K-means clustering algorithm are chosen from the non-fuzzy clustering algorithms, while automatic fuzzy algorithm and fuzzy C-means clustering algorithm are chosen as the examples for the fuzzy algorithms.

6. EXPERIMENTAL DATASETS

For evaluating the non-fuzzy (partitional) and fuzzy clustering algorithms, the following databases are used.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- The human facial images dataset (ORL database of faces) (Samaria 1994). It contains a set of face images taken between April 1992 and April 1994 at the University of Cambridge Computer Laboratory. The dataset contains about 400 images from 40 different people, with each person having 10 images taken with various facial expressions (eye open, eye close, smiling, not smiling), facial details (with glasses, no glasses), with different time, and varying lights. The size of each image is 92x112 pixels, with 256 grey levels per pixel. Fig. 1c shows some facial images from the dataset. The dataset is freely public available, and it can be downloaded from the Digital Technology Group of Cambridge Laboratories.
- The United State Post Office Advanced Technology Database Handwritten Digits dataset (CDROM 1992). The dataset contains more than 300 hand written digital images. There are 10 classes, with each class representing one digit of 0 to 9. This database was collected by the Research Center at the University at Buffalo, State University of New York. Fig. 1b shows the examples of the handwritten digits from the database.

- The Natural Scene Dataset from the Computational Visual Cognition Laboratory at Massachusetts Institute of Technology, the USA (Lazebnik et al. 2011). It contains eight categories of natural scene images: forests, mountains, open countries, coasts, inside cities, tall buildings, highways and streets, with each category containing 200 to 400 images. All the images are in JPG format and colored. The average of the image sizes is 256x256 pixel. The main sources of the images were from commercial databases, including Google images and personal photographs. Fig. 1a shows some natural scene examples from the dataset. The dataset is freely public available and it can be downloaded from comp putational Visual Cognition Laboratory.

a. Natural scene dataset

b. Handwritten dataset

c. Human facial image dataset

Fig. 1. Sample images from three the four datasets

7. EXPERIMENTAL RESULTS

The four chosen algorithms are implemented using Matlab R2013. The experiments with the three databases were conducted in order to evaluate the performances of the algorithms. The comparisons were conducted based on the data size, cluster numbers, execution time k-cross validation and the classification accuracy.

7.1. Performance Assessment

In this study, the cross validation, and accuracy are used to evaluate the performances of the four algorithms.

- K-cross-validation: in pattern recognition, k-cross- validation is a very popular measure to evaluate the performance of a classification method. It is used to estimate the quality of a classification method by dividing the number of the correctly classified results by the total of the cases. A dataset is divided into k-alternately exclusive subsets of an equal size. One subset is used as the testing set, while others are considered as the training sets. All the subsets are tested and the accuracy of the classification is calculated. In this work, the 10-cross-validation is used. The average of the overall results for the subset testing is computed.

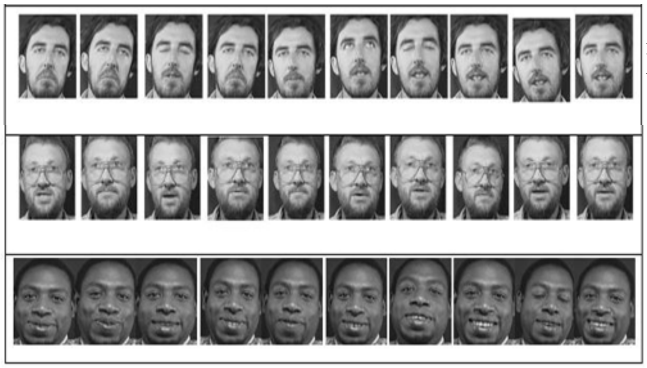
$$\text{Performance}=\frac{1}{10}\sum_{1}^{10}\text{accuracy}^{(k)} \quad (2)$$

where $\text{accuracy}^{(k)}$ is the accuracy for the kth iteration (k=1, 2,….10).

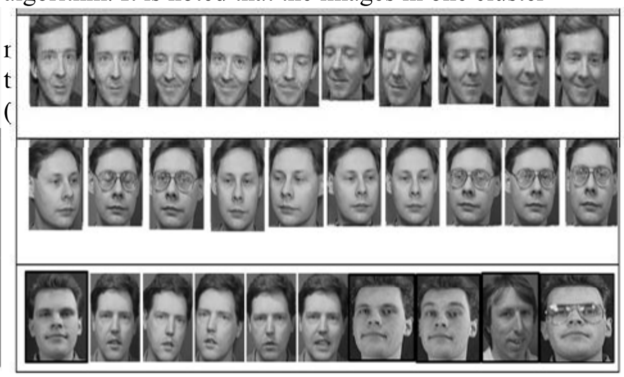
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- Classification accuracy: is the number of the correctly classified decisions divided by the total number of the cases.

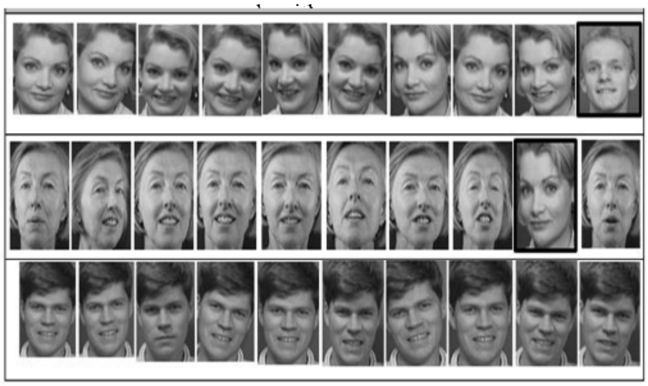
images from different persons, for example in row 3 ([3,1] [3,7][3,8] [3,9] and [3,10]). As shown by the squared images. Fig. 2c, presents the samples of clustering results obtained by the automatic fuzzy algorithm. It is noted that the images in one cluster



a. Automatic partitional clustering



b. Partitional k-means clustering algorithm



c. Automatic fuzzy clustering algorithm

d. Fuzzy c-means clustering algorithm

Fig. 2. The Samples Of Clustering Results Obtained By The Four Clustering Algorithms For The

7. 2. Experiment 1

In this first experiment, the four algorithms were applied to the human face images database. The database was divided into two groups. The first group contained 50 images which were selected randomly, while the second group contained all the 400 images in the database. Each algorithm was executed twice, firstly by using a small set of 50 images and the other one by the full 400 images. Figs. 2 shows some typical clustering results that were obtained from the four algorithms, using the two groups of the human faces. From the results in Fig. 4a, it is concluded that the automatic partitional clustering algorithm constructed the correct clusters as the same number of the persons, and each cluster contained the right images belonging to that one person. Fig. 2b shows the samples of clustering results obtained by the partitional k-means clustering algorithm. It is noted that it is possible one cluster would include mixed image images from different persons as showing by squared images. The results in Fig. 2d, show that some clusters obtained by the fuzzy c-means clustering algorithm would share similar images from different persons with similar features, such as having beards or wearing glasses. For example, the second and third clusters contain the same image from the one same person (shown in [2, 9] and [3, 10]). From the experimental results, we can see that the images belonging to one person were correctly grouped into one cluster by the automatic partitional clustering algorithm and partitional k-means clustering algorithm. There were no wrongly clustered images. However, from the results obtained by the automatic fuzzy algorithm and the FCM algorithm, we notice that the algorithms would group some images from two or even more persons into one cluster. For instance, the clusters in Fig. 2c and 2d include different face images that

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

belong to different persons. The results demonstrate that the non-fuzzy algorithms are more accurate and the performances of the algorithms are better.

Fig.3 and Fig.4 show, the execution time and accuracy of the four algorithms. From Fig 3, we can notice that the non-fuzzy algorithms are outperformed the fuzzy algorithms using two different sizes of datasets. However, regarding execution time,

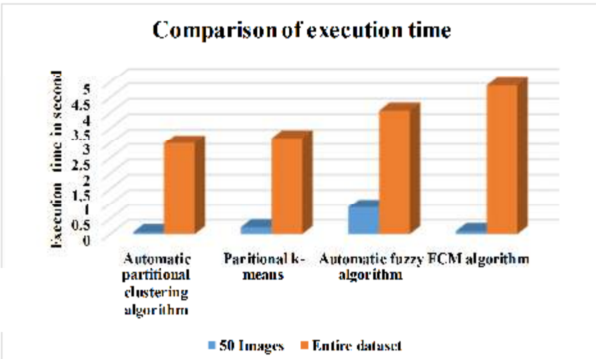


Fig. 3. Classification accuracy of the four algorithms

1) Regarding the execution time, the non-fuzzy automatic partitional clustering algorithm recorded the lowest execution time in comparing with the other algorithms. This reflects the speed of the algorithm in the clustering process whilst the FCM algorithm was recorded the biggest execution time.

2) The number of clusters which are constructed by the automatic partitional clustering algorithm has no difference

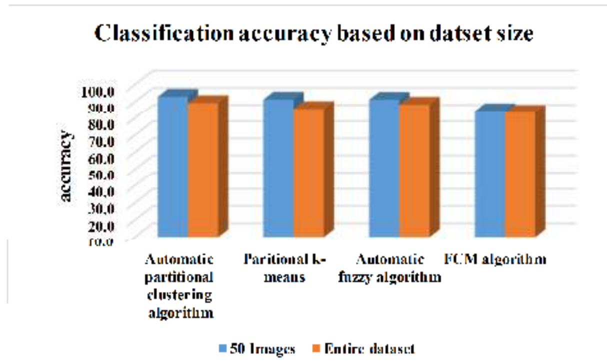


Fig. 4. Comparison of execution time among the four algorithms

Table 1 clustering number comparison

Algorithm	Data Size	The actual number of clusters	The obtained number clusters
Automatic partitional clustering algorithm	50	5	5
	400	40	45
Partitional K-means clustering algorithm	50	5	6
	400	40	47
Automatic fuzzy clustering algorithm	50	5	6
	400	40	38
FCM algorithm	50	5	7
	400	40	48

the automatic partitional clustering algorithm and partitional k-means clustering algorithm are recorded the lowest execution time using 400 face images compared with others. The 10-cross validation of the four algorithms were 91%, 87%, 86% and 81%, respectively. Table 1 shows the comparison results obtained against the size of the dataset in terms of the number of clusters after conducting the four algorithms with the face images database.

The results in Table 1 reflect the characteristics of the four algorithms in term of number of clusters. From the obtained results we can see that:

from the real clusters for the small dataset. For the large dataset of 400 images, the automatic fuzzy clustering algorithm yields the smallest difference from the real cluster number. In contrast the cluster numbers resulted by the FCM algorithm are affected by the dataset size.The accuracy of the automatic partitional clustering algorithm is the highest among the four clustering algorithms whilst the partitional K-means algorithm and the automatic fuzzy algorithm show a good result with the small dataset.

3) It appears that the non-fuzzy clustering algorithms: automatic partitional clustering algorithm and automatic K-means clustering

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

algorithm, are more effective for clustering large datasets.

4) The FCM algorithm is sensitive to large datasets. It executes fast for small datasets, but getting slow with large sizes of datasets. Overall, the algorithm is not suitable for clustering the face image datasets as their accuracies are also the lowest ones.

partitional k-means in Fig. 5b, it is noted that there are mixed images from different digits grouped into one cluster. The classification accuracy is slightly affected by the size of the data for the partitional k-means clustering algorithm for the dataset. From Fig. 5c, it is also seen that there are mixed images from different digits grouped into one cluster. Based on the results shown in Fig.5d, the two clusters contain mixed images that belong to different digits. It appears that the main problem lies in the recognition of the four digits of 0, 2, 9 and 8 for the FCM algorithm, with better results for other digits. Based on Fig. 6,

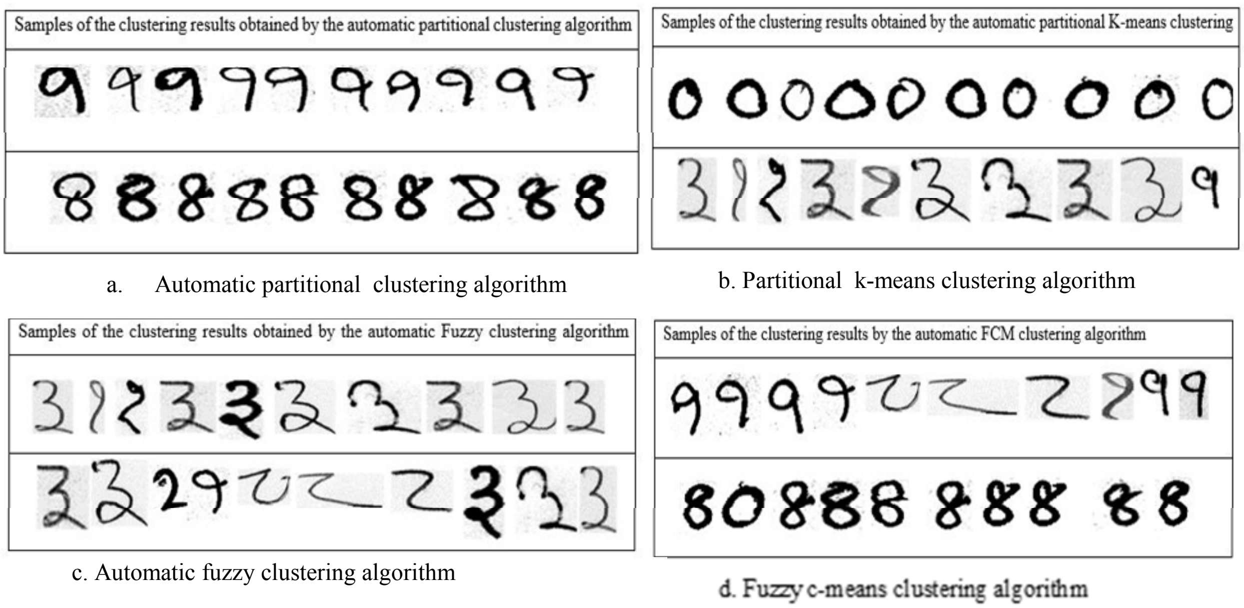


Fig. 5. Samples Of The Clustering Results Obtained By The Four Clustering Algorithms For The Hand Written

7.3. Experiment 2

The second experiment is to assess the performances of the four algorithms using the United State Post Office advanced technology handwritten digits database for clustering handwritten digits (0-9) images. Each algorithm was executed twice: one with a small dataset of 100 images randomly selected from the whole database; another one with the whole dataset of 300 images. Each cluster is expected to contain all the images from the same digit number (handwritten character). Fig 5 below illustrates the samples of the clustering results that were obtained from implementing the four algorithms on the handwritten dataset. From Fig. 5a, the automatic partitional clustering algorithm demonstrated high performances in clustering the handwritten images by categorizing the digits from 0-9 into 10 correct clusters. However, the results obtained by the classification accuracy is low by the automatic fuzzy clustering algorithm for the dataset among the four algorithms. We noticed that the FCM algorithm resulted in low performances when it dealt with the handwritten images. The accuracy was 68% and 60%, respectively, when the sizes of the datasets are 100 and 300 images. However, the dataset size and the type of a dataset did not have significant impacts on the execution time of the non-fuzzy algorithms; automatic paritional clustering and parititional K-means clustering algorithms. The automatic partitional clustering algorithm classified the images that belong to one digit into one cluster. It generated 10 clusters correctly with each cluster representing 30 images of one same digit according to the results in Table 2. The other three algorithms resulted in clusters with mixed images from different digits.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figs. 6 and 7 present comparisons among the four algorithms based on the execution time and the classification accuracy against dataset size. From Fig. 6 the automatic partitional clustering algorithm and prtitional k-means can achieve high accuracy compared with others using all images in dataset-2. In terms of execution time, Fig 7 shows the comparison among the four algorithms. We can notice that the non-fuzzy algorithms yield the lowest execution time compared with the fuzzy algorithms. The four algorithms achieved 80%, 74%, 70% and 61%, respectively, after applying 10-cross validation. Table 2 illustrates the comparisons among the four algorithms based on cluster number. From Table 2 we can notice that the partitional k-means and automatic partitional clustering algorithms obtained the minimum number of clusters.

From the obtained results we can find that:

(1) The automatic partitional clustering algorithm gives better performances than other three algorithms, with the lowest execution times and the highest classification accuracies for the two datasets.

(2) The performances of the automatic fuzzy algorithm and the FCM algorithm are more sensitive to the sizes of datasets and the types of images. For the handwritten digits images database, the clustering accuracies of the two algorithms decreased.

(3) In general, based on the results in Tables 1 and 2 the non-fuzzy algorithms are more suitable for different types of image datasets.

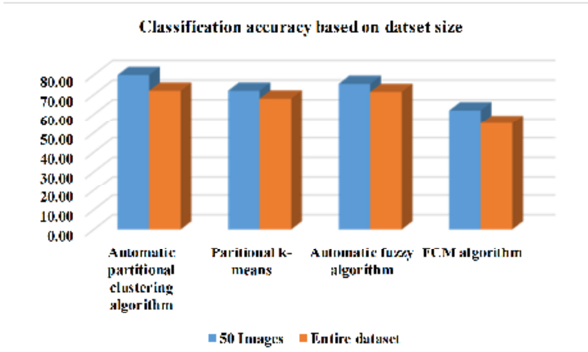


Fig. 6. Classification accuracy of the four algorithms

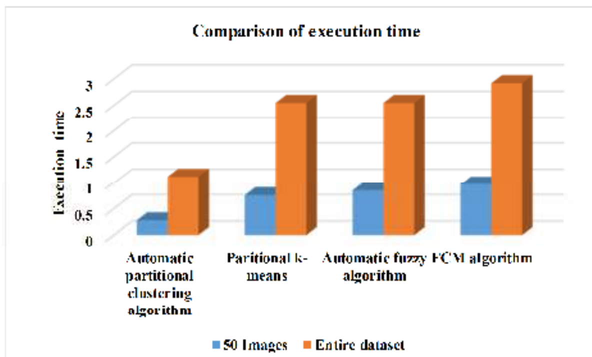


Fig. 7. Comparison of execution time among the four algorithms

Table 2 Comparison Between The Four Algorithms

The algorithm	Dataset size	The actual number of clusters	The obtained number clusters
Automatic partitional clustering algorithm	100	10	12
	300	30	32
Partitional K-means clustering algorithm	100	10	11
	300	30	35
Automatic fuzzy algorithm	100	10	13
	300	30	33
FCM algorithm	100	10	13
	300	30	38

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7.4. Experiment 3

Scene images clustering is considered a very challenging problem in computer vision and classification process [4, 6, 25, 28].

In this section we present the performances of the four algorithms using the scene database, consisting of three categories of images from forests, tall buildings, and coasts. For each category 100 images are selected as the experimental image data. The images of all the three categories are given as the input. The images are to be classified into three clusters of coasts, forests, and tall buildings. The purpose of this experiment is to investigate the performances of the four algorithms on complex datasets. Figs. 8 shows the experimental results.

The experimental results show that the non-fuzzy clustering algorithms produce reasonable results. Fig. 8a and Fig. 8b show that the automatic partitional clustering algorithm and the automatic fuzzy algorithm can classify the images that belong to one category into one correct cluster. As shown in Fig. 8b the partitional K-means results in a small rate of errors for grouping two forest images into the coast images cluster. The results in Fig. 9 and 10 demonstrate that the automatic partitional clustering algorithm remains having better performances than other three algorithms, whilst the fuzzy C-means algorithm results in the lowest accuracies. In addition, the non-fuzzy algorithms: the automatic partitional clustering algorithm and the partitional k-means algorithm were recorded relatively lower execution times, whilst the fuzzy C-means has the highest execution time. The comparison in terms of accuracies are presented in Fig. 9. The fuzzy algorithms achieve the highest accuracy compared with non-fuzzy algorithms. That proves the ability of the fuzzy algorithms to classify different types of images. From Fig. 10 the automatic partitional clustering algorithm and partitional k-means algorithm record the lowest execution time although different types of images are used.

The 10-cross validation results from the four algorithms were 79%, 70%, 69% and 58% respectively. Table 3 shows the comparison among the four algorithms. The lowest cluster number was obtained from automatic partitional clustering algorithm while the partitional k-means clustering algorithm outperformed the other algorithms.

Table 1 provides more details about the comparison results in term of number of clusters.

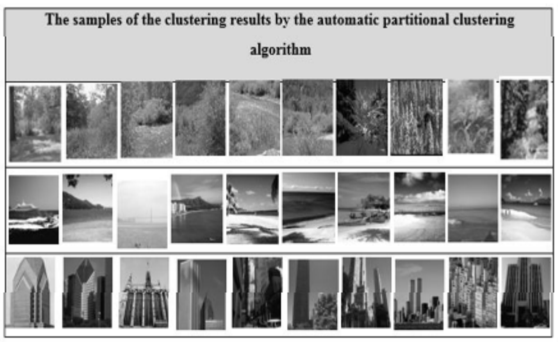
From the obtained results we can find that:

1) Comparing with the results from Experiments 1 and 2, all the four algorithms resulted in lower accuracies for this complex scene images database. In general the automatic partitional .
2) clustering algorithm still provides better accuracy results than the other three algorithms.

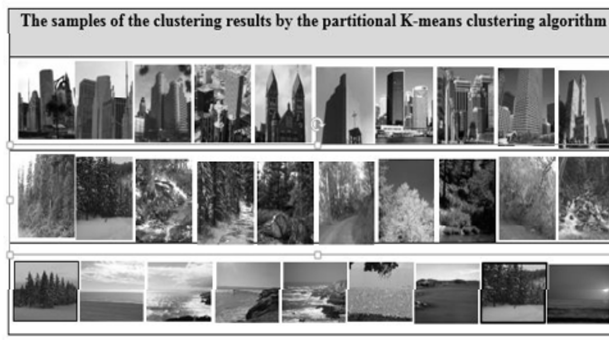
3) The partitional clustering algorithm records a smaller execution time of 0.2810s and 1.124s when the number of samples are 100 and 300, respectively. This indicates that the size of images doesn't have significant effects on the speed of the algorithm.
4) When comparing the fuzzy algorithms, the non-fuzzy algorithms give better results for different types of databases and the size of the datasets, with lower average execution times.

Table 3 Comparison among the four algorithms

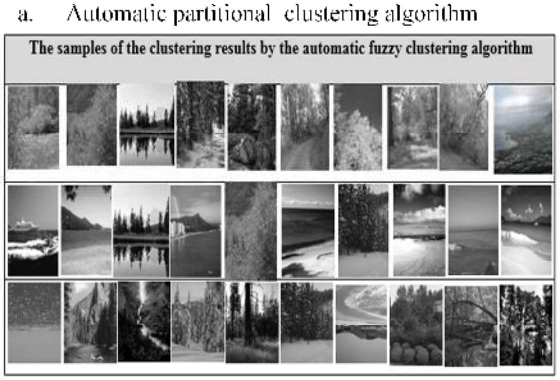
The algorithm	Dataset size	The actual number of clusters	The obtained number clusters
Automatic partitional clustering algorithm	100	10	12
	300	3	3
Partitional K-means clustering algorithm	100	10	15
	300	3	3
Automatic fuzzy algorithm	100	10	16
	300	3	4
FCM algorithm	100	10	14
	300	3	4



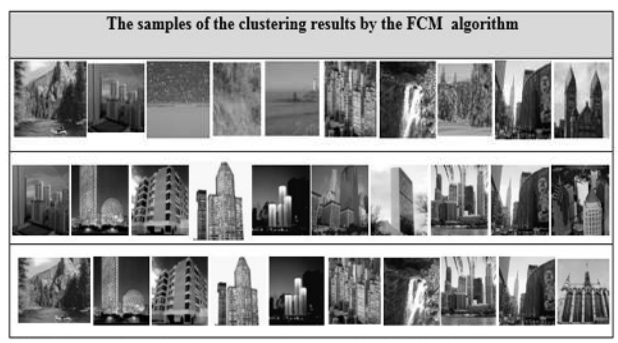
a. Automatic partitional clustering algorithm



b. Partitional k-means clustering algorithm



c. Automatic fuzzy clustering algorithm



d. Fuzzy c-means clustering algorithm

Fig. 8. The Samples Of Clustering Results Obtained By The Four Algorithms For The ~~handwritten~~ Natural Scene Images

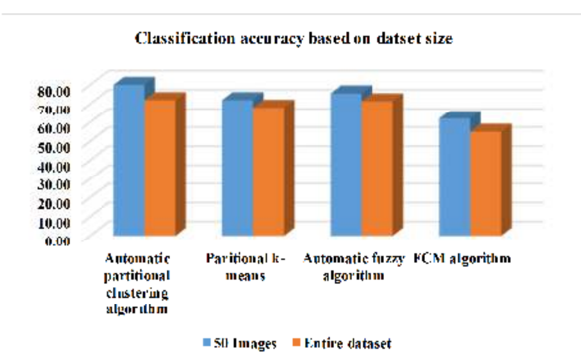


Fig. 9. Classification accuracy of the four algorithms

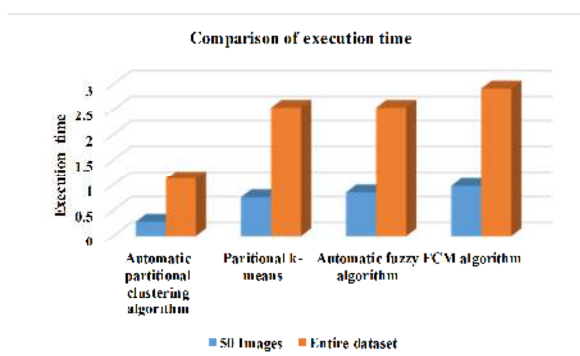


Fig. 10. Comparison of execution time among the four algorithms

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

8. CONCLUSIONS:

This paper evaluates the performances of fuzzy and non-fuzzy clustering techniques through four algorithms: the automatic partitional clustering, partional k-means clustering, automatic fuzzy clustering and the fuzzy C-means clustering algorithms. Three image databases are used in the experiments. From the results in Experiments 1 to 3, we can conclude the following:

1) The automatic partitional clustering and partitional k-means clustering algorithms give paritional clustering, namely: each image is classified into one and only one correct cluster. The algorithms provide correct clustering results.
2) The automatic partitional clustering algorithm results in a reasonable execution time during the three experiments with different types of databases. It is more feasible with large datasets.
3) The automatic fuzzy algorithm and the FCM algorithm result in overlapping clusters, with one type of images may be classified into different clusters.
4) The automatic partitional clustering algorithm and the automatic fuzzy algorithm are automatic types of clustering algorithms - the number of the clusters are obtained automatically through the implementation of the algorithms. The partitional k-means clustering algorithm and the FCM algorithm are non-automatic algorithms as the number of obtained clusters are given a priori.
5) The automatic clustering algorithms whether partitional or fuzzy, generally give a low execution time, and acceptable clustering accuracy results.
6) The partitional k-means clustering algorithm is suitable and provides a good results with the large
7) Imaging datasets, but the accuracy of the algorithm would be decreased for different types of images.

REFERENCES:

[1] Abbas, O.A. 2008. Comparisons between data clustering algorithms. Int. Arab J. Inf. Technol 5, 320-325.
[2] Banerji, S., Sinha, A., and Liu, C., 2012. Scene image classification: Some novel descriptors, Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on. Publishing: 2294-2299.
[3] Bataineh, K., Naji, M., and Saqer, M. 2011. A comparison study between various fuzzy clustering algorithms. Editorial Board 5, 335.
[4] Bautell, M., Luo, J., and Gray, R.T. 2003. Sunset scene classification using simulated image recomposition, Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on. Publishing, I-37-40.
[5] Bezdek, J.C., Ehrlich, R.,z nd Full, W. 1984. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences 10, 191-203.
[6] Cai, D., He, X., & Han, J. 2005. Document clustering using locality preserving indexing. Knowledge and Data Engineering, IEEE Transactions on 17, 1624-1637.
[7] CDROM, C., 1992. 1, USPS Office of Advanced Technology. Publishing.
[8] Chattopadhyay, S., Pratihar, D.K., & Sarkar, S.C.D. 2012. A comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms. Computing and Informatics 30, 701-720.
[9] Elavarasi, S.A., Akilandeswari, J., and Sathiyabhama, B. 2011. A survey on partition clustering algorithms. International Journal of Enterprise Computing and Business Systems 1.
[10] El-Hamdouchi, A., and Willett, P. 1989. Comparison of hierarchic agglomerative clustering methods for document retrieval. The Computer Journal 32, 220-227.
[11] Fritz, H., García-Escudero, L.A., and Mayo-Iscar, A. 2013. Robust constrained fuzzy clustering. Information Sciences 245, 38-52.
[12] Fung, G., 2004. A Comprehensive Overview of Basic Clustering Algorithms. Publishing.
[13] Garces, E., Munoz, A., Lopez-Moreno, J., and Gutierrez, D., 2012. Intrinsic images by clustering, Computer Graphics Forum. Publishing: 1415-1424.
[14] Jain, A.K., Murty, M.N., and Flynn, P.J. 1999. Data clustering: a review. ACM computing surveys (CSUR), 31, 264-323.
[15] Namratha, M., and Prajwala, T.R. 2012. A Comprehensive Overview of Clustering Algorithms in Pattern, Journal of Computer Engineering Recognition 4, 6:23:30.
[16] Kaur, M., and Kaur, U. 2013. Comparison between k-means and hierarchical algorithm using query redirection. International Journal of Advanced Research in Computer Science and Software Engineering, 3.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[17] Kim, T.-h., Bhattacharyya, D., and Bandyopadhyay, S.K. 2011. Supervised chromosome clustering and image classification. Future Generation Computer Systems 27: 372-376.
[18] Lazebnik, S., Schmid, C., and Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. Publishing, 2169-2178.
[19] Li, Y., Li, P., Wu, B., Jiao, L., and Shang, R. 2013. Kernel clustering using a hybrid memetic algorithm. Natural Computing 12, 605-615.
[20] Samaria, F.S. 1994. Face recognition using hidden Markov models. University of Cambridge.
[21] Sarsoh, J.T., and Hashem, K.M. 2012. Classifying of Human Face Images Based on the Graph Theory Concepts. Global Journal of Computer Science and Technology 12.
[22] Sarsoh, J.T., and Hashem, K.M. 2007. Clustering of Human Face images with different rotation angles.
[23] Wang, Y., and Pan, Y. 2014. Semi-supervised consensus clustering for gene expression data analysis. BioData mining, 7: 1-13.
[24] Wu, K.-L., and Yang, M.-S. 2002. Alternative c-means clustering algorithms. Pattern recognition 35: 2267-2278.
[25] Xia, K., Wu, Y., Ren, X., and Jin, Y. 2013. Research in Clustering Algorithm for Diseases Analysis. Journal of Networks 8: 1632-1639.
[26] Yu, J., Hong, R., Wang, M., and You, J. 2014. Image clustering based on sparse patch alignment framework. Pattern Recognition 47: 3512-3519.
[27] Zadeh, L.A. 1965. Fuzzy sets. Information and control, 8, 338-353.
[28] Zang, M., Wen, D., Wang, K., Liu, T., and Song, W. 2015. A novel topic feature for image scene classification. Neurocomputing 148:467-476.
[29] Zhang, D.-Q., and Chen, S.-C. 2004. A novel kernelized fuzzy c-means algorithm with application in medical image segmentation. Artificial Intelligence in Medicine 32: 37-50.
[30] Zhang, Y., Wang, W., Zhang, X., and Li, Y. 2008. A cluster validity index for fuzzy clustering. Information Sciences 178:1205-1218.