Design A System For Image Matching By Using Fuzzy C-Means Clusters

Ali A. Yassin
Department of Computer Science
College of Education
University of Basrah

Essam Z. Ebady
Polymer Research Center
University of Basrah

Aqeel A. Yassin
Islamic University in Lebanon

الخلاصة :

يهدف هذا البحث إلى الاستفادة من إمكانية خوارزمية (FCM) ذات التجمعات العنقودية المضببة في مطابقة الصور نتيجة لكونها تحمل صفات المنطق المضبب الذي يتمتع بإمكانية عالية وقوة في التعامل مع المسائل المعقدة فضلا عن كونه ا تستطيع ان تثير الغموض والابهام بفعالية عالية . كما و تتمتع التجمعات العنقودية بإمكانية التعامل مع مجموعة من البيانات المنتخبة عبر كتلة البيانات التي تتعامل معها وبذلك تعطي دقة بالنتائج وتقلل من حجم البيانات التي سوف نتعامل معها ، هذه البيانات في مجال البحث سوف تتجه نحو بيانات الصورة ال تي سوف نختار مجموعة من بياناتها كممثلات رسمية عن الصور التي تخضع لنظام المطابقة ثم بعد ذلك تتفذ عليها السياسة المتبعة في عملية التطابق للحصول على النتائج فضلا عن كوننا نقوم بعملية الدمج في طريقة الانتخاب العنقودي المضبب في اختيار بيانات نموذجية من نقاط تقع على حواف الصورة وليس بيانات الصورة بشكل أساسي. هذا ما تمت مناقشته في البحث.

Abstract

This research aims to take advantage of the possibility of an Fuzzy cluster means (FCM) algorithm with a cluster fuzzy groupings in matching images as a result of carrying fuzzy logic specification who enjoys high potential and strength in dealing with complex issues as well as being able to manage uncertainty and ambiguity effectively high. And also enjoy the possibility of clusters dealing with a range of data across the block selected data dealing with it and thus give accurate results and reduce the volume of data that will deal with this data in the area of research will be moving toward a data image that will choose a set of

formal statements by representatives of the photos that are subject to the system of matching and then That it implemented the policy in the process of matching to get results as well as the fact that we merger process between the method of election to choose a fuzzy cluster data model of pixels located on the edges of the image and not the image data is essential and this is what will discuss it later.

Keywords: fuzzy c-means, fuzzy clustering, classification, pattern recognition.

1- Introduction

One of the classic problems in computer vision is the face recognition problem. In general this problem can take on a wide variety of forms, Therefore the work of a number of researchers at several methods for the purpose of the work algorithms or software simulates the issue of discrimination Faces.

Some used algorithm genetic Some of them used the format engineering bilateral or three-dimensional, as well as several other ways. In addition to the progress in each of them must deal with a database containing the information to be discrimination with image.

The first part of this research includes the definition of matching images and the basic structure of the system of images matching The second part is knowing and cluster analysis algorithm (FCM) and third Part contains a general structure of the system using the method of matching. Fourth Part contains the use of two identical images in the first deal with the image data in general and then make them pay the basic environment for algorithm to select a set of pixels of each image in a match to become the official representative of those images to make the matching process, second method will be interest from some features a image extraction pixels that fall on photo Edge, where will we discuss here the issue of the fact that the pixels which will deal with FCM algorithm lies on edge of image, which are either black or white as the application of an algorithm calculating the edges of the image, where would we explain how their dealings with the fuzzy logic and then we matched based on the way, where note the difference in the results of both methods through a series of experiments,
which will adopt. And then we shall discuss the results and future work.

2- literature survey

The term *cluster analysis* (first used by Tryon, 1939) actually encompasses a number of different classification algorithm. A general question facing researchers in many areas of inquiry is how to *organize* observed data into meaningful structures, that is, to develop taxonomies[1,2]. For example, biologists have to organize the different species of animals before a meaningful description of the differences between animals is possible. According to the modern system employed in biology, man belongs to the primates, the mammals, the amniotes, the vertebrates, and the animals. Note how in this classification, the higher

the level of aggregation the less similar are the members in the respective class. In fact, cluster analysis is not as much a typical statistical test as it is a "collection" of different algorithms that "put objects into clusters." The point here is that, unlike many other statistical procedures, cluster analysis methods are mostly used when we do not have any a priori hypotheses, but are still in the exploratory phase of our research. In a sense, cluster analysis finds the "most significant solution possible. Clustering techniques have been applied to a wide variety of research problems. Hartigan (1975) provides an excellent summary of the many published studies reporting the results of cluster analyses. For example, in the field of medicine, clustering diseases, cures for diseases, or symptoms of diseases can lead to very useful taxonomies. In the field of psychiatry, the correct diagnosis of clusters of symptoms such as paranoia, schizophrenia, etc. is essential for successful therapy. In archeology, researchers have attempted to establish taxonomies of stone tools, funeral objects, etc. by applying cluster analytic techniques[2,3].

3- Cluster analysis

Basically, cluster analysis is able to recognize patterns by which the data is grouped, i.e. its aim is to partition a given set of data or objects into clusters (subsets, groups, classes). Figure 1 shows a simple 2-dimensional space with several points in it. For humans, it is obvious that one can group those points into two classes1. A data/cluster analysis algorithm should be able to perform the same task, i.e. search for a model or a structure with analyze data by merely looking at it. the data. We can easily think of a simple way of doing this. The Euclidean distance between points can be used in order to determine a measure of their dissimilarity. The more two points are away from each other, the more probable it is that they fall into deferent clusters. Thus, we can consider this process as an optimization problem. An objective function is used to assign each possible cluster partition a quality or error value that has to be minimized or maximized. An ideal solution is the cluster partition that obtains the best evaluation. This technique is called *objective function based cluster analysis*[3,4].
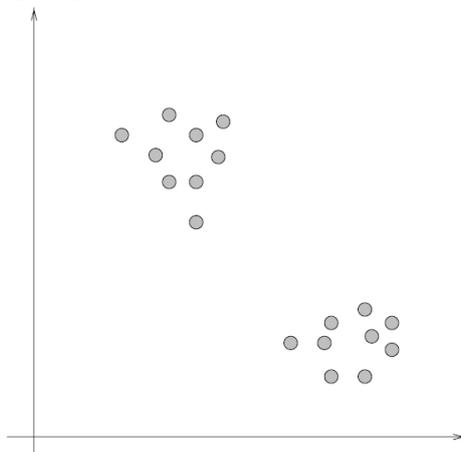
Figure (1): show two clusters of points in a 2-dimensional space

4- Fuzzy clustering

It is often the case that the data we want to partition can not be classified into fixed clusters. Especially in real-world problems, clusters almost always overlap. So a hard partition, as it is called, is not possible. This is where a fuzzy technique can help. Instead of grouping each object into fixed clusters, we can use a mapping on our data that assigns values between 0 and 1 to every object[4,5]. This creates a fuzzy set and the values are the degree of membership of a certain cluster. Let *X* be the set of points from figure 1. If we search for two clusters, a good fuzzy set *C*1 would map the points from the upper-left cluster to values near 1 and points from the lower-right region to values near 0, whereas the second set *C*2 would do the opposite, i.e. give lower-right points higher and upper-left points smaller values[5].

5- The fuzzy c-means algorithm

The fuzzy c-means algorithm was one of the first clustering techniques whose final version was presented by Bezdek in 1973. It is an iterative clustering approach that finds spherical clouds of points in a *p*-dimensional space. There are two constants that have to be specified at the beginning, namely a number of clusters *c* we want to search for and a value for the *fuzzifer m*. The latter is a measure of how much the membership value of every vector in the search space is blurred" around the centers of the clusters. The greater *m*, the fuzzier the results will be. If we choose *m* = 1, we would get a hard

partition of the data, whereas a high value for *m* is used when we know that the clusters mostly overlap and are hardly distinguishable. This value is commonly set to 2. The main part of the fuzzy c-means algorithm is the objective function. The algorithm tries to minimize that function which keeps track of all distances of points to their prototype. A prototype is the name for the center of a cluster in analysis space. So, in the end, the algorithm finds the *c* centers in analysis space such that all points of that space have minimal distances to their prototypes [6].

6- Image Matching

Image Matching: - is the process of finding similarities between the two percentage or more. The process of matching an important pixel in many areas of computing vision where there are many techniques that stem from identical images, as follows [1]:
1- Identical fingerprints.
2- Matching faces and some parts of the body.
3- Control devices which are directed at the gates of airports and skyscrapers, which relies principle to the matching process.
4- In the medical field.

Thus, the identical images is a very important component in applications that require comparable images based on their components. There are two important phases in matching images [1,2,6,7]:

The first type space representation of images, The second measure of the proportion of congruence between the images intended for submission to the process of matching.

There are many measures of matching used in this area which is dependent on low levels in operations such as compatibility matching calculate the difference between the images or matching pixel to pixel where this treatment requires a relatively long time as well as a fall in parts of the image be unnecessary in the process of matching. The average levels that depend on the reconciliations are the qualities to draw images to be subject to the process of matching such as determining the pixels located on the edges of images or more pixels intensity lighting (Brightness) or pixels more variation (Darkness). Finally, high level of representation by Depends on the marks given to each type of images may sometimes depend on the color or intensity levels as well as lighting techniques rely on accurate calculations (Soft Computing), which is represent a fuzzy logic, neural networks and groupings fuzzy cluster. The basic structure of the system of matching as show in figure (2) [1,2,3,7]: -

```
┌─────────────────────┐
│    Input Images     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Preprocessing    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│   Fuzzy Clustering  │
│      C-Means        │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│   Image Matching    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Output Result    │
└─────────────────────┘
```
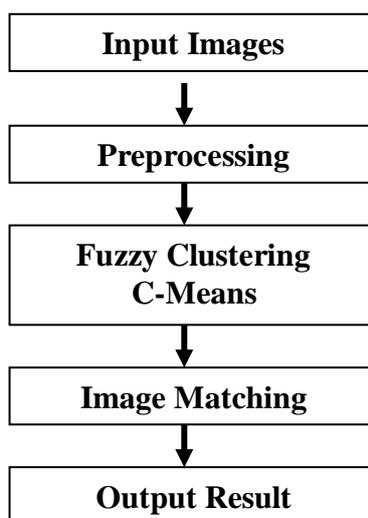
Figure (2): The general structure of the system matching images using a FCM method

Where: --

Input Images: mean the input of all the images that chosen to process of matching, which some of which may be different in size or prone to reflection and rotating operations or noise... etc.

Preprocessing: in this segment will implement a series of initial treatment, which means the unification of the volume of images the same size image basic rules as well as some simple operations such as removing some noise from the image and Sleeking and balancing discrepancy.

Features Extraction: This section contains the qualities to draw in the image (such as the noise control, rotating, identifying terminals pixels, class photo... etc). The process of normalization (it means the measure of image then compared with those whose standard features, which are extracted from the image of the system. After these operations coming the process of creating the image data drawn advantages and expressed either a matrix or reading areas include adjacency or most gray or rely on the areas that fall on the edges of the image and therefore will attend the data ready for matching called (Data Provide).

Matching: is the mechanism used to calculate the discrepancy between input images and essential image in the system. Where the essential image in the system is the image that will take place the rest of the process of matching images, as possible on the basis that we have found more than one image in the system within a database containing a set of data was The data staff in an institution or company, especially banks and airports to identify each employee through the introduction of its own data and then identify that person across the identical image and this process is very effective and important especially in the security institutions and criminal [7,8].

7- The Clusters Analysis with image processing

The Cluster analysis used in many scientific areas are characterized by patterns, drawing attributes, data compression... etc.[3,4]. The cluster analysis is a branch of the exact calculation (Soft Computing) where the division of data into clusters so that each cluster gathered a group of comparable data. Modus operand of cluster analysis algorithm is to find the centers of clusters to reflect on those Clusters and finally reflect the input data in general [3], data which provided to the algorithm FCM in this research are represent the photo required identification and then after that we dealt with using the appropriate method of matching them.

7-1 Using Cluster Analysis in Image Matching

The cluster analysis division of data to a cluster groupings such a way that each collect a set of comparable data, which represents the elite, selected from the data that is focused on the task of data and this provides the speed and accuracy in treatment outcomes. Modus operandi of cluster analysis algorithm is finding a center of clusters for use data are essential for the treatment process, where we adopted in the search algorithm (Fuzzy C-means), which is characterized as being used fuzzy logic in the field of treatment [4,5,8]. Where the limits of the system is designed to allow for overlapping data elements belonging to collect more than cluster fuzzy situational relying on a function of belonging used in the algorithm.

7-2 Algorithm of Clusters Fuzzy (FCM)

This algorithm depends on the division of data foggy, the borders between the overlapping of clusters, because each element will belong to more than combining clustering, but to varying degrees based on the outcome of a function of belonging used. The basis of the work of this algorithm is to reduce the objective function.

(J) Objective Function as in the equation (1) [6,7,9] : -

$$J = \sum_{I=1}^{C} J_I = \sum_{I=1}^{C} \sum_{J=1}^{N} U_{ij}^{M} d_{ij}^{2} \qquad \qquad ...(1)$$

Where: --
C: - Number of Clusters.
N: - Number of inputs.
Uij: - Matrix affiliation between (0,1).
m: - fuzzy measure.
dij: - the distance between destined j with Cluster i.
Our way in the search relied on the law of Euclid in calculating distances (Euclidean distance) as in the equation (2): -

$$dij = d\,(XI,\ Yj) = [\sum_{P=1}^{P} (\mathbf{X_{ip}} - \mathbf{Y_{jp}})^2]^{1/2} \qquad ...(2)$$

Going along with P (Xj,, Yj).
Cj represent a center, calculated as the equation (3): -

$$C_j = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}\, Xi}{\sum_{i=1}^{n} U_{ij}} \qquad ...(3)$$

Then calculates the value of new Uij by equation (4): -

$$Uij = \frac{1}{\sum_{k=1}^{m} \left(\dfrac{\|x_i - C_j\|^2}{\|x_i - C_k\|^2} \right)^{\frac{2}{m-1}}} \qquad ...(4)$$

Repeated the process leading to $Uij^{K+1} - Uij^{K} < E$, where K a number of times the way repetition of the condition above, the E representing the small value between of '0 'and'1' and possible expression of the algorithm as follows [7,8]: -

Algorithm

1. *Initialize U=[Uij] Matrix, U(0)*
   At K_step : Calculate The Centers Vectors

2. $C^{(k)}$ *= [Cj] With* $U^{(k)}$*;* C_j *Is Computed In Equation (3)*

3. *Update* $U^{(k)},U^{(k+1)}$ *;*U_{ij} *is Computed In Equation (4)*
   If $\| u^{(k+1)} - u^{(k)} \| < E$ *Then Stop ; Otherwise return to Step2*

8- Image Processing and Fuzzy C-means Clustering Method (FCM)

When using an algorithm FCM in identical images, this algorithm will divide the image to the cluster groupings overlap to allow each element in the image that belongs to collect more than situational through the creation of every degree of affiliation collected and then finding algorithm based center of clusters so as to reflect each Center for the assembly cluster which belongs to it, use these basic data center in the process of matching images. This means that the number of data used in the matching process will shrink to a large extent, example if we had the image size (512 * 512) pixel, the algorithm FCM will define for example (300) pixel Center used data are essential in matching the image [2,8,9].

9. Image Matching By Using Fuzzy C-Means (FCM)

Most algorithms for matching images of all data dealing with the enter image, including what divides the image to the group of blocks based on certain specifications which requires a relatively long time deal depends on the size of the image that deal with it, at this algorithm would have benefited from FCM algorithm, which is one of roads cluster used In the different areas where data will be entered into which will operate in this algorithm is the basic image of the essential system which will depend on the selection of centers, through which builds the matrix model representing selected pixels of the image representative of the rest of the other pixels, when using an algorithm FCM Matching images in this algorithm will divide the image to the cluster groupings overlap to allow each element that belongs to collect more than situational through the creation of a membership application across each collected a function of certain affiliation. Then the algorithm to find centers of clusters so as to reflect on each centre on the Rally Cluster, which belongs to him. These centers used groupings data are essential in the process of matching the images, this means that the number of data used in the matching process will shrink to a large extent, example if we had the image size (100*100) and we are want to finding (200) Center to reflect the image data in the

treatment process will We deal with recent data that reflect the basic image data as in Figure (4), but the overall structure of the system of matching images using a method FCM show as figure (3)[2,7,9]: -
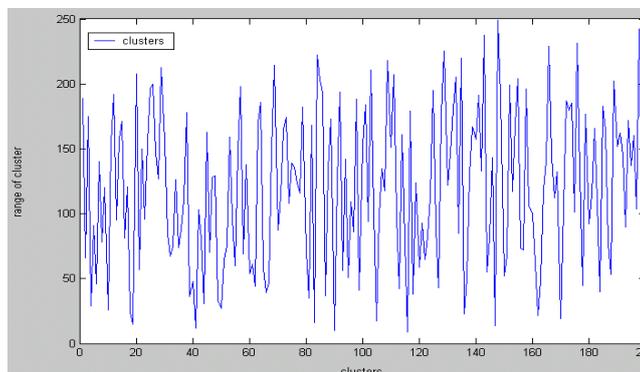
Image   *Cluster Of Image*

Figure (3): which shows how C-Means fuzzy on the image

The basic algorithm· to calculate the discrepancy between the images using FCM are as follows: -

Image Matching With Fuzzy Clustering Algorithm
Inatialize U=[Uij] Matrix,U$^{(0)}$
ImageRead(im1,im2,...,imn)
[N M] = Size(im1)
ImageResize(im2,im3,...,n,N,M)
Vector (im1,im2,...,imn,n,N,M,X1,X2,...,Xn)
For i:= 1 To n Do

Step2 :
At K-Step Calculate the Centers Vector C$^{(k)}$ = [Cj] With U$^{(k)}$

$$Cj = \frac{\sum_{i=1}^{n} U_{ij}^{m} \cdot Xi}{\sum_{i=1}^{n} U_{ij}^{m}}$$

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

$$U_{ij} = \frac{1}{\sum_{i=1}^{c} \left(\frac{\|Xi-Ci\|^2}{\|Xi-Ci\|^2} \right)^{\frac{2}{m-1}}}$$

If $\| U^{(k+1)} - U^{(k)} \|$ $< E$ Then Stop

Else Return To Step2
End if
Center[i] = Cj ; j=1,2,...,m
Matching (Center,m,n,R,Index)
Output "Ratio =",R,Index
End Image Matching With Fuzzy Clustring Algorithm

10- Image Matching by using FCM And Edge Detection Algorithm

The edges of the image is a set of pixels located on the edges of the image is calculated based on a certain threshold value, as follows:

If the pixel is greater than that threshold carrying value 1, which represents the white color, else when they carry value (0), which represents black color. The image resulting from the edges of the image are calculating the values of either black pixels (0) or white (1) and cluster analysis algorithm FCM used fuzzy logic in the transformation of values between zero and one?

We benefited from the algorithm for calculating the end of the image but instead of dealing with the values of either zero or one, we have a pixel about the value of compensation that are greater than the threshold value instead of the one with natural value in the form and format include Gray, figure (4) shows that: -
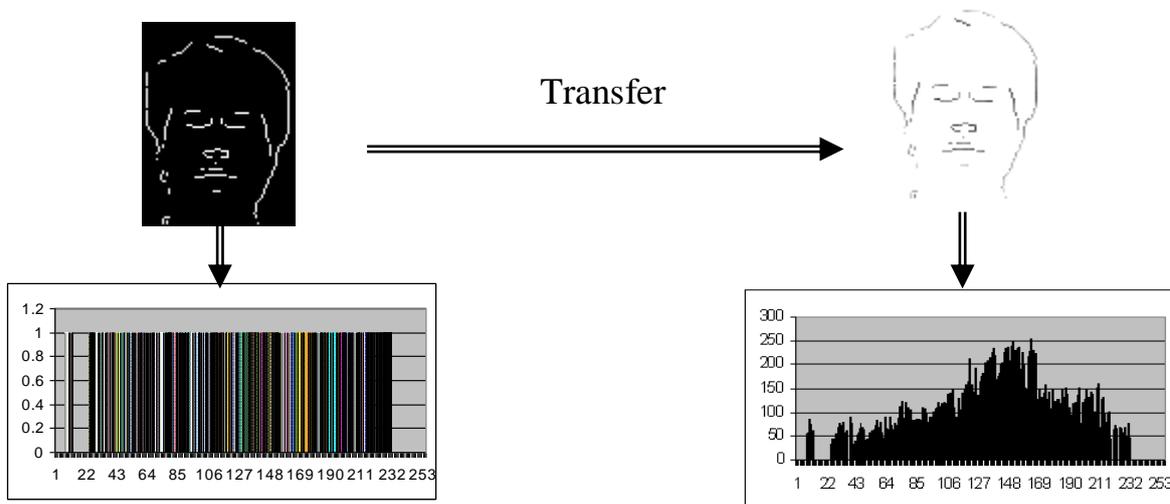
Figure (4): Explain the difference in the distribution of color when the difference in the amount of compensation expanse in the last image

Characterized a system identical photos by using an algorithm fuzzy FCM amended on photos edge shown in Figure (5), by high speed and accuracy of the result where the will be divided image data to a group of Clusters, where similar pixels of the image belongs to the same cluster

and different pixels belong to different clusters, These Clusters represent the summery of collected those pixels. Data will be built Clusters which instead of being all pixels of the image as in the previous algorithm will be located the pixels on the borders of the image, as in Figure (6) which illustrates how the distribution and concentration of cluster in (25) Center for the edges of the image.

Input Images

↓

Preprocessing

↓

Image Edging

↓

Fuzzy Clustring
C-Means

↓

Image Matching

↓

Output Result
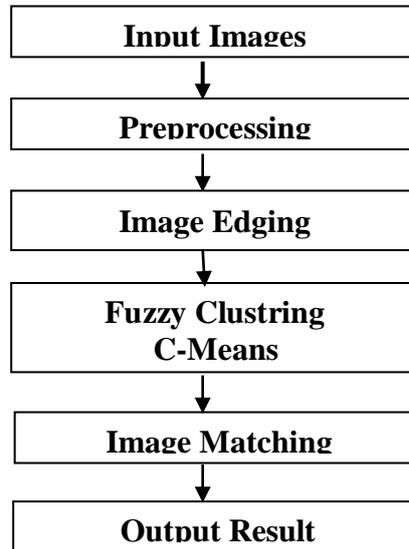
Figure (5): Explain the mechanism of algorithm matching using photo Edge with   FCM method

Image

conver

Cluster Of Image
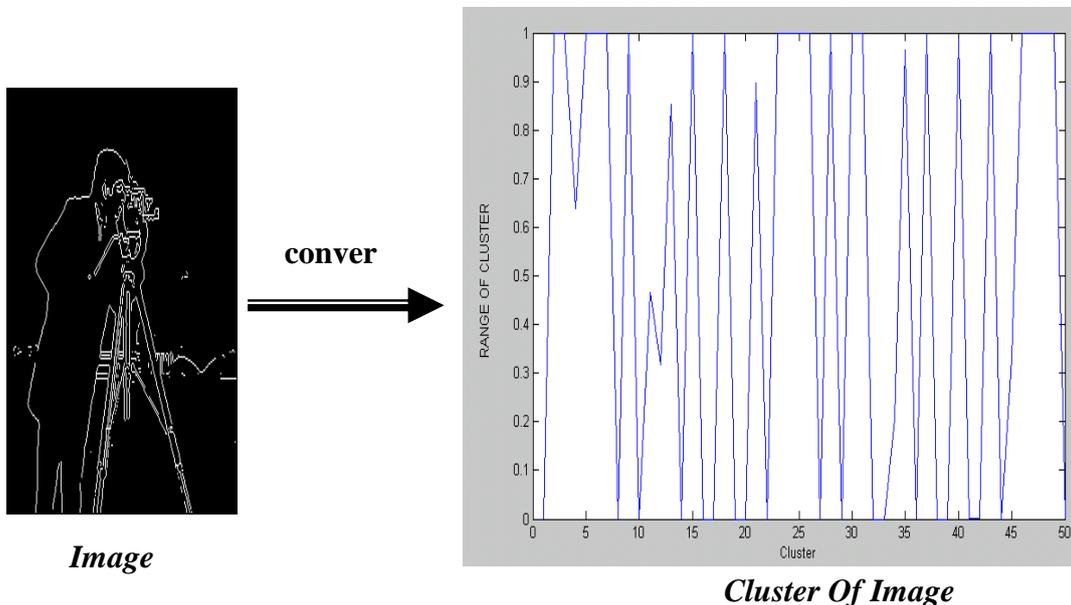
Figure (6): Show the FCM method on the image edge

11- Experiential Results
Experience (1)

In this paragraph will take group of different images in terms of the proportion of noise and note the results of applying both algorithms

which are illustrated in table (1) on a series images in the form (7). Where the first photo image is stored in the system and the rest of images represent cases different input and then compared the results between the algorithms across the table (1).

Table (1): Show experiential Results to both algorithm in figure (7)

Algorithms	Image1	Image2	Image3	Image4	Image5
1st Algorithm	Image found in the system	95.833	92	66.3	77
2nd Algorithm	Image found in the system	74.576	50	51.64935	63

Iamge1 *Image2* *Image3* *Image4* *Image5*

Figure (7) Show noise ratio in different images

Experience (2):

In this paragraph will take a group of different images in terms of addition texts to images and note the results of applying both algorithms which are illustrated in Table (2) to set images in the figure (8).

Table (2): Show experiential Results to both algorithm in figure (9)

Algorithms	Image1	Image2	Image3	Image4	Image5
1st Algorithm	Image found in the system	100	100	85	0
2nd Algorithm	Image found in the system	100	100	87	0

Image1   Image2   Image3   Image4
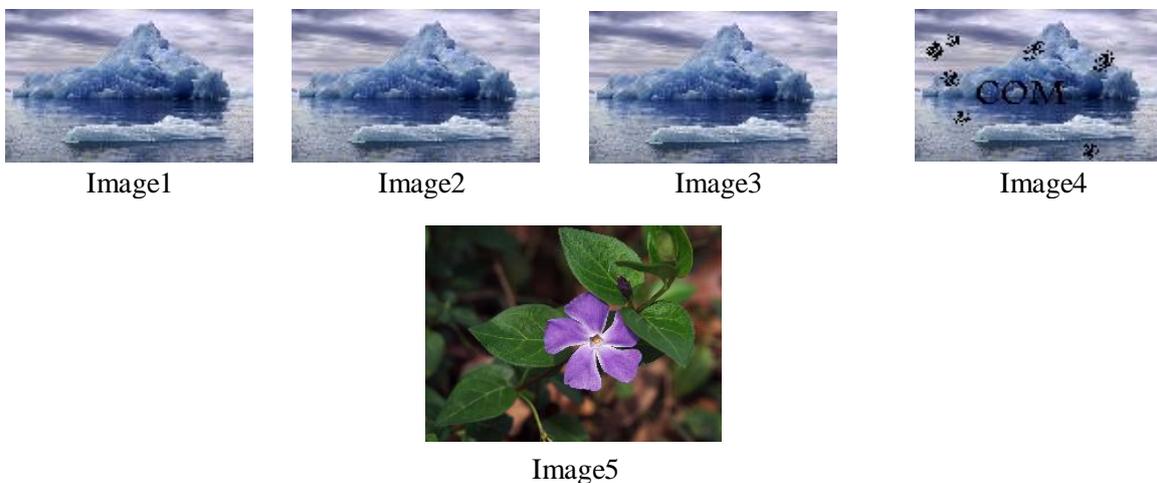
Image5

Experience (3):

Figure (8): Show group of input images in the algorithms

In this paragraph will take a group of different images and note the results of applying both algorithms which are illustrated in Table (3) to set images in the form (9).

Table (3): Show experiential Results to both algorithm in figure (10)

Algorithms	Image1	Image2	Image3	Image4	Image5
1st Algorithm	Image found in the system	100	43.0003	0	0
2nd Algorithm	Image found in the system	100	31.0001	0	0

Image1

Image2

Image3

Image4

Image5

Figure (9): Show group of input images in the algorithms

8- <u>Conclusion</u>:

By the results we have had on these algorithms, and through using an extensive database on the additional tests as well as the theme chosen for the research noted the following:

1) The establishment of a hybrid system depends on the draw qualities of the image and use sophisticated methods to elect a tremendous amount of data as is the case in the photos and used in determining the identity or characterize a particular person is very effective and distinctive.

2) The speed in process and accuracy of results, The speed based the election of a part of the image data, then used in the matching

process and precision built on draw qualities of the image and focus on important areas rather than dealing with all the image data.

3) Algorithm which relies on identifying the edges of the image in this research included a two important partial must be conducted first mentioned the development of the algorithm edges of the image, through the production of the image containing black areas and areas include gray instead of white to enable them to deal with the algorithm FCM This is the second partial. This algorithm is better than the first by the results we have obtained through a series of tests.

REFERENCES

1) Christian B. U. Perwassa, G. Sommerb, "A Fuzzy Logic Algorithm for Dense Image Pixel matching", Institut f ¨ ur Informatik CAU Kiel, 24105 Kiel, Germany.
   a-christian@perwass.com
   b-gs@ks.informatik.uni-kiel

2) HRTizhoosh, "Fuzzy image processing", University Of Waterloo. WWW.Pami.waterloo.ca / tizhoosh / wfip.htm

3) Hong J. 2002, "Generalized Fuzzy Clustering Model With fuzzy C-means", Computer Science and Engineering University Of South Corolina, Columbia, www.Case.Sc.edu

4) Martin. ES, "Pap-Smear Classification", Orested-Dtu, Automation Beth Bjerregaard Department Of pathology Herlev University Hospital.

5) N. Efford 2000, "Digital Image Processing A Practical Introduction Using Java", Pearson Edition Limited.

6) A. L. Poul 1997, "Digital Singal Processing With Computer Application", John Wiley & Sons Ltd., 2nd 1997.

7) S. E. Umbagugh 1998, "Computer Vision And Image Processing", Prentice-Hall, Inc., USA.

8) B. Konar A. & K-Mukherjee 2001, "Image Matching With Fuzzy Moment descriptors", Department Of Electronics And Tel-Communication Engineering Jadavpur University.
   WWW.elsevier.com

9) C. B. Perwass, G. Sommer 2001, "A Fuzzy Logic Algorithm for Dense Image Pixel matching", Proceeding Of VI 2001, Ottawa Canada.

10) Ali. A. Alamre 2005, "Image Matching", M.Sc.Thesis, college Of science Basrah University.