International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 9 Issue 8, August-2020, Impact Factor: 6.754

Using Traditional and Fuzzy Clustering to Image Retrieval

Shahbaa I. Khaleel

Software Dept., College of Computer Science and Mathematics, Mosul University, Iraq

ABSTRACT

In this research, a content based image retrieval system was designed and implemented that assesses the similarity of each image in its data store with the image of the query in terms of color and its properties, and restores images within the required similarity range. Here, color content extraction, known and powerful techniques, color histogram CHM, color moments CMM and color corrlogram CCM methods are all used to get effective results. This research introduces clustering based retrieval, a new scheme for image retrieval, based on a fairly simple assumption: similar semantic images tend to cluster in some feature area. Retrieval based clustering for an image attempts to recover coherent and similar image sets. Here the query image comparison is not done with all the images in the database, but only with a small subset. In this research, the traditional clustering process represented by the algorithm of the kmeans KM was used to retrieve content based color images and was good at work, but to obtain better results, the fuzzy clustering method (fuzzy c-means FCM) was used to perform the retrieval process of color images that extracting their characteristics or features using three methods to extract the characteristics of color images represented by CHM, CMM and CCM. Then using the traditional and fuzzy clustering method to retrieve images and the last method was the best of all methods.

Keywords: Image retrieval, feature extraction, color histogram, color moments, color correlogram, clustering algorithm , fuzzy logic.

1. INTRODUCTION

A typical content-based image retrieval system deals not only with various sources of information in different formats but also with user's requirements. Basically it analyzes both the contents of the source of information and the user queries, and then matches these to retrieved those items that are relevant. The major functions of such a system are to[1,2]: first, analyze the contents of the source information, and represent the contents of the analyzed sources in a way that will be suitable for matching user queries (space of source information is transformed in to feature space for the sake of fast matching in a later step). This step is normally very time consuming since it has to process sequentially all the source information (images) in the database. However, it has to be done only once and can be done off-line[3,4]. Second, analyze user queries and represent them in a form that will be suitable for matching with the source database. Part of this step is similar to the previous step; however, it is applied only to the query image[[5,6]. Third, define a strategy to match the search queries with the information in the stored database, and then retrieve the information that is relevant in an efficient way. This step is done online and is required to be very fast. Modern indexing techniques can be used to reorganize the feature space to speed up the matching processing[7,8].

It is evident from the above discussion that on the one side of a content based image retrieval system, there are sources of visual information in different formats, and on the other there are the user queries. These two sides are linked through a series of tasks as illustrated in figure(1).

In this research, the traditional clustering process represented by the algorithm of the kmeans was used to retrieve content based color images and was good at work, but to obtain better results, the fuzzy clustering method was used to perform the retrieval process of color images that extracting their characteristics or features using three methods to extract the characteristics of color images represented by color histogram, color moment and color correlogram. Then using the traditional and fuzzy clustering method to retrieve images and the last method was the best of all methods.

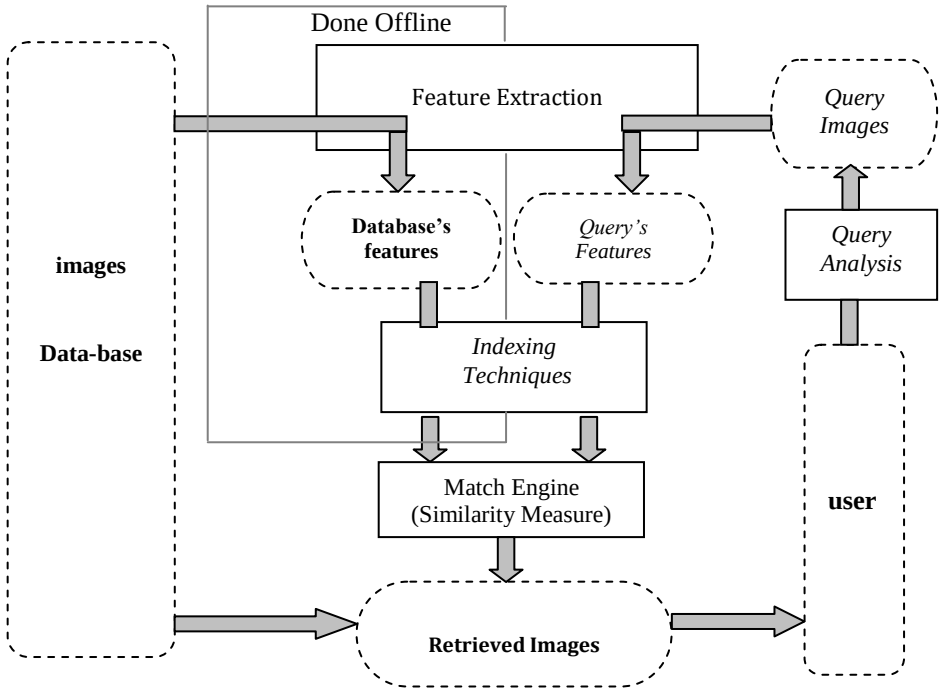

Figure (1) Broad outline of a contentbased image retrieval

2. COLOR FEATURES

Color is a simple and straightforward feature for all kinds of color images. The human eye is much more sensitive to color shades than gray-level intensities in an image. The colors of different objects are also largely resolution and view invariant. Selecting an appropriate color space and the used color quantization are key issues for color feature extraction. Color quantization is used to reduce the number of distinct colors in an image. It is used to reduce both computational complexity of color feature extraction and the dimensionality of the resulting feature vectors[9,10].Color has been the most commonly-used feature type in content based image retrieval. Basic color features are easy to implement and usually yield reasonable and predictable results which can then be improved by including other types of features. The standard representation for color information in content based image retrieval has been the color histogram. The color histogram describes the distribution of different colors in an image in a simple and computationally efficient manner. Other commonly used color features include color moments and the color correlogram [3,10]].

2.1 COLOR HISTOGRAM

One of the widely used methods for querying and retrieval by color content is color histograms. The color histograms are used to represent the color distribution in an image. Mainly, the color histogram approach counts the number of occurrences of each unique color on a sample image. Since an image is composed of pixels and each pixel has a color, the color histogram of an image can be computed easily by visiting every pixel once. By examining the color histogram of an image, the colors existing on the image can be identified with their corresponding areas as the number of pixels [5,10]. One possible way of storing the color information is to use three different color histograms for color channel. This is called Histogram of Intensity, a histogram with a large number of bins will not only increase the computational cost, but will also be inappropriate for building efficient indexes for image databases. Another possible method is to have a single color histogram for all of the color channels. In the latter approach, the color histogram is simply a compact combination of three histograms and the empty slots can be discarded easily. The histogram approach is commonly used in most of the existing systems supporting query-by-color content[8].The retrieval of similar images is based on the similarity between their respective color histogram. A common similarity metric is based on the Euclidean distance between the abstracted feature vectors that represent two images, and it is defined as:

$$d(Q,I) = \sqrt{\sum_{j=1}^{n} \left(h_j^Q - h_j^I\right)^2}$$

Where Q and I represent the query image and one of the images in the image set, and h_j^Q and h_j^I represent the values of the feature vectors of these images respectively, i.e. color histogram feature. While the distance between intensity histogram features is calculated as follows:

$$d(Q,I) = \sqrt{\sum_R \sum_G \sum_B \left(h_{rgb}^Q - h_{rgb}^I\right)^2}$$

Where, h_{rgb}^Q , h_{rgb}^I represent the histogram of query image and one of the images in the image set. Note that a smaller distance reflects a closer similarity match [7,10].Even though color histogram based techniques have been quite successful in given settings, they have some notable shortcomings. Firstly, the color histogram is too finely quantized in color space, and hence, does not take into consideration the fact that the human visual system can only perceive a few colors at a time. The large number of feature vectors is often difficult to index and raises the retrieval cost. Another shortcoming of color histogram is that it does not incorporate any spatial information. Several color descriptors have been proposed with the effort to include spatial information, including the compact color moments and the color correlogram and have also been used to reduce the dimensionality of the feature vectors [9].

2.2 COLOR MOMENTS

Color moments have been successfully used in many retrieval systems, especially when the image contains just the object. The first order (mean), the second (variance or standard deviation) and the third order (skewness) color moments have been proved to be efficient and effective in representing color distributions of images. Mathematically, the first three moments are defined as:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij} \quad \sigma_i = \left[\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij} - \mu_i\right)^2\right]^{1/2} \qquad S_i = \left[\frac{1}{N}\sum_{j=1}^{N}\left(f_{ij} - \mu_i\right)^3\right]^{1/3}$$

Where f_{ij} is the value of the i^{th} color component of the image pixel j , N is the number of pixels in the image , and μ , σ , S represent mean, standard deviation , and skewness respectively[11].Since only 9 (three moments for each of the three color components) numbers are used to represent the color content of each image, color moments are a very compact representation compared to other color features. Usually, color moments can be used as the first pass to narrow down the search space before other sophisticated color features are used for retrieval. The distance measure d_{CM} between query image Q and target image T is calculated by: [12]

$$d_{CM}(Q,T) = \sqrt{\sum_{i\in(R,G,B)} \left[\mu_i(Q) - \mu_i(T)\right]^2 + \left[\sigma_i(Q) - \sigma_i(T)\right]^2 + \left[S_i(Q) - S_i(T)\right]^2}$$

2.3 COLOR CORRELOGRAM

A color correlogram expresses how the spatial correlation of color changes with distance. A color histogram captures only the color distribution in an image and does not include any spatial information. Thus, the correlogram is one kind of spatial extension of the histogram[10,13].Color correlogram is very similar to Co-occurrence matrix developed some 20 years ago for gray scale texture classification. The Correlogram represents the distribution of the Co-occurrence matrix, but it is calculated using several different distances (d) instead of using a single distance; and it uses eight directions (angles) as shown in figure (2).

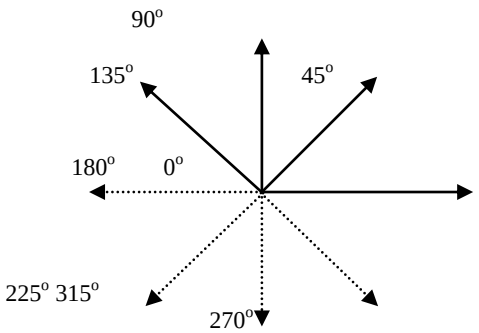

Figure (2): The Eight Directions for Correlogram

Let [D] denote a set of D fixed distances $\{d_1,...,d_D\}$, and f be an image consisting of M different colors $c_1, c_2,..., c_M$. For a pixel $p(x,y) \in f$, let $f(p)$ expresses the pixel color and $f_c = \{p | f_{(p)} = c\}$. Therefore, $p \in f_c$ is equivalent to $p \in f$, $f_{(p)} = c$. For simplicity, $L_\infty - Norm$ is used to measure the distance between pixels. This measure is computed for two pixels $p_1(x_1, y_1)$, $p_2(x_2, y_2)$ as follows: $|p_1 - p_2| = \max\{|x_1 - x_2|, |y_1 - y_2|\}$

Assume that k is a specified distance and $i, j \in \{1,...,M\}$. The correlogram of f is defined by:

$$\gamma(i,j,k) = \underset{p_1 p_2 \in f}{pr}\{p_2 \in f_{cj} | p_1 \in f_{ci}, |p_1 - p_2| = k\}$$

$\gamma(i,j,k)$ denotes the probability of finding pixels with color c_j at the distance k of the pixel with color c_i [19].In this research we calculate five features from the correlogram that is :

$$Energy = \sum_{ci}\sum_{cj}\sum_{k}\left(\gamma^k_{cicj}\right)^2 \qquad Entropy = \sum_{ci}\sum_{cj}\sum_{k}\gamma^k_{cicj}\log\left(\gamma^k_{cicj}\right) \quad Contrast = \sum_{ci}\sum_{cj}\sum_{k}(ci-cj)^2\gamma^k_{cicj}$$

$$Inverse-Difference-Moment = \sum_{ci}\sum_{cj}\sum_{k}\frac{\gamma^k_{cicj}}{|ci-cj|^2}, ci \neq cj$$

$$Correlation = \frac{\sum_{ci}\sum_{cj}(cicj)\gamma_{cicj} - \mu_x\mu_y}{\sigma_x\sigma_y}$$

Where, means and standard deviations are defined as:

$$\mu_x = \sum_{ci} ci \sum_{cj}\gamma_{cicj} \qquad \mu_y = \sum_{cj} cj \sum_{ci}\gamma_{cicj}$$

$$\sigma_x = \sum_{ci}(ci-\mu_x)^2\sum_{cj}\gamma_{cicj} \qquad \sigma_y = \sum_{cj}(ci-\mu_y)^2\sum_{ci}\gamma_{cicj}$$

The computation order of the correlogram is $O(M^2 d)$ where d represents the number of different distances. The computation time required in image retrieval is proportional to the length of the feature vector used in image database indexing. Decreasing the number of the levels (M) in the images is done by quantizing the images before computing of the correlograms. Using this method, we achieve two important benefits: 1) Decreasing the computational cost and 2) Quantization of the image into sets of similar levels generalizes the image content, which improves the retrieval results[2].The main advantages of using color correlogram are: ***i)*** taking into consideration the spatial correlation of colors, ***ii)*** describing the global distribution of local spatial correlation of colors, ***iii)*** being easy to implement, and ***iv)*** producing fairly small size index.

3. CLUSTERING

Clustering is an unsupervised learning problem, which tries to group a set of points into clusters such that points in the same cluster are more similar to each other than points in different clusters, under a particular similarity metric [14].The goal of clustering is to identifying the clusters, which can be considered as classes[15]. Clustering can be used to produce an effective image index as follows: After clustering, each cluster is represented by its centroid or sometimes a single representative data item and, instead of the original data items, the query point is compared to the centroids or the cluster representatives. The best cluster or clusters, according to the used similarity measure, are then selected and the data items belonging to those clusters are retrieved also according to the used similarity measure [16,17,18].

3.1 K-MEANS CLUSTERING METHOD

The k-mean algorithm is the most frequently used clustering algorithm due to its simplicity and efficiency. K-means is a partitional clustering algorithm. It performs iterative relocation to partition a dataset into k cluster [19]; and it is based on the minimization of a performance index which is defined as the sum of the squared distances from all points in a cluster domain to the cluster center. This algorithm consists of the following steps: [20,21]

Step1: Choosing K initial cluster centers $z_1(1), z_2(1),..., z_K(1)$. These are arbitrary and are usually selected as the first K samples of the given sample set.

Step2: Distributing the samples $\{x\}$ at the k^{th} iterative step among the K cluster domains, using the relation:

$$x \in S_j(k) \quad if \quad \|x - z_j(k)\| < \|x - z_i(k)\|$$

for all $i = 1,2,...,K, i \neq j,$ where $S_j(k)$ denotes the set of samples whose cluster is $z_j(k)$.

Setp3: Computing the new cluster centers $z_j(k+1), j = 1,2,...,K$, such that the sum of the squared distances from all points in $S_j(k)$ to the new cluster center is minimized. In other words, the new cluster center $z_j(k+1)$ is computed so that the performance index:

$$J_j = \sum_{x \in S_j(k)} \|x - z_j(k+1)\|^2 \qquad , j = 1,2,...,K$$

is minimized. The $z_j(k+1)$ which minimizes this performance index is simply the sample mean of $S_j(k)$. Therefore, the new cluster center is given by: $z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x \qquad , i = 1,2,...,K$

where N_j is the number of samples in $S_j(k)$. The name "K-means" is obviously derived from the manner in which clusters are sequentially updated.

Sept4: If $z_j(k+1) = z_j(k)$ for $j = 1,2,...,K$, the algorithm has converged and the procedure is terminated. Otherwise one should go to step2.

The behavior of the k-means algorithm is influenced by the number of cluster centers specified, the choice of initial cluster centers, the order in which the samples are taken, and, of course, the geometrical properties of the data. Although no general proof of convergence exists for this algorithm, it can be expected to yield acceptable results when the data exhibit characteristic pockets which are relatively far from each other. In most practical cases the application of this algorithm will require experimenting with various values of K as well as different choices of starting configurations [22].

4. FUZZY CLUSTERING

Fuzzy algorithms can assign data to multiple clusters. Fuzzy algorithms are based on fuzzy logic [23] or fuzzy clustering is an extension of the concept of fuzzy logic which was developed by LotfiZadeh. In fuzzy logic, Zadeh suggested that a statement could be partially true and partially false [24]. Deterministic membership functions assign each datum to a particular cluster, that is, the membership functions map the membership of the datum in a cluster to either 0 or 1. Fuzzy membership functions, on the other hand, map the memberships to the real interval [0.0 1.0]. The degree of membership in the cluster depends on the closeness of the datum to the cluster center. High membership values indicate less distance between the datum and the cluster center[25].Clustering involves the process of arranging data points in such a way that items sharing similar characteristics are grouped together. The goal is to find the natural grouping of data points without prior knowledge of class labels (unsupervised). Fuzzy C Means (FCM) is a feature clustering technique wherein each feature point belongs to a cluster by some degree that is specified by a membership grade.

4.1 FUZZY C-MEANS METHOD

FCM algorithm, is also known as fuzzy ISODATA[21], which was introduced by Bezdek as extension to Dunn's algorithm to generate fuzzy sets for every observed feature[26]. The fuzzy c-means clustering algorithm is based on the minimization of an objective function called c-means functional.

Fuzzy c-means algorithm is one of the well known relational clustering algorithms. It partitions the sample data for each explanatory (input) variable into a number of clusters. These clusters have "fuzzy" boundaries, in the sense that each data value belongs to each cluster to some degree of Membership. Having decided upon the number of such clusters to be used, some procedure is needed to locate their centers (or more generally, mid-point) and to determine the associated membership for the data points[27,28].Fuzzy clustering methods allow for uncertainty in the cluster assignment. FCM is an iterative algorithm to find cluster centers (centroids) that minimize a dissimilarity function[24,29]. Rather than partitioning the data into collection of distinct sets by fuzzy partitioning, the membership matrix (U) is randomly initialized.

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall_j = 1, 2, ..., n$$

The dissimilarity function (or more generally the objective function), which is used in FCM.

$$j(U, c_1, c_2, ..., c_c) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^{m} \, d_{ij}^{2}$$

Where, μ_{ij} is between 0 and 1; c_i is the centroids of cluster i;

$d_{ij} = \|x_j - v_i\|$ represents Euclidean Distance between i[th]. Centroids c_i and j[th]. Data point.

$m \in [1, \infty]$ is a weighting exponent which is a real number greater than 1. There is no prescribed manner for choosing the exponent parameter, "m". When m is close to 1, then the cluster closest to the point is given much larger weight, then the clusters and the algorithm is similar to K-means.

μ_{ij} is named partition matrix, and satisfies the following three conditions[26,30]:

1) $0 \le \mu_{ij} \le 1 \quad \forall i, j,$ 2) $0 < \sum_{j=1}^{n} \mu_{ij} < n \quad \forall \mathrm{i},$ 3) $\sum_{i=1}^{c} \mu_{ij} = 1 \quad \forall \mathrm{j}$

A conventional FCM algorithm includes the following steps:

1. Initialize the cluster center $V = \{v_1, v_2, ...v_c\}$, or initialize the membership matrix $U = \mu_{ij}$ with random value and make sure it satisfies conditions (1) and (2) and then calculates the centers.

$$U^{(0)} = \begin{bmatrix} \mu_{11}^{(0)} \; \mu_{12}^{(0)} \; ... \; \mu_{1n}^{(0)} \\ \mu_{21}^{(0)} \; \mu_{22}^{(0)} \; ... \; \mu_{2n}^{(0)} \\ ... \quad ... \quad ... \\ \mu_{c1}^{(0)} \; \mu_{c2}^{(0)} \; ... \mu_{cn}^{(0)} \end{bmatrix} = \begin{bmatrix} \mu_1^{(0)}(x_1) \; \mu_1^{(0)}(x_2) ... \mu_1^{(0)}(x_n) \\ \mu_2^{(0)}(x_1) \; \mu_2^{(0)}(x_2) ... \mu_2^{(0)}(x_n) \\ ... \quad ... \quad ... \\ \mu_c^{(0)}(x_1) \; \mu_c^{(0)}(x_2) ... \mu_c^{(0)}(x_n) \end{bmatrix}$$

2. compute the fuzzy centers v_j by using:

$$v_j = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mu_{ij})^m \, x_i}{\sum_{i=1}^{n} (\mu_{ij})^m} \quad \forall j = 1, ..., c$$

3. updated the fuzzy membership μ_{ij} by using :

$$\mu_{ij} = \left[\sum_{\mathrm{k}=1}^{\mathrm{c}} \left(\frac{\mathrm{d_{ij}}}{\mathrm{d_{ik}}} \right)^{\frac{2}{m-1}} \right]^{-1} \qquad \ldots\ldots$$

Where, $d_{ij} = \| x_i - \mathrm{v_j} \|, \; \forall i = 1, ..., n, \forall j = 1, ...c$

4. Repeat steps 2) and 3) until the minimum J value is achieved or the centroids from the previous iteration are identical to those generated in the current iteration .
5. Finally, defuzzification is necessary to assign each data point to a specific cluster (i.e. by setting a data point to a cluster for which the degree of the membership is maximal)[24,31].

By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the "right" location within a data set. FCM does not ensure that it converges to an optimal solution, because the cluster centers are randomly initialized. Though, the performance depends on initial centroids, there are two ways as described below for a robust approach in this regard.

1) Using an algorithm to determine all of the centroids.
2) Run FCM several times each starting with different initial centroids.

The process stops when the centroids from the previous iteration are identical to those generated in the current iteration.

5. PROPOSED SYSTEM

Here, we present a technique for image retrieval based on color from large databases. The goal is to group similar images into clusters and to compute the cluster centers, so that during retrieval, the query image need not be compared exhaustively with all the images in the database. To retrieve similar images for a given query, the query image is initially compared with all the cluster centers. Then a subset of clusters that have the largest similarity to the query image is chosen and all the images in these clusters are compared with the query image.The Image Retrieval Based on

Clustering IRBC system has a three-step approach to retrieve images from the databases. The first step is *indexing*: for each image in a database, feature vector is computed and stored in feature space. The second step is *clustering*: the images in the database are grouped into clusters of images with similar color content using clustering algorithm (traditional and fuzzy algorithms). The third step is *searching*: given a query by a user, its feature vector is computed and the system retrieves images having feature vectors with a small subset of clusters that best match the query feature vector. Figure (3) shows the steps of the system that retrieved color images based on clustering.

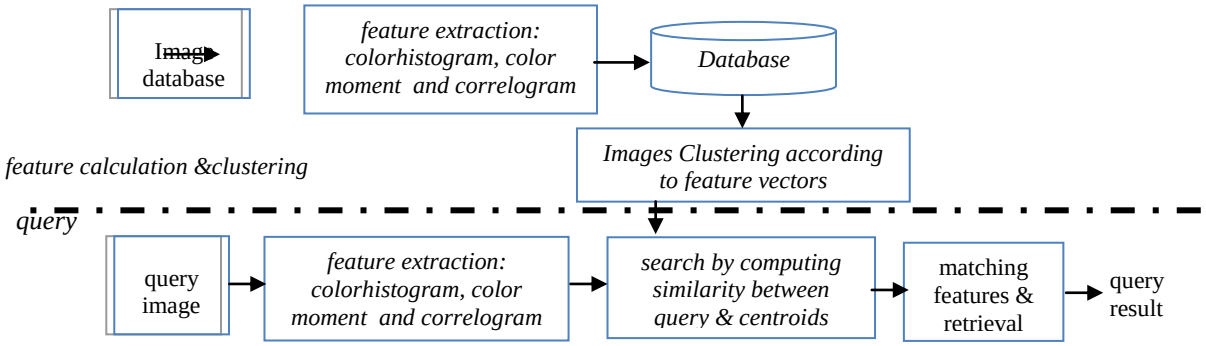

Figure (3) The steps of image retrieval based on clustering

In this research used color image and applied three types of feature extraction colorhistogram, colormoments, color correlogram, and then employed the clustering method k-means clustering algorithm(KM), fuzzy c-means clustering algorithm(FCM)to decrease the comparison of query.

6. EXPERIMENTAL AND RESULTS

Color is a simple and straightforward feature for all kinds of color images. Color features are easy to implement and usually yield reasonable and predictable results which can then be improved by including other types of features. The standard representation for color has been the color histogram. Other color features include color moments and color correlogram. These three features extraction are applied on color image in this research, Moreover, the traditional clustering algorithm (KM) and fuzzy clustering algorithm(fuzzy c-means FCM) are also applied to all these methods.

6.1-COLOR HISTOGRAM METHOD

A single color histogram for all of the color channels are used in this research, and the two clustering algorithms are also applied. The FCM is the best algorithm. This can be noticed in table (1), the abbreviation CHM is mean color Histogram method.

Table (1) The ResultsKM & FCM clustering in Color Histogram Method

Images	Relevant retrieved	Relevant number	Retrieved number	Recall	Precision	Method
Baloon1	17	25	32	0.680000	0.531250	CHM
	19	25	19	0.760000	1.000000	CHM-KM
	23	25	23	0.920000	1.000000	CHM-FCM
Rabit1	8	11	16	0.727273	0.500000	CHM
	8	11	8	0.727273	1.000000	CHM-KM
	11	11	11	1.000000	1.000000	CHM-FCM
Rabit2	5	11	13	0.454545	0.384615	CHM
	7	11	7	0.636364	1.000000	CHM-KM
	10	11	10	0.909090	1.000000	CHM-FCM
Tiger	24	25	39	0.960000	0.615385	CHM
	18	25	21	0.720000	0.857143	CHM-KM
	22	25	22	0.880000	1.000000	CHM-FCM

6.2- COLOR MOMENT METHOD

Color moments are used to represent the color content of each image. Three moments for each of the three color components are extracted; these are Mean, Standard deviation and Skewness. FCM is better than traditional clustering algorithm when applied in this method as illustrated in table (2), CMM means Color Moment method.

Table (2) The Results of KM&FCM in Color Moment Method

Images	Relevant retrieved	Relevant number	Retrieved number	Recall	Precision	Method
Baloon2	6	25	22	0.240000	0.272727	CMM
	1	25	15	0.040000	0.066667	CMM-KM
	22	25	23	0.880000	0.956522	CMM-FCM
Lion1	21	21	63	1.000000	0.333333	CMM
	14	21	14	0.666667	1.000000	CMM-KM
	21	21	21	1.000000	1.000000	CMM-FCM
Monkey1	14	15	78	0.933333	0.179487	CMM
	6	15	18	0.400000	0.333333	CMM-KM
	12	15	12	0.800000	1.000000	CMM-FCM
Livestock	14	14	59	1.000000	0.237288	CMM
	11	14	16	0.785714	0.687500	CMM-KM
	13	14	13	0.928571	1.000000	CMM-FCM

6.3-COLOR CORRELOGRAM METHOD

We calculate five features from the correlogram: Energy, Entropy, Contrast, Inverse-difference-moment, and Correlation. This method also prove that FCM is better than other methods, see table (3), the abbreviation CCM mean Color Correlogram.

Table (3) The Results of KM & FCM in Color Correlogram Method

Images	Relevant retrieved	Relevant number	Retrieved number	Recall	Precision	Method
Lion2	21	21	85	1.000000	0.247059	CCM
	10	21	11	0.476190	0.909091	CCM-KM
	21	21	21	1.000000	1.000000	CCM-FCM
Monkey2	12	15	27	0.800000	0.444444	CCM
	3	15	7	0.200000	0.428571	CCM-KM
	12	15	12	0.800000	1.000000	CCM-FCM
Rabit3	10	11	91	0.909091	0.109890	CCM
	3	11	15	0.272727	0.200000	CCM-KM
	11	11	11	1.000000	1.000000	CCM-FCM

And the images that were used in this research, i.e. the database on which all the research methods were implemented, included a group of different animal color pictures and the following images representing the query images.

7. CONCLUSIONS

In this research we designed and implemented a content based image retrieval system that evaluates the similarity of each image in its data store to a query image in terms of color and textural characteristics, and returns the images within a desired range of similarity. Here, the color content extraction, a well-known and powerful techniques, color

histogram, color moment and color correlogram are used all these features to produce efficient results.This research introduces a cluster based retrieval of image, a novel image retrieval scheme, based on a rather simple assumption: semantically similar images tend to be clustered in some feature space. Clusterbased retrieval of an image attempts to retrieve semantically coherent image clusters from unsupervised learning of how images of the same semantics are alike. At search time, the query image is not compared with all the images in the database, but only with a small subset. Here we used k-means clustering algorithm and fuzzy c-means algorithm this algorithm is an efficient algorithm and produce efficient retrieval.We conclude that color methods are useful in content based image retrieval. and it further improves the performance if the fuzzy c-mean clustering algorithm are used. These methods proved an efficient performance when used in contentbased image retrieval.

REFERENCES

[1] Tran L.V.,(2003), *Efficient Image Retrieval With Statistical Color Descriptors*, PhD Thesis, Department of Science and Technology, Linköping University, Sweden.
[2] Al-Taey, S. I. 2006. Efficient content-based multimedia retrieval with neural network, Ph.D. thesis, Computer science-Faculty of Computers & Mathematics Sciences, University of Mosul, Iraq.
[3] KosKela M., (2003), Interactive Image Retrieval Using Self-Organizing Maps, PhD Thesis, Department of Computer Science and Engineering, University of Technology.
[4] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Image Retrieval Based on Content by using Clustering Algorithms , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 7, July-2019.DOI:http://www.doi.org/ste/19.0703
[5] Konak E., (2002), A content-Based Image Retrieval System For Texture And Color Queries, MSc thesis, Department of Computer Engineering, University of Bilkent.
[6] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Content Based Video Retrieval Using Back propagation Neural Network and Radial Basis Function Network , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 12, December - 2019.DOI:http://www.doi.org/ste/19.1202.
[7] Chikara V., (2001), Color-Based Image Retrieval Using Compact Binary Signatures, Technical Report TR 01-08, Department of Computing Science, University of Alberta, Canada.
[8] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Colored Image Retrieval Based on Content by Using Integration of Features for Many Color Space Models , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 8, August-2019.DOI:http://www.doi.org/ste/19.0803
[9] Chen J., (2003), Perceptually-Based Texture and Color Feature For Image Segmentation And Retrieval, PhD Thesis, Department of Electrical and Computer Engineering, University of Northwestern.
[10] Dr. Shahbaa I. Khaleel, Dr. Nidhal Hussein, Color Image Retrieval Based on Content by Using Linear and Nonlinear Combiner, International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 8, August-2019.DOI:http://www.doi.org/ste/19.0802.
[11] Long F., Zhang H., Feng D., (2002) Fundamentals of Content-Based Image Retrieval, Ch 1.
[12] Manjunath B., Ohm J., Yamada A., (2001), Color and Texture Descriptors, IEEE Transactions on Circuits and Systems for Video Terchnology, Vol 11, pp. 703-715.
[13] Hung J., Kumar S., Zabih R., (1999), Image Indexing Using Color Correlograms, Inernational Journal of Computer Vision, pp. 245-268.
[14] Dr. Shahbaa Ibrahim Khaleel, Color Image Compression Based on Fuzzy Clustering , International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 8 Issue 7, July-2019.DOI:http://www.doi.org/ste/19.0702.
[15] Manar Y. Kashmola, Bayda I. Khaleel, 2012, Clustering and Detecting Network Intrusion Based on Fuzzy Algorithms, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 9 , Issue (2), Pages 125-138 , Publisher: Mosul University, DOI: 10.33899/csmj.2012.163706.
[16] Krishnamachari S., Abdel-Mottaleb M., (1999), Hierarchical Clustering Algorithm For Fast Image Retrieval, Conference on Storage and Retrieval For Image and Video Databases, pp. 427-435.
[17] BaydaaIbraheemKhaleel, 2014,Image Clustering based on Artificial Intelligence Techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (1), Pages 99-112 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163735.
[18] Shahbaa I. Khaleel, Karammohammedmahdisaleh, 2012, Detection of network anomaly based on hybrid intelligence techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 9 , Issue (2), Pages 81-98 , Publisher: Mosul University, DOI: 10.33899/ csmj.2012.163720.
[19] Basu S., (2003), Semi-Supervised Clustering: Learning With Limited User Feedback, PhD Thesis, University of Texas, Austin.
[20] Shahbaa I. Khaleel, Baydaa I. Khaleel, Alaa I. khaleel, Image Compression Based on Artificial Intelligent Techniques, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 6 , Issue (3), Pages 75-109 , Publisher: Mosul University, DOI: 10.33899/csmj.2009.163839.
[21] Nada N. Seleem, Baydaa I. Khaleel, Shahbaa I. Khaleel, Artificial Intelligent Techniques with Watermarking, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 6 , Issue (2), Pages 229-266 , Publisher: Mosul University, DOI: 10.33899/csmj.2009.163810.
[22] Tou J., Gonzalez R., (1974), Pattern Recognition Principles, Addision-Wesley Publishing Company, USA.
[23] Dr. ManarYounis K., Dr. BaydaIbraheemKhaleel, 2013, Hybrid Fuzzy and Neural Network for Intrusion

Detection System, *JOURNAL OF EDUCATION AND SCIENCE*, Volume 26, Issue 4, Pages 152-170, Publisher: Mosul University, DOI: 10.33899/edusj.2013.89971.

[24] AL-Taey, baydaaIbraheemKhaleel, 2012, Artificial Intelligent Techniques for Intrusion Detection and Classification, Ph.D. Thesis, Computer Science, College of Computer Sciences and Mathematics, University of Mosul.

[25] Manar Y. Ahmed, Bayda I. Khaleel, 2013, Using Artificial Intelligence Techniques For Intrusion Detection System, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 10 , Issue (1), Pages 157- 175, Publisher: Mosul University, DOI: 10.33899/csmj.2013.163433. Fifth Scientific Conference Information Technology 2012 Dec. 19-20.

[26] Baydaa I. Khaleel, 2014, Using Artificial Intelligence Techniques for Image Compression, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (2), Pages 65-81 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163750.

[27] Shailendra S. , PremNarayan A. , (2012), Comparison of K-means and Modified K-mean algorithms for Large Data-set , International Journal of Computing, Communications and Networking, Volume 1, No.3, pp. 106-110 .

[28] Shahbaa I. Khaleel and Ragadwaleedkhaled, 2014, Selection and Prioritization of Test Cases by using Bees Colony, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 11 , Issue (1), Pages 179-201 , Publisher: Mosul University, DOI: 10.33899/csmj.2014.163746.

[29] Shahbaa I. Khaleel, Ashraf Abdulmunim Al Thanoon, 2013, Design a Tool for Generating Test Cases using Swarm Intelligence, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 10 , Issue (1), Pages 421-444 , Publisher: Mosul University, DOI: 10.33899/csmj.2013.163468.

[30] Fardos Adnan Abdalkader, Shahbaa I. Khaleel, Nada N. Saleem, 2010, Hybrid intelligent watermark System, AL-Rafidain Journal of Computer Sciences and Mathematics Volume 7 , Issue (2), Pages 125-138 , Publisher: Mosul University, DOI: 10.33899/csmj.2013.163902.

[31] Huang J., (1998), Color-Spatial Image Indexing And Applications, PhD Thesis, Cornell University.