



University of AL_Qadisiya

College computer science and information tecgnology

computer science department

Content based Image Retrieval (CBIR)

**Project submitted to the college of Computer Science and
information technology as part of the requirements to get**

B.SC

Submitted

Zainab Musa Abd

Haneen Mohmmmed Abd-hadi

Haneen Rasem Oada

Supervision by

A.L Mohammed Hamzah Abed



Dedicated

Everyone who contributed to our arrival through to the end
To all who taught me something new And fueled an intellectual
science and knowledge To each of the stop beside us and
helped us In all the hardships To our professors



Thanks

We must and we move the last steps in the university life to
go back to the years we spent in the university campus with the
esteemed our professors who have given us so much great
efforts in building tomorrow's generation to send the nation
again.offer our deepest gratitude and appreciation and love to
those who carried the message holiest in life.To those who
paved the way for us science and knowledge...

thanks to All professors Distinguished ...

index Contents

1.1 Introduction	1
1.2 Aim of the project	2
1.3 Visual features used in retrieval	2
1.4 image database	3
2.1 Introduction	4
2.2 Techniques types	5
2.3 image retrieval techniques	5
2.3.1 Texture	5
2.3.1.1 Wavelet Texture Features	6
2.3.1.2 Haar wavelet transform	7
A. Steps of Haar Wavelet Transform	8
B. The advantages of Haar Wavelet transform	10
2.3.1.3 Extraction of feature	10
2.3.1.4 Algorithm	12
2.3.2 Shape	12
2.3.3 color	13
2.2.3.1 color histogram	13

2.2.3.2 Histogram of Prime Value	13
2.2.3.3 Color Image Moment	14
2.2.4 Proposed work	14
2.2.5 Similarity measurement	14
2.2.5.1 Euclidian Distance	15

Index of figure

Fig(2,1) After applying Haar Wavelet Transformation	11
Figure(2.2) image retrieval proposal work	16

Index of equation

equation(2,1)the values of first average sub.....	7
equation(2,2) the approximation coefficients.....	8
equation(2,3) detail coefficient.....	8

equation(2,4) The work of low pass filter.....	10
equation(2,5) highpass filter.....	10
equation(2,6) The Wavelet signature	12
equation(2.7) the mean	14
equation (2.8) standardeviation.....	14
equation(2,9) Euclidian Distance.....	17
equation(2,10) Cosine.....	17

Chapter one

1.1 Introduction

content-based image retrieval has gained tremendous momentum during the last decade . A lot of research work has been carried out on Image Retrieval by many researchers, CBIR or Content Based Image Retrieval is the retrieval of images based on visual features such as colour, texture and shape[1] . Reasons for its development are that in many large image databases, traditional methods of image indexing have proven to be insufficient, laborious, and extremely time consuming. These old methods of image indexing, ranging from storing an image in the database and associating it with a keyword or number, to associating it with a categorized description, have become obsolete. This is not CBIR. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image. It involves two steps[2]:

- 1.Feature Extraction: The first step in the process is extracting image features to a distinguishable extent.
2. Matching: The second step involves matching these features to yield a result that is visually similar.

The course textile images plays a vital role in content based image retrieval (CBIR). Thus the task is capably storing, processing and retrieving textile image records. "Content-based" means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image meta data itself. The content-based image retrieval system retrieves the stored images from the database by comparing the features of the query image against the images in the collection. The system first extracts and stores the features of the query image then it go through all images in the database and extract the features of each image. The results are the images that its features are most similar to the query image. This is led to solve problems text-based image retrieval technology that have two difficulties in the retrieval: First, it has been impossible to note each image; second, the subjectivity and non-precision of image an nota-tion may lead to the adaptation in the retrieval process. In order to overcome these problems, Content-Based Image Retrieval (CBIR) was proposed in the 1990's.

1.2 Aim of the project

When processors become increasingly strong memory becomes increasingly cheaper. And the evolution of the large picture of the diverse applications of rules of business data in the technical and industrial satellites, medicine and other .this has attracted a lot of Customer For example Medicine, Geo, architecture, fashion, design and advertising To obtain a copy of the p database efficiently and effectively is now very necessary . The (CBIR) can used for Various applications[3]

prevention: Automatic face recognition systems, used by police forces.

Security Check: Finger print or retina scanning for access privileges.

Medical Diagnosis: Using CBIR in a medical database of medical images to aid case . diagnosis by identifying similar past

Intellectual Property: Trademark image registration, where a new candidate mark is .compared with existing marks to ensure no risk of confusing property ownership

The aim of this project is to review the current state of the art in content-based image retrieval (CBIR), a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. Our findings are based both on a review of The relevant literature and on discussions with researchers in the field.

1.3 Visual features used in retrieval

Visual features were classified in into primitive features such as color or shape, logical features such as identity of objects shown and abstract features such as significance of scenes depicted[4] . Color has been the most useful and effective feature and almost all systems employ colors. Although most of images that used through internet and social media are in the RGB (Red, Green, Blue) color space, this space is only rarely used for indexing and querying as it does not correspond well to the human color perception. It only seems reasonable to be used for images taken under exactly the same conditions each time such as trademark images. Other spaces such as HSV (Hue, Saturation, and Value) spaces are much better with respect to human perception and are more frequently used. This means that differences in the color space are similar to the differences between colors that humans perceive [4].Texture partly due to the imprecise understanding and definition of what exactly visual texture actually is, texture measures have an even larger variety than color measures. Some of the most common measures for capturing the

texture of images are wavelet and Gabor filters. These texture measures try to capture the characteristics of the image or image parts with respect to changes in certain directions and the scale of the changes. This is most useful for regions or images with homogeneous texture [4-5]. Shape is an important visual feature and it is one of the basic features used to describe image content. However, shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost[6] .

1.4 image database

We have used a database in this research, which that is a set of 17 types of roses vary in color, shape and texture between them. We got this database from MATLAB help environment , which consists of 1360 pictures, this data base will be used in our search to find the best matching.

Chapter two

2.1 Introduction

With the development in the computer technologies and the advent of the internet, there has been bang in the amount and the difficulty of digital data being produced stored, conveyed, analyzed, and accessed. The lots of this information are multimedia in behavior, comprising digital images, audio, video, graphics, and text information. In order to construct use of this enormous amount of data, proficient and valuable techniques to retrieve multimedia information based on its content need to be developed. In all the features of multimedia, image is the prime factor . Image retrieval techniques are splitted into two categories **text and content** .based categories. **The text** based algorithm comprises some special words like keywords. Keywords and annotations should be dispenses to each image, when the images are stored in a database. The annotation operation is time consuming and tedious[7]. In addition, it is subjective. Further more, the annotations are sometimes incomplete and it is possible that some image features may not be mentioned in annotations . In a CBIR system, images are automatically indexed by their visual contents through extracted low level features, such as shape, texture, color, size and so on[7.8] .However, extracting all visual features of an image is a difficult task and there is a problem namely semantic gap in the semantic gap, presenting high level visual concepts using low level visual concept is very hard. In order to alleviate these limitations, some researchers use both techniques together using different features. This combination improves the performance compared to each technique separately[9.10]. there are two steps for answering a query to retrieve an image. First, some keywords are used to retrieve similar images and after that some special visual features such as color and texture are extracted. In other words, in the second step, CBIR is applied. Color moments for color feature and co-occurrence matrix for extraction of texture features have been computed.

2.2 Techniques types

basically CBIR used two approaches for retrieving the images from the image data base.

1.text-based approach (index images using keywords)

Text based method used the keywords descriptions as a input and get the desired output in the form of similar types of images .**Examples:-** (Google, Lycos, etc.)[11].

2.content.based approach (index images using images)

Content based approach using image as an input query and it generate the output of similar types of images[11]

2.3 image retrieval techniques

There are various techniques have been proposed to retrieve the image effectively and efficiently from the large set of image data in which some of the methods are described below:

2.3.1 Texture

Texture analysis is an important field of image analysis, and plays a role in many tasks, such as medical imaging, machine vision, and content indexing of large image databases. Whilst numerous methods have been proposed through the last few decades, the problem shown to perform to a high level in all situations, over a wide variety of both natural and artificial textures[12]. Much initial focus on texture analysis relied on the assumption that texture could be characterized by the local statistical properties of pixels' grey level values. First order statistics, such as mean, variance and histograms, as well as second level statistics such as co-occurrence matrices have been used with varying levels of success. The conjecture that such statistics sufficed for texture classification was soon rejected, and a variety of new techniques emerged, such as Markov field models, fractals, and connected component analysis[13]. A common weakness of all of these methods is that the texture is analysed on only a single scale, a limitation that has been quickly overcome by the use of multi-resolution analysis methods. This approach is further justified by studies of the human visual system, which have shown that certain cells in the visual cortex respond only to particular spatial frequencies and/or orientations [13.14]

More recently, Gabor filters have been utilised in a variety of ways to solve the problems of texture segmentation and classification . Typically, a bank of such filters are constructed, with each tuned to a specific spatial frequency and orientation. The energy of each filter output is then used to either segment or classify the image.

2.3.1.1 Wavelet Texture Features

When used for the purpose of image analysis, the wavelet transform is typically computed by applying a separable filter bank to the image, given by

$$A_n = (H_x * (H_y * A_{n-1})_{2,1})_{2,1}$$

$$H_n = (G_x * (H_y * A_{n-1})_{2,1})_{2,1}$$

$$V_n = (G_x * (H_y * A_{n-1})_{2,1})_{2,1}$$

$$D = (G_x * (G_y * A_{n-1})_{2,1})_{2,1}$$

where * represents two-dimensional convolution, $\bar{\cdot}$ represents down-sampling by the given factor in each dimension, and H and G are the low and highpass filters, respectively. The approximation image A is obtained by lowpass filtering in both directions, while the detail coefficients H_n , V_n , and D_n are obtained via highpass filtering in one or more directions, thus providing information about a particular scale in either the horizontal, vertical or diagonal direction. Due to the subsampling involved in this computation, the total number of coefficients is equal to the number of pixels in the original image, and images at each subsequent lower resolution level are half the size of the previous level. A simple and powerful feature extracted from the wavelet coefficients is the average energy of each detail image. This is defined as the sum of the squares of each detail image, normalised for the total number of coefficients in the image. Such energy signatures coefficients is the average energy of each detail image. This is defined as the sum of the squares of each detail image, normalised for the total number of coefficients in the image. Such energy signatures provide a good indication of the total energy contained at specific spatial frequency levels and orientations. In addition to energy signatures, more complex features have also been extracted from the wavelet detail coefficients. Such features include histogram parameters of the individual coefficients, as well as second order statistics such as co-occurrence matrix features. These methods have shown to provide further classification accuracy, with error rates as low as 5% reported

2.3.1.2 Haar wavelet transform

Haar Transform is the simplest and basic transformation from the space domain to a local frequency domain and can work as an example for orthonormal wavelet transforms . It is the first known Wavelet and was proposed by Alfred Haar .The Haar Wavelets family for $m \in (0, 1)$ is defined by

$$h_i(m) = \begin{cases} 1 & 0 < m < 0.5 \\ -1 & 0.5 < m < 1 \\ 0 & \text{other wise} \end{cases}$$

For This is also called mother Wavelet. Haar Wavelets are orthogonal, and their forward and inverse transform require only additions and subtractions. So it can be easily implemented on computer. A HT decomposes each signal into two components, one is called average or trend and the other is known as difference or fluctuation. A precise formula for the values of first average sub

Signal. $A^1 = (a_1, a_2, \dots, a_{N/2})$ at signal, length N i.e $f = (f_1, f_2, \dots, f_N)$ is

$$A_n = \frac{f_{2n-1} + f_{2n}}{\sqrt{2}}, n=1,2,3,\dots,N/2$$

and the first detail sub signal,

$$A_n = \frac{f_{2n-1} - f_{2n}}{\sqrt{2}}, n=1,2,3,\dots,N/2 \dots \dots \dots (1.1)$$

To understand how wavelets work, let us start with a simple example. Assume we have a 1D image with a resolution of four pixels, having values

[6 12 8 6]

Haar wavelet basis can be used to represent this image by computing a wavelet transform Consider the pairs of pixels (6, 12) and (8,6) take the average of each pair, 9 and 7, and then record this in the next line. Then record the difference of the averages from the first

value of the pair. This process is then applied to this new string resulting in the line, where the differences are just carried down. As follows:

$$\{6 \ 12 \ 8 \ 6 \}$$

$$\{9 \ 7 \ -3 \ 1 \}$$

The differences recorded to the right-hand side are known as the detail coefficients. Thus, the original image is decomposed into a lower resolution (two-pixel) version and a pair of detail coefficients. The recursive process of averaging and differencing is called a filter bank. The original image can be reconstructed by recursively adding and subtracting the detail coefficients from the lower resolution versions.

A. Steps of Haar Wavelet Transform:

To calculate the Haar transform of an array of n samples:

- 1 .Compute the average of each pair of samples .(n/2 averages)
- 2 .Compute the difference between each average and the samples it was calculated from. (n/2differences)
- 3 .Write the first half of the array with averages
4. Write the second half of the array with differences.
5. Repeat the process on the first half of the array. While doing this the array size should be power of two.

For 2D Haar Transform the procedure remains the same. For example, apply 2D HT to the following finite 2D signal.

$$\begin{bmatrix} 2 & 3 & 3 & 4 \\ 6 & 7 & 5 & 2 \\ 4 & 8 & 7 & 3 \\ 3 & 9 & 2 & 3 \end{bmatrix}$$

Using 1D HT along first row, the approximation coefficients are

$$\frac{1}{\sqrt{2}} (1+2) \text{ and } \frac{1}{\sqrt{2}} (3+4).....(2,2)$$

and the detail coefficient are

$$\frac{1}{\sqrt{2}}(2-3) \text{ and } \frac{1}{\sqrt{2}} (3-4).....(2,3)$$

The same transform is applied to the other rows of I

By arranging the approximation parts of each row transform in the first two columns and the corresponding detail parts in the last two columns we get the following results

$$\frac{1}{\sqrt{2}} \begin{cases} 5 & 7 & : & -1 & -1 \\ 13 & 7 & : & -1 & 3 \\ 12 & 10 & : & -4 & 4 \\ 12 & 5 & : & -6 & -1 \end{cases}$$

In which approximation and detail parts are separated by dots in each row. By applying the following step of 1D HT to the columns resultant matrix, we find that the resultant

$$\frac{1}{\sqrt{2}} = \begin{array}{|c|c|c|c|} \hline 18 & 14 & -2 & 2 \\ \hline 24 & 15 & -10 & 3 \\ \hline -8 & 0 & 0 & -4 \\ \hline 0 & 5 & 2 & 5 \\ \hline \end{array}$$

Thus we have $A = \begin{array}{|c|c|} \hline 18 & 14 \\ \hline 24 & 15 \\ \hline \end{array}$

$$V = \begin{array}{|c|c|} \hline -2 & 2 \\ \hline -10 & 3 \\ \hline \end{array}$$

$$H = \begin{array}{|c|c|} \hline -8 & 0 \\ \hline 0 & 5 \\ \hline \end{array}$$

$$D = \begin{array}{|c|c|} \hline 0 & -4 \\ \hline 2 & 5 \\ \hline \end{array}$$

Each piece shown in example has a dimension (number of rows/2) × (number of columns/2) and is called A, H, V and D respectively. A approximation area) includes

information about the global properties of analysed image. Removal of spectral coefficients from this area leads to the biggest distortion in original image. H (horizontal area) includes information about the vertical lines hidden in image. Removal of spectral coefficients from this area excludes horizontal details from original image. V (vertical area) contains information about the horizontal lines hidden in image. Removal of spectral coefficients from this area eliminates vertical details from original image. D (diagonal area) embraces information about the diagonal details hidden in image. Removal of spectral coefficients from this area leads to minimum distortions in original image. To get the value at next level, again HT is applied as earlier on A. Thus the HT is suitable for application when the image matrix has number of rows and columns as a multiple of 2.

B.The advantages of Haar Wavelet transform:

- 1 .In terms of computation time, it produced Best performance
2. Speed of Computation is very high.
3. Haar Wavelet Transformations deals with Simplicity in working.
4. It is an efficient method for Image compression.

2.1.3.3Extraction of feature

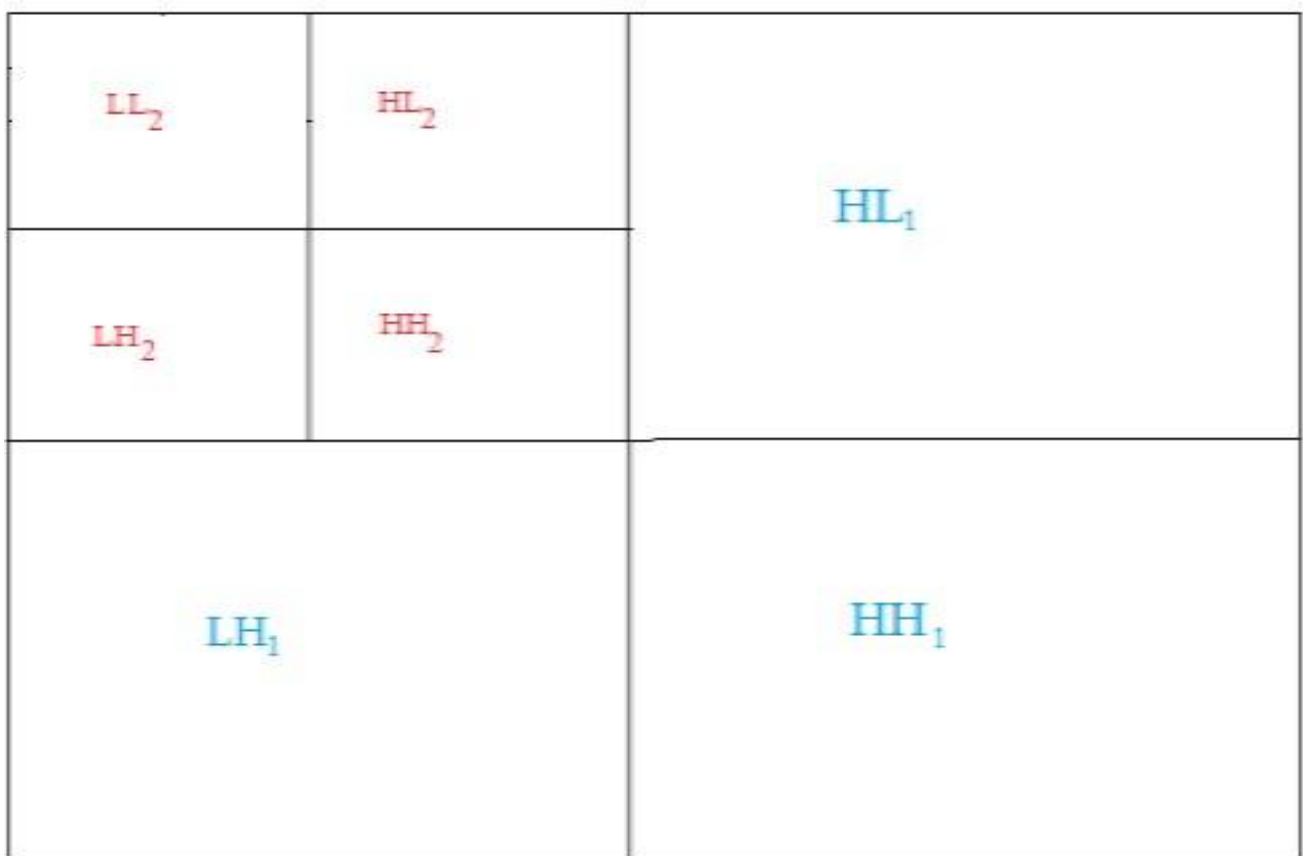
The main purpose of this method is to show the impact of the discrete Haar wavelet transformation on an image as Feature Extraction. This is the combination of a sequence of low-pass and high-pass filters, known as a filter bank. The work of low pass filter performs an averaging/blurring operation, and is written as:

$$H = \frac{1}{\sqrt{2}}(1,1) \dots (2,4)$$

and differencing operation is performed by highpass filter and written as:

$$G = \frac{1}{\sqrt{2}}(1,-1) \dots (2,5)$$

on any adjacent pixel pair. After performing transformation, image composed in four new sub-images. This leads to blurred image and it removes the details, but in remaining three images it represented separately. Using this technique we can easily store and transmit images. we compute feature vectors using Haar wavelets because they are the fastest in computing and have been found to perform well in practice. It helps us to speed up the wavelet computation phase for thousands of sliding windows of different sizes in an image. They also facilitate the development of efficient incremental algorithms for computing wavelet transforms for larger windows in terms of the ones for smaller windows. Production of large number of signatures is the disadvantage of Haar wavelets. We proposed the modified Haar wavelet transformation that reduces signatures only by calculating 10 for the image in our method. In the proposed method, we apply Haar Wavelet Transform to Texture based Images. This can also be extended to other features of the images. In our feature vector computation process, we applied Wavelet Transformations only three times to get 10 sub images of input image in the following way.



After applying Haar Wavelet Transformation
for 3 iterations

2.3.1.4 Algorithm

Algorithm for calculating the signatures of Wavelets

- 1 .Image is for size $r \times r$
- 2 .Divide Image based on Haar Wavelet Transform into BB, CB, CC, BC.
- 3 .Compute signature fr for BB, CB, CC, BC.
- 4 .Divide image into size $r/4 \times r/4$.
5. Compute signatures fr for CB, CC, BC.
- 6 .Divide image into $r/8 \times r/8$.
7. Obtain 10 signatures and stop

The Wavelet signature (texture feature representation) is computed from sub image as follows,

$$F_r = \frac{C_{ij}^2}{I \times J} \quad \dots (2,6)$$

fr is the computed Wavelet signature (texture feature representation) of the sub image. C_{ij} is the representation of the intensity value of all elements of sub image. $i \times j$ is the size of the sub image.

2.3.2 Shape

Shape may be defined as the characteristic surface configuration of an object; an outline or contour. It permits an object to be distinguished from its surroundings by its outline[12]. Shape representations can be generally divided into two categories [1].

1. Boundary-based

2. Region-based

Boundary-based shape representation only uses the outer boundary of the shape. This is done by describing the considered region using its external characteristics; i.e., the pixels along the object boundary. Region-based shape representation uses the entire shape region by describing the considered region using its internal characteristics; i.e., the pixels contained in that region. Methods of Representation For representing shape features mathematically, we have Boundary-based Polygonal Models, boundary partitioning Fourier Descriptors Splines, higher order constructs Curvature Models Region-based: Superquadrics Fourier Descriptors Implicit Polynomials Blum's skeleton The most successful representations for shape categories are Fourier Descriptor and Moment Invariants The main idea of Fourier Descriptor is to use the Fourier transformed boundary as the shape feature. The main idea of Moment invariants is to use region-based moments, which are invariant to transformations as the shape feature.

2.3.3 color

. Content-Based Image Retrieval (CBIR) According to the user-supplied bottom characteristics, directly find out images containing specific content from the image database. The basic process: First of all, do appropriate preprocessing of images, and then extracted image features needed from the image according to the contents of images to keep in the database. When retrieval has been occur to identify the image, extract the corresponding features from a known image and then retrieve the image database to identify the images which are similar with it, also it can give some of the characteristics and features based on a queried requirement, then retrieve out the required images based on the given eigenvalues. In all retrieval process features extraction is the most sensitive thing, because depend on the features and got the good matches image from the image's database [5-13]

2.2.3.1 color histogram

Content The intensity histogram shows how individual brightness levels are occupied in an image; the image contrast is measured by the range of brightness levels. The histogram plots the number of pixels with a particular brightness level against the brightness level. For 8-bit for each color band in all pixels, the brightness ranges from zero (black) to 255 (white) [14]. Color histogram represents the joint distribution of three color channels in an image (Red, Green and Blue). Therefore it characterize the global color information in an image [15]. In this proposal work will used RGB color space and for each band will calculate the set of features depend on the histogram. The histogram is used a model of the probability distribution of gray level. The first order histogram probability $P(g)$ is defined as follow:

$$p(g)=N(g)/M$$

Where $P(g)$ a probability of a gray level g in image

$N(g)$ number of pixel with gray level g in image

M total number of pixel in image

2.2.3.2 Histogram of Prime Value

A histogram for 8 bits per each color of the pixel 24 bit per pixel start from black color with 0 value into 255 (white) and between 0 and 255 there is 54 prime number. In some cases in this proposal work the retrieval based on prime's values will depend on these values of primes and their frequency to compare between query image and database's images.

2.2.3.3 Color Image Moment

Image moments and their functions have been used as features in many image processing applications, viz., pattern recognition, image classification, target identification, and shape analysis. Moments of an image are treated as region-based shape descriptors [2]. One of the moment that paper considered it in this work is mean is the average value, so tells us something about the general brightness of the image. The mean can be defined as follows:

$$Mean = \frac{1}{N} \sum_{j=1}^M \sum_{i=1}^N f(i,j)$$

$$\mu_X = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \dots\dots(2,7)$$

And the second moment has been used it is a standard deviation which also known as the square root of the variance tell something about the contrast. It describes the spread in the data, image with high contrast should have a high standard deviation and it is defined as

$$\sigma = \sqrt{\frac{1}{n} \sum_{j=1}^M \sum_{i=1}^N (x_{i,j} - \mu)^2}$$

$$\sigma_X = \sqrt{\frac{1}{n} \left\{ \sum_{i=1}^n X_i^2 - \frac{1}{n} \left(\sum_{i=1}^n X_i \right)^2 \right\}} \dots\dots(2.8)$$

And the third and fourth features that consider it is variance and entropy is a measure of how evenly energy is distributed in an image or data, entropy provides a measure of the amount of energy that cannot be used to do work and it can be formulated

$$e = - \sum_{i=1}^L P(z_i) \log P(z_i)$$

2.2.4 Proposed work

In this proposal work have been used three ways to retrieve image from database of images all of them depend on color histogram with add some extra features to increase the efficiency of work

1 .Image Retrieval using Color Histogram

A color image histogram with 24 bits per pixel starting from 0 to 255 from black to white for each color band (Red , Green and Blue) and because the changing in color value + 15

or -15 not visible with human eyes. So in the first Experiment work in this paper divided histogram into sub block histogram with 17 block each block contain 15 gray scale color for each of the RGB component of each image read from database. Thus each image will have 3 histogram associated with it.

2 .Image Retrieval using Prime value of Color Histogram

The second experiment in our work is take the frequency of the primes number between 0-255 for all images in database and query image and apply the Euclidian distance between two features vector for each color (R,G and B) so in the actual work will compare 162 value of each vector and the smallest Euclidian value select it as the optimal match.

3.Image Retrieval using statistical measurment

The third experiment depend on the histogram value for three band (Red, Green and Blue) will compute the set of moment (mean , Entropy , Variance and Standard Deviation)of the color image for each band and apply the measurement distance to check which image is more closer to the query image from database.

In general the algorithm that followed it in this work as the following outline steps with in each time different features used.

- 1.Read images in database and extract RGB format pixel information from images.
- 2 .Extract the features from the images in database.
3. Read in a query image and extract RGB format pixel information
- 4 .Create histograms for each of the RGB components of the query image.
- 5 .Compute a Euclidean distance by comparing the query image histograms to that of each image in the database.
6. Sort images in database in order of ascending Euclidean distance to query image and return as result, and the image with smallest Euclidean distance is the optimal match.

Note the different techniques that used it depend on the features and how can extract it from the images. And the figure (2.1) show the proposal work.

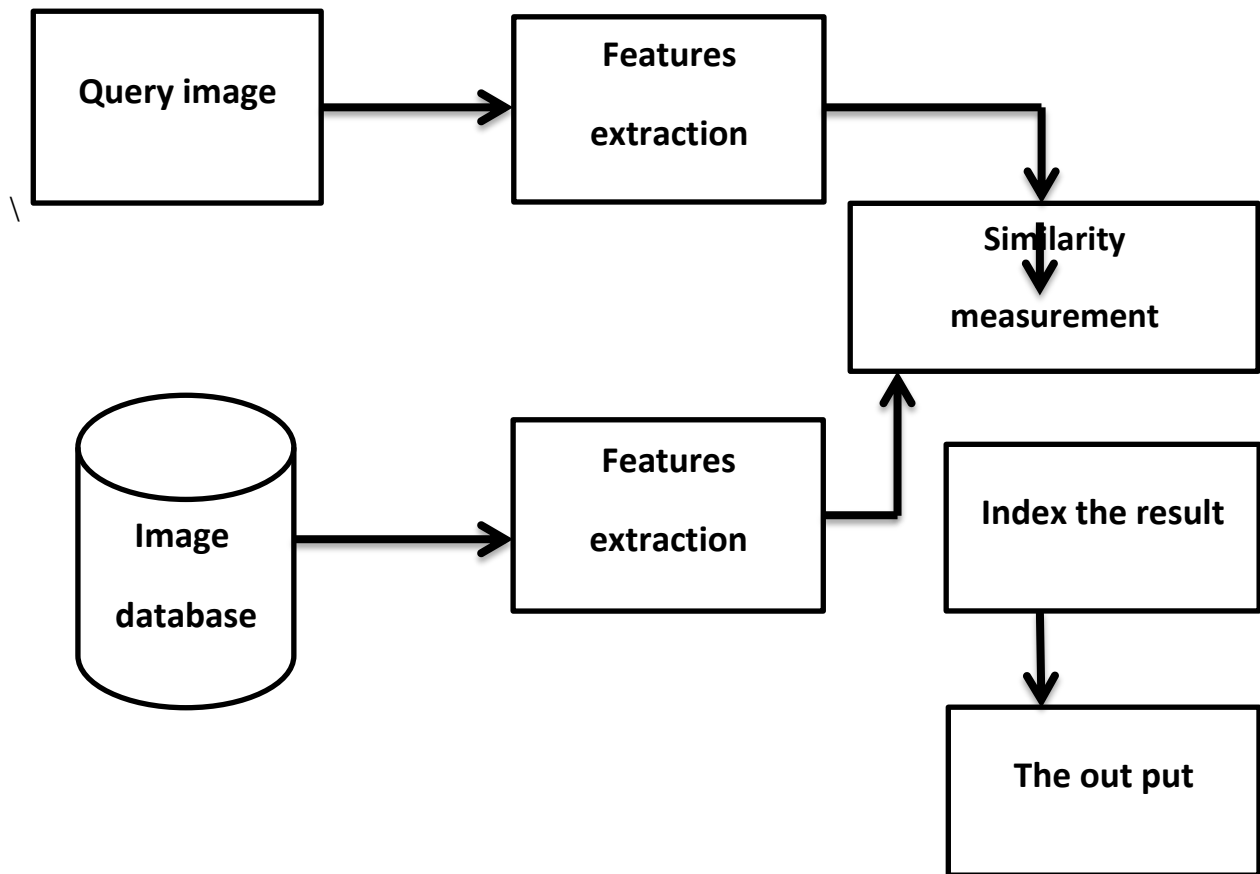


Fig2.2: image retrieval proposal work

2.2.5 Simllarity measurement

In the retrieval process , features of the query image specification are compare with features of the image database to determine which image match correctly (similar) with the given features. The matching process is based on similarity measure between query image and images in database. There are several similarity measures that are proposed and used for varied purpose. For retrieval base on similarity of two features distance between the two can be defined as a match measure[16] . It is the way that compute the similarity measure between query image vector and the vector of database's images, where each vector contain the features that have been selected it in different proposal work in this paper have been used the Euclidean distance to select which images are closer to the query image.

2.2.5.1 Euclidian Distance

It's the most common metric for measuring between two vectors and is given by the square root of the summation of the squares of the different between vectors component Euclidian Distance $(A.B) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$ Where $A = [a_1, a_2, a_3 \dots a_n]$ $B = [b_1, b_2, b_3, \dots, b_n]$ Smallest Euclidian Distance between two vectors of features means the most probable matching and will be selected.

$$WED = \sum_{i=1}^N \frac{(f_i - f_i^{(k)})^2}{(\delta^{(k)})^2} \dots \dots \dots (2,9)$$

2.2.5.1 Cosine

The Cosine distance measure the similarity between two vector with non-zero vector integer or Boolean component, in the space a point may be through of as a direction. The cosine measurement distance between two vector is an angle between point and vector that make. This angle's degree between 0 to 180, depend on how many dimensions the space has. Where the Cosine of 0 is 1 and it is less than one for any other angle in positive space. If we have two vectors with the same values a cosine similarity of 1, and two vectors at 90° have a similarity of 0, and two vectors diametrically opposed have a similarity of -1. The equation number () show the formula of Cosine distance.

$$\cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \dots \dots \dots (2,10)$$

Conclusion

In this part of our work ,we study and implement the first part of content based image retrieval that based on the color and histogram of the color . basically we used three experiment first one based on histogram . second based on only the prime value between [0_255] of histogram and third some feature extracted form histogram

References

1. Sharmin Siddique, "A Wavelet Based Technique for Analysis and Classification of Texture Images," Carleton University, Ottawa, Canada, Proj. Rep. 70.593, April 2002.
2. Thomas Seidl and Hans-Peter Kriegel, "Efficient User-Adaptable Similarity Search in Large Multimedia Databases," in Proceedings of the 23rd International Conference on Very Large Data Bases VLDB'97, Athens, Greece, August 1997.
3. FOLDOC, Free On-Line Dictionary Of Computing, "co-occurrence matrix," May 1995, Available at: <http://foldoc.doc.ic.ac.uk/foldoc/foldoc.cgi?cooccurrence+matrix>
4. Henning Muller, Nicolas Michoux, David Bandon and Antoine Geissbuhler , A Review of Content Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions, Division for Medical Informatics, University Hospital of Geneva.
5. Rishav Chakravarti, Xiannong Meng , A Study of Color Histogram Based Image Retrieval , Sixth International Conference on Information Technology: New Generations IEEE 2009.
6. Reshma Chaudhari, A. M. Patil , Content Based Image Retrieval Using Color and Shape Features, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2012.
7. H. Mohamadi, A. Shahbahrami, J. Akbari, " Image retrieval using the combination of text-based and contentbased algorithms", Journal of AI and Data Mining, Published online: 20, February-2013
8. Pabboju , S. and gopal, R. (2009). A Novel Approach For Content- Based Image Global and Region Indexing and Retrieval System Using Features. International Journal OF Computer Science and Network Security. 9(2), 15-21.
9. . Li, X., Shou, L., Chen, G., Hu, T. and Dong, J. (2008). Modelling Image Data for Effective Indexing and Retrieval In Large General Image Database. IEEE Transaction on Knowledge and Data Engineering. 20(11), 1566-1580.
10. Demerdash, O., Kosseim, L. and Bergler, S. (2008). CLaC at Image CLEFphoto 2008, ImageCLEF Working Notes
11. Patheja P.S., Wao Akhilesh A. and Maurya Jay Prakash, "An Enhanced Approach for Content Based Image Retrieval", International Science Congress Association,

Research Journal of Recent Sciences, ISSN 2277 – 2502 Vol.1(ISC-2011), 415-418, 2012.

and Designers. New York: 12. P. Brodatz, Textures: A Photographic Album for Artists Dover Publications Inc., 1966.

13. P. P. Ohanian and R. C. Dubes, "Performance evaluation for four classes of textural features," Pattern Recognition, vol 25, pp. 819-33, 1992.

14. Mark S. Nixon and Alberto S. Aguado , Feature Extraction and Image Processing, British Library Cataloguing in Publication Data ,2002 first edition.

15. Xiuqi Li ,Shu Ching Chen Mei Ling Shyu Borko Furht , an effective content based visual image retrieval system , Department of Electrical and computer Engineering ,University of Miami ,2002.

16. Sridhar R. Avula ,Jinshan Tang ,Scott T. Acton , Image retrieval using segmentation , Proceedings of the 2003 Systems and Information Engineering Design Symposium ,2003.