Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 05, May -2018

CONTENT BASED IMAGE RETRIEVAL USING NUMERICAL SHARING OF COLOR HISTOGRAM METHOD

Dr P.J.Arul Leena Rose

Associate Prof, Dept of Computer Application, FSH, SRMIST

ABSTRACT:- *Content based image retrieval is aimed at efficient retrieval of relevant images from large image databases based on automatically derived imagery features. These features are typically extracted from shape, texture or color properties of query image and the images in the database (target images). The relevance between the query image and any target image is ranked according to a similarity measure computed from the features.*

This research paper makes a Content based image retrieval scheme with an enhanced algorithm. In this paper, I propose a new method by approximating the statistical distribution of color histogram for image indexing and comparison and the method is called "Content Based Image Retrieval by Statistical Distribution of Color Histogram Method". This research work is developed in IDL 6.3 for its better support in digital image processing.

Key words : *Content based image retrieval, Windows Metafile*

1. DIGITAL IMAGE PROCESSING

The field of image processing continues, as it has since the early '70s, on a path of dynamic growth in terms of popular and scientific interest and number of commercial applications. Considerable advances have been made over the past 30 years resulting in routine application of image processing to problems in medicine, manufacturing, entertainment, law enforcement, and many others. Examples include mapping internal organs in medicine using various scanning technologies (image reconstruction from projections), automatic fingerprint recognition (pattern recognition and image coding), and HDTV (video coding), to name a few.

Content based image retrieval (CBIR) has been largely explored in the last decade. In CBIR context an image is represented by a set of low level visual features, which have no direct correlation with high level semantic concepts, and the gap between high level concepts and low level features in the major difficulty that hinders further development of CBIR systems.

Compared with other machine learning problems, CBIR online learning has three challenges. (1) Small size of the training set (2) Intrinsic asymmetry (3) Fast response requirement.

The term image, refers to a two-dimensional light intensity function f(x,y), where x and y denote spatial coordinates and value at any point (x, y) is proportional to the brightness of the image at that point.

Digital Image

A digital image can be considered as a matrix whose row and column indices identify a point in the image and the corresponding matrix element values identifies the gray level at that point.

Based on the way that image data is saved, images can be split into 3 different types:
1) Bitmap **2)** Vector **3**) Metafile

Bitmap

Bitmaps images are exactly what their name says they are: a collection of bits that form an image

Vector

In vector graphics, the co-ordinates of images (lines and curves) are saved as mathematical data.

Metafile

Metafile graphics are simply 2D graphics that are made up of both vector and bitmap**.**

The elements of the general-purpose system capable of performing the image processing operations are:

1. Image Acquisition
2. Image Storage
3. Processing the image
4. Communication
5. Display

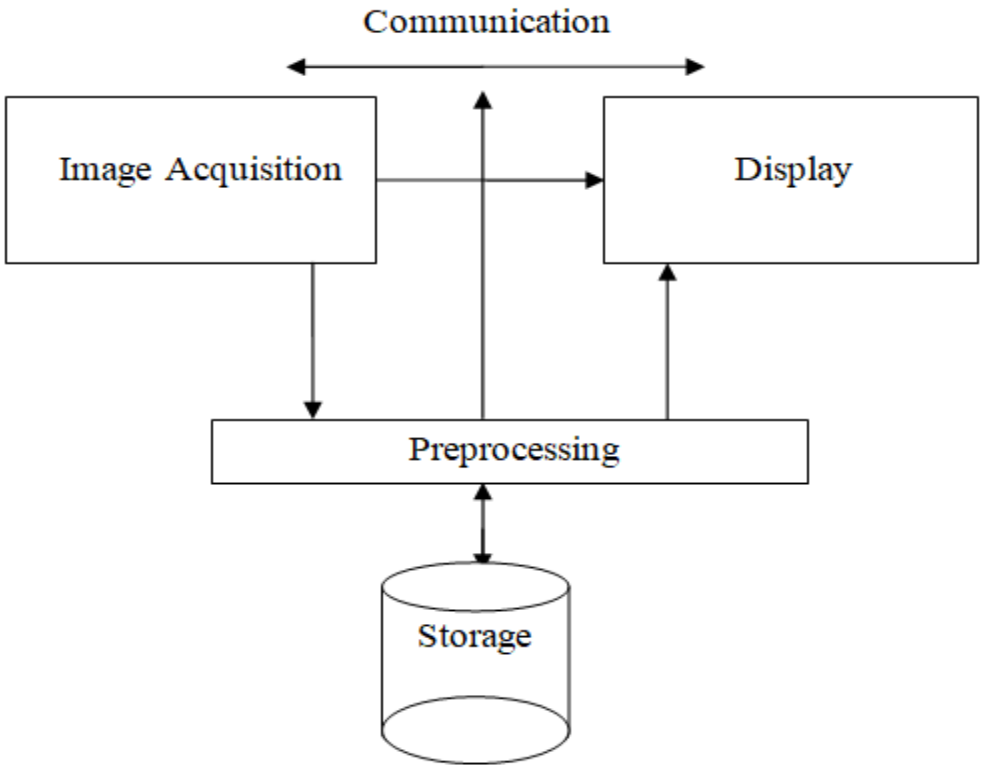


Fig. 1. Image processing operations

2. STEPS IN IMAGE PROCESSING

There are seven steps involved in the digital image processing.

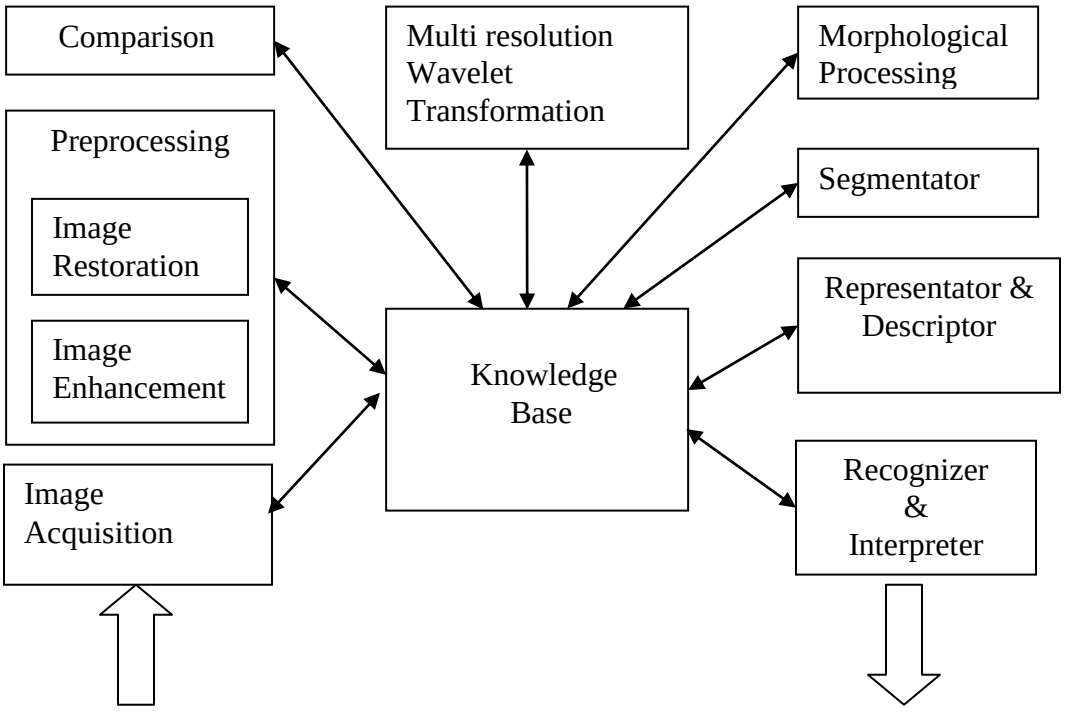


Figure 2. Steps involved in Image processing

Image acquisition is the process of acquiring a digital image. To acquire an image we require an imaging sensor and the capability to digitize the signal produced by the sensor.

Figure 3. Image Acquisition

The key function of preprocessing is to improve the image in ways that increase the chances of the other process.

(i) Image enhancement

1) To provide more effective display of data for visual interpretation human eye can distinguish up to 40 grey shades

2) Increase the visual distinction between features in a scene.

3) "Digital darkroom" techniques

(ii) Image rectification and restoration

1) Correction of geometric distortions,

2) Calibration of data,

3) Elimination of noise

4) Correction of blurring,

3. CONTENT BASED IMAGE RETRIEVAL BY STATISTICAL DISTRIBUTION OF COLOR HISTOGRAM METHOD

The main purpose of image retrieval is to find the same or the images with higher similarity from the database compared with the user provided query image. The most important factor what we concern about is the retrieval accuracy. On the other hand, what we should also take into consideration when retrieving images are the time and complexity. In the past few years; the general problems of content-based image retrieval have been widely studied. The content-based visual query (CBVQ) techniques indexed images and videos by the visual features to provide the search capabilities for a system.

Moreover, the process of content-based image retrieval has been investigated in several recent efforts such as Visage's commercial image retrieval system. The Chabot image retrieval system**,** IBM's QBIC project, and MIT's Photobook system. The QHIC system has investigated many features such as color, texture, shape and so on. In addition, the system has placed a research

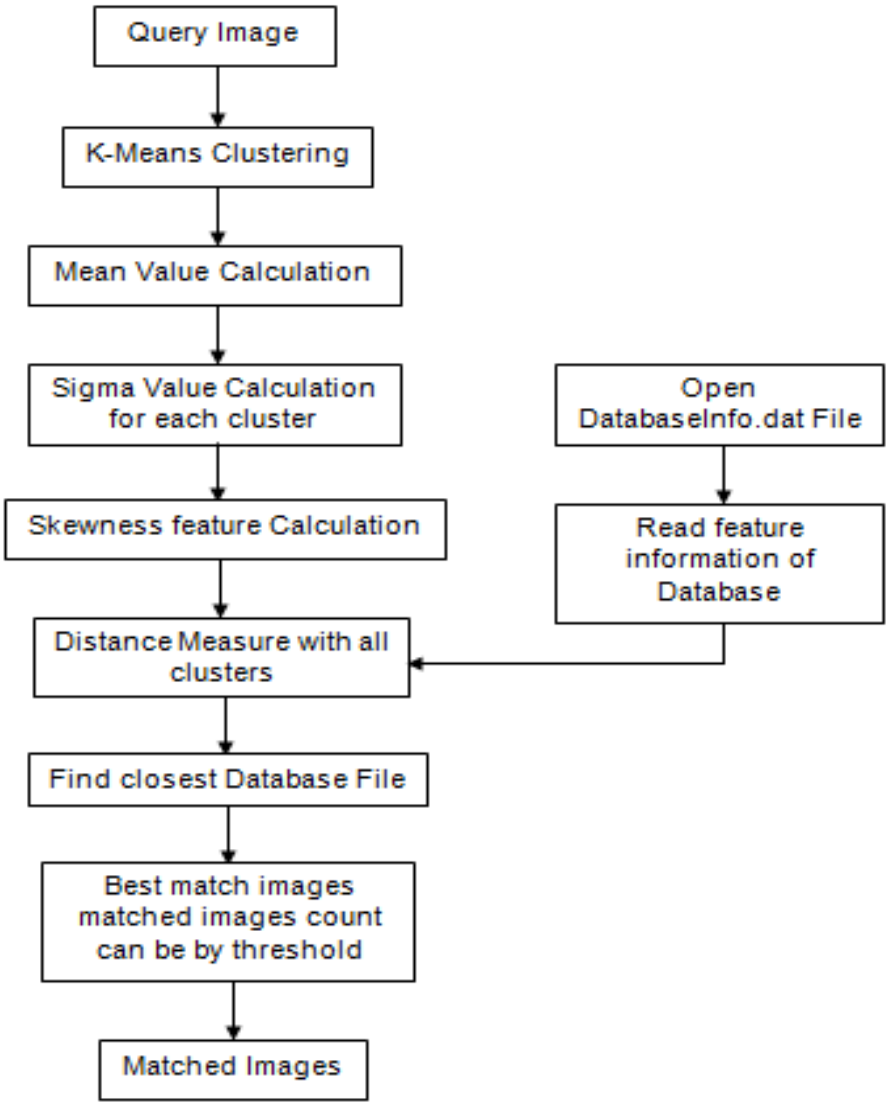


Figure 4. Block diagram for the Image Retrieval section of "Content based image Retrieval by Statistical Distribution of Color Histogram Method"

focus on designing strategies for indexing the features. Some systems allowing user to sketch the outlines of objects to find out the desired images are also considered. For example, the QVE (Query by Visual Example) system first normalizes the images in the database as regular-sized images and then processes them to get abstract images. Currently, finding effective methods to incorporate content-based image retrieval approaches into an image database system is an attractive research. Lire ignored in the histogram based retrieval methods. Hence, we consider the distribution of color histogram for image retrieval in this method.

3.1 HISTOGRAMS AND STASTICS

Histogram

Recently, the histogram comparison has become a popular technique for image and video indexing**.** Swain and Ballard proposed the histogram intersection for matching color objects based on the hypothesis that similar images almost have higher similarity among their color distributions. Besides, in comparison to the complexity of traditional techniques of pattern recognition, it seems that the histogram- based techniques have lower complexity especially for felicitating real time implementation. In the histogram-based techniques, the images with least difference of histograms are considered as the retrieval results or for further search. In general, the distance between two histograms Q and $\boldsymbol{I}$ in L^p metric can be expressed as

$$d_{L^p}(Q,I) = \sum |Q - I|^p$$

Typically, the most frequently metric used in the indexing techniques is the absolute error *(L^1)* or square error *(L^2)*. In the histogram intersection technique, the intersection of two histograms is defined as

$$h_{c_i}(Q \cap I) \triangleq \frac{\min\{H_{c_i}(Q), H_{c_i}(I)\}}{|Q|}$$

where *Q* and ***I*** are the histograms of query image and I image in the database, $H_{c_i}(Q)$ and $H_{c_i}(I)$ are defined as the amount of pixels for color c_i in the query image and image in the database

Some statistics properties

The histogram can be considered as *a* kind of empirical distribution of probability for a random variable. In this method we focus on the viewpoint of making use of this property.

Basic Statistics: In general, some of the well known statistics properties such as mean (μ), standard derivation (σ) and skewness (S_k, Pearson's coefficient of skew ness) are widely used for the feature representation of images. They are listed as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \;,\; \sigma = \left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2\right]^{1/2}, \text{ and } S_k = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3$$

The coefficient of skewness is used for evaluating the skew direction and degree for a unimodal curve. It is sometimes called *a m*easure of asymmetry. It can be shown that the skewness of a symmetrical distribution is equal to zero.

Mixture probability density function

Generally speaking, the probability density function can be considered as the representation of a distribution of random variable. For example, the probability density function with mean, μ and variance σ^2 of Gaussian distribution is defined as

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right]$$

In this method, we consider the histogram *as a* kind of distribution represented by mixture Gaussian probability density function. The mixture Guassian probability density function is defined as follows:

$$p(z) = \frac{p_1}{\sigma_1\sqrt{2\pi}}\exp\left[-\frac{(z-\mu_1)^2}{2\sigma_1^2}\right] + \ldots + \frac{p_N}{\sigma_N\sqrt{2\pi}}\exp\left[-\frac{(z-\mu_N)^2}{2\sigma_N^2}\right]$$

where μ_i is the mean value of each probability density function, σ_i is the standard derivation about the mean for i=1, …, N, and $p_1, p_2, \ldots, p_N$ satisfies the constraint $p_1 + p_2 + \ldots + p_N = 1$. An example of mixture Gaussian distribution is shown in the following figure.

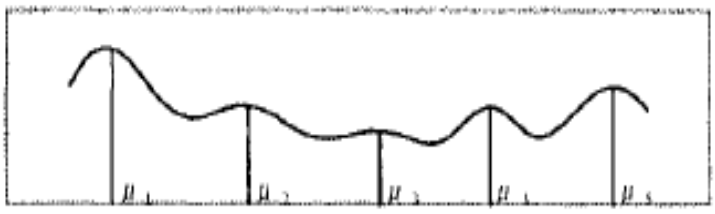


Figure. Mixture Gaussian probability density function

The histogram comparison has become a popular indexing technique for image retrieval. However, we *can* find that either the L^p metric or histogram intersection for comparing images docs not take the relationship among neighboring color bins into consideration. Their distance measure only concentrates on the difference of each bin itself while ignores their relationships. Hence, this method makes use of the relationship to improve the performance of retrieval.

In this method, the histogram bins are first classified into *K* clusters by the well-known K-means algorithm. For each cluster, we evaluate its mean value as the axial value C_i. Then, we evaluate the feature values mean μ_i standard derivation σ_i, and skewness S,, for the *K* axes according to the histogram. They are calculated from the neighboring *N* histogram bins of each axis. The feature vectors of an image in this method are defined *as* $f_i=(C_i, \mu_i, \sigma_i, S_{ki})$ for $i=1,\ldots, K$. They can be considered *as* the representation of some properties of the mixture Gaussian distribution for the histogram.

Image retrieval

Applying the extracted feature vectors for image retrieval, the feature vectors of an image are first extracted as the indices for each image in the database. When we want to find suitable images from the database for a query image *Q*, the feature vectors based on each axis of histogram bin of *Q* are first evaluated. Then, the *K* feature vectors of each image in the database are compared to the corresponding axial feature values of the query image. Those images with smaller differences are then selected as the query result of the image. The technique can be also applied for the partial image query. The main reason is that the histogram of partial image will be comprehended in the histogram of the original image. They will almost have the same distribution properties. Hence, it is applied to the partial image query directly.

Distance Measures

The distance measure of this method is defined as follows:

$$dis = \sum_{i=1}^{K}\left[(\mu_{c_i}-\mu_{q_i})^2 + w_1(\sigma_{c_i}-\sigma_{q_i})^2 + w_2(S_{k_{c_i}}-S_{k_{q_i}})^2\right]$$

where μ_{c_i}, σ_{c_i} and $S_{k_{c_i}}$, arc the feature values of image in the database on axis i, μ_{q_i}, σ_{q_i} and $S'_{k_{q_i}}$, are the feature values of query image on axis i. and w_1 and w_2 are the weighting coefficients. In this method, the weighting coefficient w_1 and w_2 are set to 100. The main purpose of using these factors are that the standard derivation and skewness are usually with a small value.

This methodology is independent with rotation and scaling. The comparison with the previous method can be view from the following table:

	Full images				Partial images			
	Full	Rota-tion	Scal-ing	Aver-age	Par-tial	Rota-tion	Scal-ing	Aver-age
Histogram intersection	100%	90%	100%	96%	37%	33%	26%	32%
Proposed method	100%	93%	100%	98%	70%	63%	57%	63%

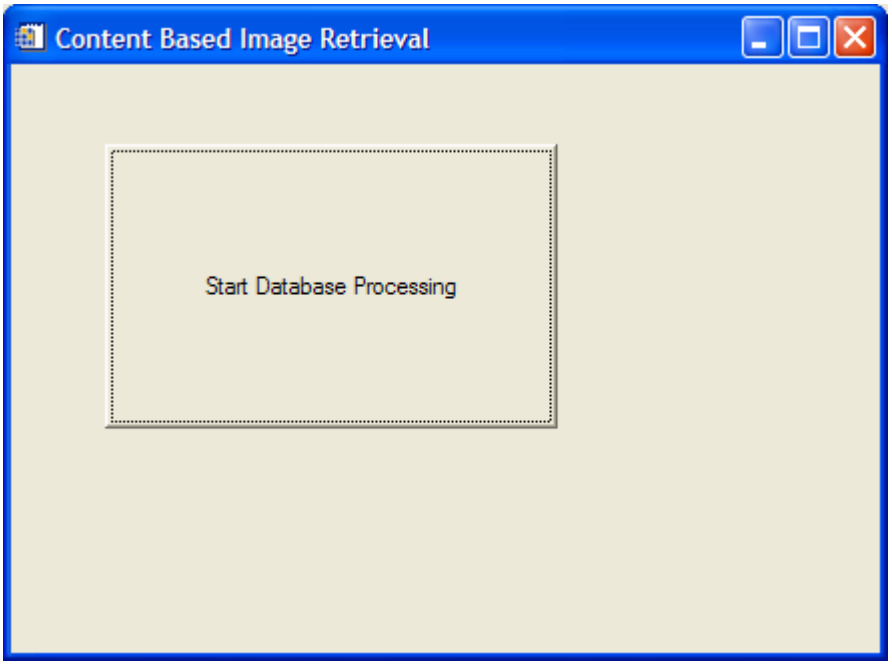
Table: 1. Comparison of different methods

In this method, we propose an efficient and accurate image retrieval technique by approximating the statistical distribution of color histogram for image indexing and comparison. The main advantage is that the performance of the basic histogram based indexing technique has been improved by this method. It becomes simple to compute the distance measure of two images by this method. Also, this method tolerates the generally considered problems when retrieving images or partial images such as transition, scaling, and rotation. The represented features of the approximated distribution of color histogram can also be used as the indices for image database. Experimental results show that this method is quite effective not only for the performance but also better results for image indexing and retrieval.

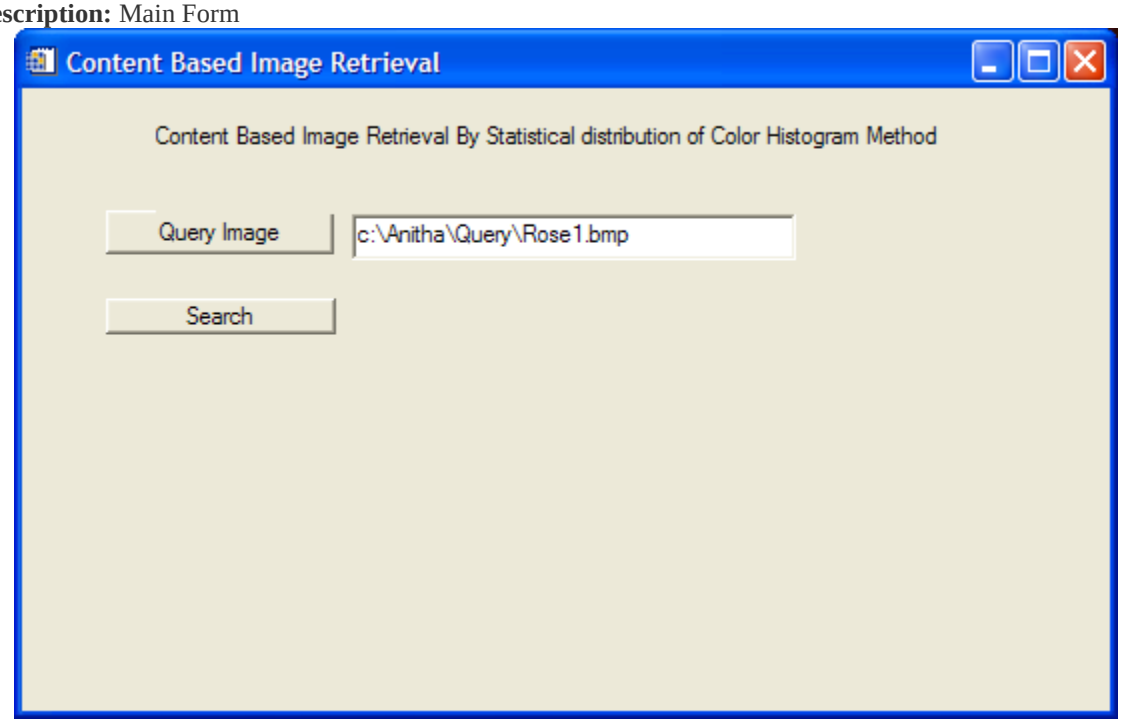
SAMPLE SCREENS :SAMPLE SCREEN SET FOR ROSE1.BMP

Method Name: Content Based Image Retrieval by Statistical

Image Description: Database Processing Main Form



Method Name: Content Based Image Retrieval by Statistical
Image Description: Main Form



Method Name: Content Based Image Retrieval by Statistical Distribution of Color Histogram Method
Image Description: Input Query image

Method Name: Content Based Image Retrieval by Statistical Distribution of Color Histogram Method
Query Image Name: Rose1.bmp
Image Description: Clustered Image

4. ANALYSIS

This section deals with the Analysis of Content based image retrieval by statistical distribution of color histogram method

An analysis on time taken for the three Content Based Image Retrieval Methods namely, "Content based Image Retrieval by Statistical Distribution of Color Histogram Method", "Content based Image Retrieval by Euclidean Distance Method" and "Content based Image Retrieval by Hamming Distance Method" for the test query image Rose.bmp

Test Query Image	Content Based Image Retrieval Methods	Time Taken (in Seconds)
Rose.bmp	Content based Image Retrieval by Statistical Distribution of Color Histogram Method	4.3120

An analysis on time taken for the three Content Based Image Retrieval Methods namely, "Content based Image Retrieval by Statistical Distribution of Color Histogram Method", "Content based Image Retrieval by Euclidean Distance Method" and "Content based Image Retrieval by Hamming Distance Method" for the test query image Man.bmp

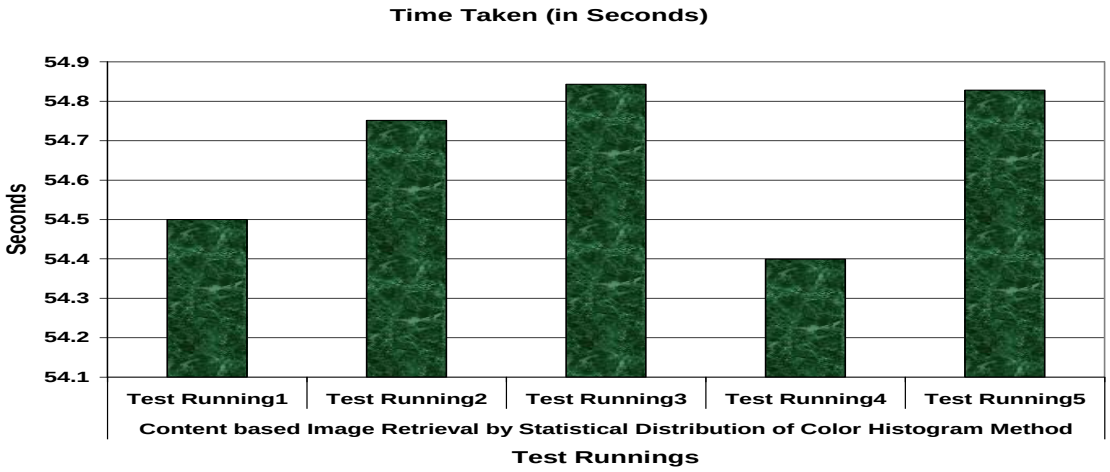
Test Query Image	Content Based Image Retrieval Methods	Time Taken (in Seconds)
Man.bmp	Content based Image Retrieval by Statistical Distribution of Color Histogram Method	4.0310

An analysis on Error Ratio for the three Content Based Image Retrieval Methods namely, "Content based Image Retrieval by Statistical Distribution of Color Histogram Method", "Content based Image Retrieval by Euclidean Distance Method" and "Content based Image Retrieval by Hamming Distance Method" for the test query image Rose.bmp

Test Query Image	Content Based Image Retrieval Methods	Error Ratio (in Percentage)
Rose.bmp	Content based Image Retrieval by Statistical Distribution of Color Histogram Method	0%

An analysis on Database Processing time taken for the "Content based Image Retrieval by Statistical Distribution of Color Histogram Method for the five time running of source code

Content Based Image Retrieval Methods	Running Time	Time Taken (in Seconds)
Content based Image Retrieval by Statistical Distribution of Color Histogram Method	Test Running1	54.500
	Test Running2	54.751
	Test Running3	54.843
	Test Running4	54.399
	Test Running5	54.8280



An analysis on Accuracy Ratio for the three Content Based Image Retrieval Methods namely, "Content based Image Retrieval by Statistical Distribution of Color Histogram Method", "Content based Image Retrieval by Euclidean Distance Method" and "Content based Image Retrieval by Hamming Distance Method" for the test query image Rose.bmp

Test Query Image	Content Based Image Retrieval Methods	Accuracy Ratio (in Percentage)
Rose.bmp	Content based Image Retrieval by Statistical Distribution of Color Histogram Method	100%

5. CONCLUSION AND FUTURE ENHANCEMENT

The main purpose of image retrieval is to find the same or the images with higher similarity from the database compared with the user provided query image (or creating one by drawing). The most important factor what we concern about is the retrieval accuracy. This paper tries to increase the retrieval accuracy.
This paper, explore the issue of feature selection in the CBIR learning context. Through Content Based Image Retrieval by Statistical Distribution of Color Histogram Method .
Instead of K-Means clustering method we use Fuzzy K-Means clustering method to get better clustering. In this paper we used 256 x 256 dimension, it can be extended to any dimension. In this paper we used Bitmap format images. It can be extended to compressed jpeg files also

6. REFERENCES

1) V. E. Ogle and M. Stonebraker, "Chabot: Retrieval from a relational database of images," *IEEE: Computer*, vol. 28, no.9, pp. 40-48, Sept. 1995.
2) K. K. Srihari, "Automatic indexing and content-based retrieval of captioned images," *IEEE: Computer*, vol. *28*, no. 9
3) D. Petkovic, P. Yanker, and C.Faloustos, "The QBIC project: Querying images by content using color, texture, and shape," In *Storage and Retrievel for image and and Video Database*, vol. 1908
4) A. Pentla, R.W. Picard, and *S.* Sclaroff', "Photobook: Tools for content-based manipulation of image database," In *Proceeding of the SPIE Storage and Retrieval Image and Video Database II*, Feb. 1994.
5) J. Hafner, H. *S.* Sawhney, W. Equitz, M. Flickner. and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEE'E Trans. Pattern Anal machine Intelli.,* July 1995.
6) T. Tako, T. Kurita, N. Otsu, and K. lirata, "A Sketch retrieval method for full color image database: Query by visual example," in *Pvoc. International Conference on Pattern Recognition,* 1992, pp. 530-533.
7) G. L. Gimel' farb and A. K. Jain, "On retrieving textured images from *an* image database," *Pattern recognition,* vol. 29, no. 9, pp. 1461-1483, 1996.
8) V. N. Gudivada and V. V. Raghavan, "Design and evaluation of algorithms for image retrieval by spatial similarity," *ACM Trans. on information Systems,* vol. 13, no. 2, 1995.
9) D. H. Huang and S. F. Chang, "A content-based image retrieval system," *Proceedings of' IPP'R Conference on Computer Vision, Graphics, and image processing,* Taiwan, pp259-266, 1996

