



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume: 6**

**Issue: IX**

**Month of publication: September 2018**

**DOI:**

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# Perceptual Hashing for Content based Image Retrieval

Meenalochini. M<sup>1</sup>, Gomathi. N<sup>2</sup>

<sup>1,2</sup>Assistant Professor, Jaishriram Engineering college, Tirupur, Tamil Nadu, INDIA.

**Abstract:** Content Based Image Retrieval (CBIR) is becoming effective source for fast retrieval on Image database for current internet world as it contains large collections of images. Content Based Image Retrieval (CBIR) is a technique which uses visual features of image such as color, shape and texture to search user required image from large image database according to user's requests in the form of a query image. In this paper, CBIR using Perceptual hashing is implemented with feature extraction techniques like Color histogram, Gabor filters, and Canny's edge detection. The hash codes are generated for feature extracted images and similarity measure is computed for retrieved images. By using hashing codes, we can achieve linear search time complexity and can speed up the retrieving process.

**Keywords:** Content based image retrieval (CBIR), Perceptual hashing, Color histogram, Gabor filter, Canny's edge detection.

## I. INTRODUCTION

Content based image retrieval is efficient and fast growing research area which helps to retrieve images from large database automatically. CBIR works directly on images and not on the text annotation of the image. With this, CBIR system can be fully automated by the indexing of the image database and free from human dependence. In general, User make query on CBIR system by submitting a query image. In the background, the system evaluates the similarity between the query image and the images in the database. In general, CBIR system use features like color and texture for image indexing and retrieval as well as shape information, spatial relationship between regions, salient points and so on. The information gained by feature extraction is used to measure similarity between the images. Images are represented by points in the high dimensional feature space. Each extent of the feature corresponds to one dimension in the feature space. A metric is defined to calculate the actual similarity between two points. This section introduces three features: color, texture and shape, which are used most often to extract the features of an image.

### A. Color

The color feature is one of the most simple and widely used visual feature in content-based image retrieval. It is relatively robust to background complication and independent of image size and orientation. The most universal method is color histogram. This technique describes the proportion of pixels of each color in an image. Color histogram is divided into global color histogram and local color histogram. A color histogram is a type of bar graph, where each bar represents a particular color of the color space being used. The bars in a color histogram are referred to as bins and they represent the x-axis. The number of bins depends on the number of colors in an image. The y-axis denotes the number of pixels that are in each bin.

### B. Texture

Texture is innate property of all surfaces that describes visual patterns, each having properties of homogeneity. It contains important information about the structural arrangement of the surface. Gabor wavelet is widely adopted to extract texture from the images for retrieval. Basically Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and specific orientation. The scale and orientation tunable property of Gabor filter makes it especially useful for texture analysis. These enable filtering in the frequency and spatial domain.

### C. 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. Shape feature can be generally divided into two categories, Boundary-based and Region-based. Boundary-based representation uses the outer boundary of the shape. So to reduce complexity, image is converted into gray scale image. Edge Detection is carried out at the sharp edges on the images. Edges in images are areas with strong intensity. Here, edge detection is done by using Canny's algorithm.

## II. RELATED WORKS

Content based image retrieval for general-purpose image databases is a highly challenging problem because of the large size of the database, the difficulty of understanding images, both by people and computers, the difficulty of formulating a query, and the issue of evaluating results properly.

Suresh Kumar natarajan et al, implemented CBIR using Perceptual hashing for medical images that result in accurate retrieval of images.

Henning Mullar et al, used techniques that help in retrieval of medical images in field of radiology. Using Hash function by Harsh Kumar Sarohi st al, gives very fast retrieval time and low computation cost. This is done by applying feature extraction and generating hash value to retrieve images through similarity measure.

Christoph Zauner analyzed different functions of Perceptual hashing. Here, Block mean value based function is used that helps to retrieve images faster and provides high performance speed. Hash value is generated as binary sequence and similarity between two hash values are measured.

Abhijeet Mallick et al, implemented feature extraction techniques with different metrics that leads to high accuracy. The feature extraction for the image was implemented in color techniques. The techniques for other features like shape and texture were also implemented for the retrieval of images.

Akshay Alex et al, improved retrieval process directly on image solving text based process. Using feature extraction, the images are retrieved faster and automatically process on image. Low-level features are implemented for feature extraction which helps to retrieve images without keywords.

Text based retrieval doesn't provide good process in sketched images. CBIR achieves using features for various applications. The edge histogram descriptor used by Pavani,S.,et al, results in poor sketches and more similarity is to be achieved.

Mangijao Singh,S.,et al, implemented the texture and color features during segmentation from pixels or small blocks. Such features do not properly represent the property of an entire region; thus it is necessary to study feature extraction from the whole region after segmentation.

Ramesh Kumar , A., et al, analyzed that the color histogram is constructed by quantizing the colors within the image and by counting the number of pixels of each color. The feature vector of an image has to be derived from the histograms of its color components. Then it has to set the number of bins in the color histogram to obtain the feature vector of desired size.

Navneet Kaur ,et al, described that the retrieval of information has to be given to the users along with the description of the result sets.

## III. METHODOLOGY

To search and retrieve images from large database, CBIR system is introduced to solve the problems of text-based image retrieval. CBIR uses low level features like color, texture and shape. The overall proposed block diagram is given in Fig.1. To extract features of image color histogram, Gabor filter and Canny's edge detection are used. The features are extracted from query image and images in database. The Hashing is used to speed up the retrieving process. Similarity measure is implemented using Euclidean distance. Then, the similar images are retrieved.

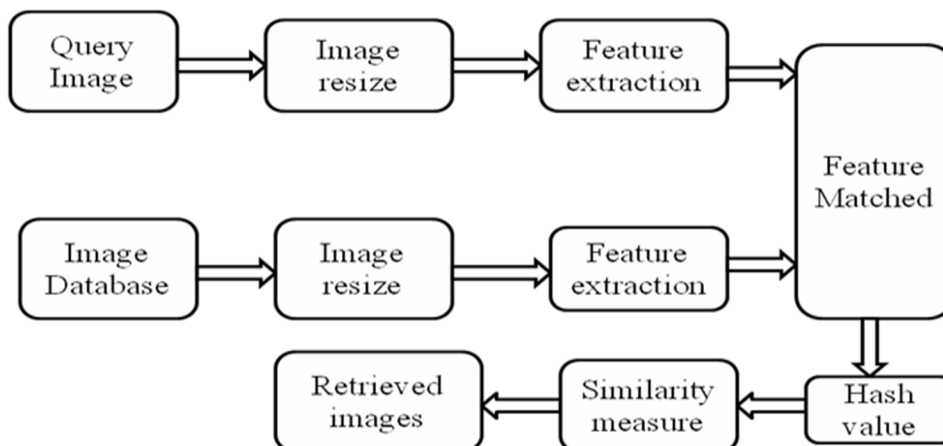


Fig.1 Block Diagram of Proposed work

#### IV. FEATURE EXTRACTION

Feature extraction plays an important role in content-based image retrieval to support for efficient and fast retrieval of similar images from image databases. The feature is defined as a function of one or more measurements, each of which specifies some quantifiable property of an object, and is computed such that it quantifies some significant characteristics of the object. The low – level features are color, texture, and shape.

#### V. COLOR HISTOGRAM

The color histogram is obtained by quantizing image colors into discrete levels and then counting the number of times each discrete color occurs in the image. In color histograms, quantization is a process where number of bins is reduced by taking colors that are similar to each other and placing them in the same bin. Each image in the database is computed to obtain the color histogram, which shows the proportion of pixels of each color within the image. The search by specifying the query image, the system registers the proportion of each color of the query image and goes through all images in the database to find the images whose color histograms match with the query most closely. A color histogram H for a given image is defined as a vector:

$$H = \{H[0], H[1], H[2], \dots, H[i], \dots, H[n]\} \quad \text{--- (1) \rightarrow}$$

Where ‘i’ represents color in the color histogram [i] is the number of pixels in color ‘i’ in that image and ‘n’ is the number of bins in the color histogram.

#### VI. GABOR FILTER

It is a linear filter used for frequency and orientation representation. Since this requires computation of binary orthogonal wavelets which result in time consuming. Therefore, filter with various scales and rotations are implemented resulting in array of magnitudes. The following mean and standard deviation of the magnitude of the transformed coefficients are used to represent the homogenous texture feature of the region.

The 2-D Gabor function can be specified by the frequency of the sinusoid  $W_x$  and the standard deviation  $\sigma_x$  and  $\sigma_y$ , of the Gaussian envelope as:

$$g(x, y) = \frac{1}{\sigma_x \sigma_y} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j W_x\right) \quad \text{--- (2) \rightarrow}$$

#### VII. CANNY’S EDGE DETECTION

In shape extraction, edge detection is carried out to get outer boundary of the image. Edge detection leads to removal of noise that result in blurred version of the image. Here edge detection is done by using Canny’s algorithm. The Canny edge detector finds noise present in the image by applying threshold. This results in boundary shape of the image.

#### VIII. PERCEPTUAL HASHING

Perceptual hashing is a robust algorithm widely used for content identification. Block-mean perceptual hashing algorithm is used where the algorithm takes the color pixels of an image to generate the hash value. The Perceptual hash is reliable and fastest algorithm to retrieve images. The hash value is generated by observing the color pixel and its average color pixel value. Hash bit is generated, with respect to the average of total color pixels.

#### IX. EUCLIDEAN DISTANCE

Euclidean Distance, a nearest neighbor classifier is used for classification. Find the maximum distances between the feature database and projected query image which provides similar images.

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^n [Q_i - D_i]^2} \quad \text{--- (3) \rightarrow}$$

Where Q and D are feature vectors of the Query image and database image.

#### X. RESULTS

In this section, the results obtained on Corel dataset. The experiments are carried out in Matlab with Intel processor 2.10 GHz with 3GB RAM.



### XI. DATASET

Sample images from Corel dataset are shown in Fig



Fig.2 Sample database images

### XII. PERFORMANCE ANALYSIS

The retrieval results are a list of images ranked by their similarities measured with the query image. The experiment was carried out with the number of retrieved images set as 10 to compute the average precision and recall.

Precision P is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images. Precision measures the accuracy of the retrieval.

$$\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of images retrieved}}$$

Recall is defined by R and is defined as the ratio of the number of retrieved relevant images to the total number of relevant images in the whole database. Recall measures the robustness of the retrieval.

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of relevant images}}$$

On an average the overall performance accuracy of the proposed method is found as 80%.

### XIII. EXPERIMENTAL RESULTS



Fig.3 result of image retrieval process (query image = 'BUS category')



Fig.4 result of image retrieval process (query image = 'TRAIN category')

DATASET	PRECISION	RECALL
Bus	0.5	0.5
Train	1.0	1.0
Buildings	0.7	0.7
Rose	0.5	0.5
Food	0.7	0.7

Table.1 The overall performance accuracy

The above fig.3 shows the result of the image retrieval using the feature set, when a query image of ‘BUS’ category is submitted to the system.

Total no. of images retrieved, T	=	10	
No. of relevant images retrieved, A	=	5	
No. of non-relevant images retrieved, B	=	5	
Precision	=	0.5	
Recall	=	0.5	

#### XIV. CONCLUSION

In this paper, content based image retrieval is done using hash codes. The Experimental results demonstrate that the proposed method has high accuracy. The result of precision and recall are shown in Table.1. The query image and retrieved results are closely matched with corresponding accuracy to 80%. Ever-changing Image database like Internet may dramatically increase retrieval time and slowly retrieve images. CBIR using Perceptual hashing overcomes these problems and retrieve images faster. Content- based Image Retrieval is an emerging technique but the application of CBIR is yet to attain advancement.

#### REFERENCES

- [1] Abhijeet Mallick, Deepak Kagate and Nikhil Vaidya(2014), A review on feature extraction techniques for CBIR ,pp:14-18.
- [2] Akshay Alex, Pranay Goyal, Tejaswinee Thorat, Mayur Sonawane and Subhash Rathod(2012), Content Based Image Retrieval Using Spatial Features , International Journal of Engineering Trends and Technology,pp:313-318.
- [3] Christoph Zauner(2010), Implementation and Benchmarking of Perceptual Image Hash Functions.
- [4] Harsh Kumar Sarohi and Farhat Ullah Khan(2012), Image Retrieval using Perceptual Hashing , IOSR Journal of Computer Engineering,pp:38-40.
- [5] Henning Muller, Nicolas Michoux, David Bandon and Antoine Geissbuhler(2007), A Review of Content- Based Image Retrieval Systems in Medical Applications – Clinical Benefits and Future Directions.
- [6] Issam,H., Laradji, Lahouari Ghouti and El-Hebri Khiari(2013), Perceptual hashing of color images using hypercomplex representations, International Conference on Image Processing,pp:4402-4406.
- [7] Liu,Y., Zhang,D., Lu,G., and Ma,W.,(2007), A survey of content based image retrieval with high-level semantics.
- [8] Mangijao Singh,S., and Hemachandran,K.,(2012), Content-Based Image Retrieval using Color Moment and Gabor Texture Feature, IJCSI International Journal of Computer Science Issues,pp:299-309.
- [9] Navneet Kaur, Meenakshi Sharma and Pankaj Sharma(2014), A content based image retrieval using statistical technique, International Journal of Computer Science Engineering and Information Technology Research, pp:199- 202.
- [10] Pavani,S., Shivani, Venkat Narayana Rao,T., Deva Shekar(2013), Similarity Analysis Of Images Using Content Based Image Retrieval System, International Journal Of Engineering And Computer Science,pp:251-258.
- [11] Ramesh Kumar,A., and Saravanan,D.,(2013) Content Based Image Retrieval Using Color Histogram, International Journal of Computer Science and Information Technologies,pp:242 - 245.
- [12] Suresh Kumar Nagarajan and Shanmugan Saravanan(2012), Content-Based Medical Image Annotation and Retrieval using Perceptual Hashing, IOSR Journal of Engineering, pp: 814-818.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)