# HPCIR: Histogram Positional Centroid for Image Retrieval

Sadan Kumar T N, *PG Student*
Department of Electronics and
Communication Engineering
University Visvesvaraya College of
Engineering
Bangalore, India

Dr A Sreenivasa Murthy,
*Associate Professor*
Department of Electronics and
Communication Engineering
University Visvesvaraya College of
Engineering
Bangalore, India

Rajani N, *Research Scholar*
Department of Electronics and
Communication Engineering
University Visvesvaraya College of
Engineering
Bangalore, India,
*rajaniap.25@gmail.com*

*Abstract*— **Content Based Image Retrival has wide application in the field of communication via internet. There exists various methods for retriving an image from the vast set of data present in the desktop, laptop, workstations and internet. The proposed Histogram positional centroid for image retrival technique uses histogrm feature extraction and position centroid. The Histogram Positional Centroid for Image Retrieval uses three feature vectors, namely histogram feature, positional centroid of each bin feature and average colour of each bin feature, for retrieval of images. Further, in the proposed method adding location and average colour description of each bin in colour histogram are incorporated. This results in speedy computation in retriving the image from the huge data base.**

*Keywords— Content Based Image Retrival; Histogram; Histogram Centroid; Location Indexing; Average Color Descriptor; Semantic Gap*

## I. INTRODUCTION

Content Based Image Retrieval is a system in which images are described by color, texture and shape. The metadata based image retrieval produces lot of irrelevant images and thus CBIR has a wide scope for image retrieval research. Hence the system in which retrieval is done using query by image content is required. The problem of finding digital images in a huge database of enormous advancement in internet accessibility of image capturing devices, digital scanners etc…As a collection of images are rapidly increasing the image database is also increasing. So there is a requirement for efficient and effective image search algorithm for retrieving, browsing and searching. The image retrieval tools are needed for the users in many domains like crime prevention, fashion designing, architecture, medicine, remote sensing, etc. Many general purpose image retrieval systems have developed for this purpose.

Content based image retrieval mainly focuses on the retrieval of images with desire characteristics from a collection of significant size of database. CBIR is a system based on image processing and computer vision where in images of specific characteristics are retrieved from database. Usually law enforcement agencies use online recognition systems to verify identity of a single person based on the subject database from a global database. The researchers try to solve the following issues in CBIR.

i) Efficient algorithms to bridge the semantic gap between human vision and computer vision

ii) Method to extract the features and storage such way that a feature retains enough information but requires less storage.

iii) A similarity measures that identifies the closeness between query and retrieved images either in the feature domain.

In CBIR to improve the retrieval accuracy the research interest is mainly start concentrating more on design of low level feature extraction algorithm to minimizing the semantic gap between richness of human semantics and visual features. Many techniques have been developed to improve the retrieval efficiency of the CBIR system. In this work histogram technique is extended for color image processing by adding its location and location centroid is taken for comparing the similarity between query image and database image. The proposed method reduces the computation complexity and also it speeds up the system.

The rest of the paper is organized as follows. Section II gives background work for CBIR also provides related work and problem definition is provided in section III. Section IV explain the algorithm and implementation. The results and performance analysis is briefly discussed in Section V. Finally, the conclusion is drawn at end of this paper.

### A. Motivation

The thrust for the recent development in technology of storage that provides a provision for large storage and extensive databases of digital images. These digital images can be million and occupying about terabytes of data. Thus, an efficient and effective method of retrieval need to be designed.

## II. BACKGROUND WORK

The process of image retrieval from the large database is Histogram centroid and histogram features. The general block diagram of the content based image retrieval system is as shown in figure1. The image features from images in database and the query image are extracted using any one of the above features Histogram centroid or histogram method. The features are stored initially at the database. Once the query image is fed as the input, the query image feature is found by using any one of the above stated method. The feature matching is done by finding the similarity measurement using Euclidean distance. The similar images are retrieved on the basis of distance measure for 'n' iterations from the image database that lead to the availability of related images called the retrieved images.
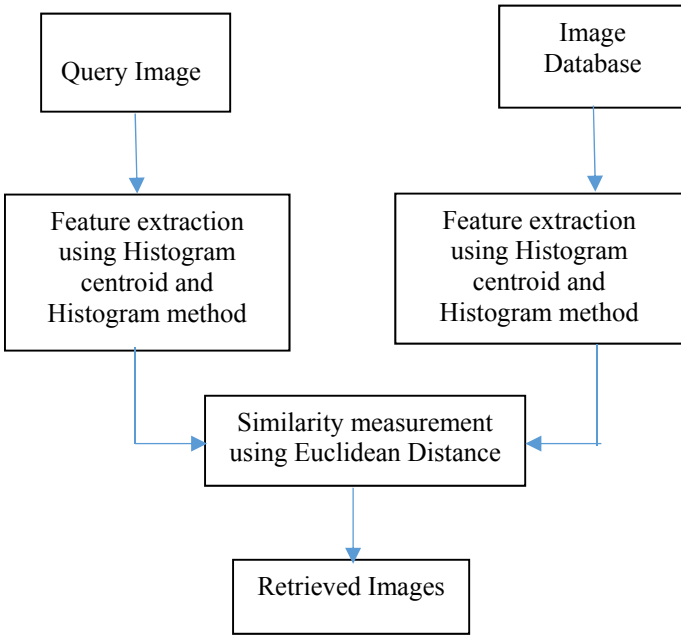


Fig 1: The General Block diagram of Content Based Image Retrieval System

### A. Related Work

P Aigrain et. al.,[1] have discussed various techniques available for CBIR which can be a good reference for beginner in the field. In S Aksoy et. al., [2] proposes graph theoretic approach for CBIR in which database search is modelled as graph clustering problem. This work uses two sets of feature: first set from line angle ratio statistics which is texture histogram that uses location of lines as well as their surroundings. The second set of feature is the variance of gray level spatial dependencies which are computed from co-occurrence matrix. Andrea Manno-Kovacs et. al. [3] proposes to use obvious feature points of the most identifiable image area through the use of salient orientation histogram. Anil K. Jain et. al. [4] focus on CBIR applications for TATTOO images in which features are extracted using scale invariant feature transform. Constraints based on the geometry are also introduced to reduce false retrieval. C.Patvardhan et. al[5] uses wavelet feature and edge histogram are used in CBIR for images having disturbances.

Jamil Ahmad, et al. [6] has attempted to capture the most influential features like color, edges and edge orientations with associated saliency values to compute a identifier to represent color images. In this paper proper weights are allocated for the features to measure the significance of image features.

### III. PROBLEM DEFINITION

The content based image retrieval should use simple and suitable best low level features that meet the requirement of a robust, efficient with computationally simple calculations. The main objective of this work is to retrieve the desired images from a huge image database and the development of an effective efficient algorithm with Histogram Positional Centroid. Let us consider the image extract Red, Green and Blue components and compute histogram for each component as represented in equation (1), then the average color of each bin and positional centroid of each bin is represented in equation (2) and (3) respectively.
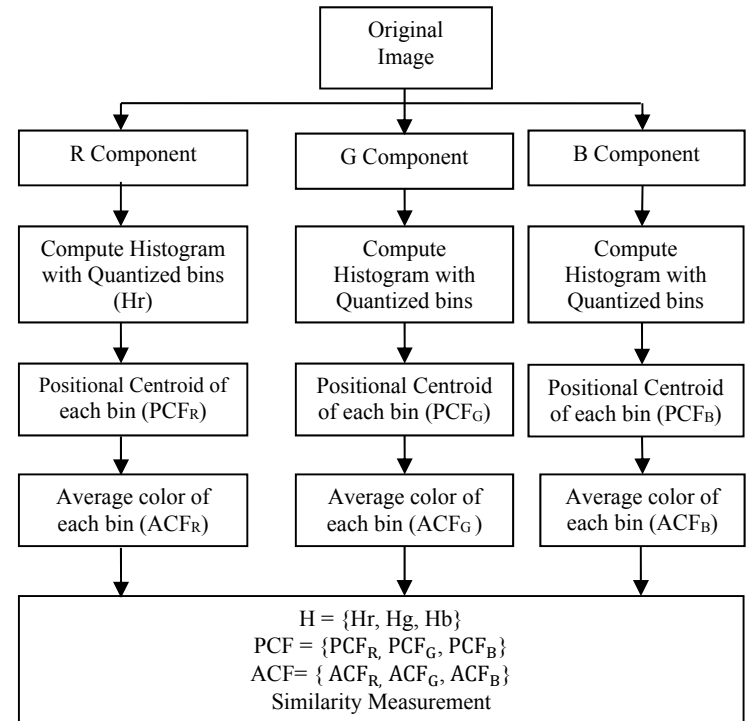


Fig 2: Block Diagram of Histogram Positional Centroid

$$H = \{H_R, H_G, H_B\} \qquad (1)$$

$$ACF = \{ACF_R, ACF_G, ACF_B\} \qquad (2)$$

$$PCF = \{PCF_R, PCF_G, PCF_B\} \qquad (3)$$

The detail description of the proposed method is as shown in Figure 2. The processing of these bins would lead in more computations and hence positional centroid of each bin within the quantized vectors is found. This lead to determine the average color bins with faster computation. The factor 'H'

defines the set of histograms given by the H, PCF and ACF components are used in the process of finding the retrieval image. The method of retrieval in this format would lead to determine the images at the data set that may lead in undefined images with the quantized format. The images outset would be blurred and would not convey the proper set images. Hence the quantized image need to be polished to convey the information more effectively. The method of polishing the image to convey the certain proper information would be done using similarity measurement. The distance is important metric and it plays a most important role in the CBIR system since the retrieval result is very sensitive with the chosen distance metric. The similarity measurement is the distance computation used to measure the similarity matching between two images [7]. The image matching between two images is done by calculating the distance between the query image and the images in the database based on their corresponding features. The distance metric used to find the similarity between the images is Euclidean distance. The feature vectors of the images with small distance value are considered as most similar to the query images.

## IV. ALGORITHM AND IMPLEMENTATION

Analysis and improvement in content based image retrieval is the need at the present internet era with lots of computational cost affecting the minds of the designer in designing the suitable algorithm for CBIR. The Table 1 shows the new proposed HPCIR algorithm that lead in few simple steps that decreases the computational cost with more precision. These steps only give the brief idea of computing the histogram and the set found should be very clear to state the status of the image.

Histogram is one of the basic method for knowing the image intensity levels and it gives computationally inexpensive image features. In this work image intensity levels are found for every twenty six bins and later the position of the intensity levels is taken and stored. These average of locations are chosen to get the positional centroid as image feature. Thus, the simplest form of histogram having fixed number of bins are defined. A bin corresponds to a fixed range of intensity values.

The range of intensity values defines the width of a bin. Generally, all bins are of equal width. The intensity (color) changes in the same image may shift the bin-membership of pixels altering the intensity distribution.

Table 1: HPCIR Algorithm

| | |
|---|---|
| Step 1: | Input the query Image |
| Step 2: | Determine the RGB components of the Image |
| Step 3: | Compute the histogram of each Quantized bins |
| Step 4: | Determine the positional centroids of the image |
| Step 5: | Average color of each bin is found |
| Step 6: | Compute the set found in step2,3,4, and 5 for all |
| Step 7: | Store the features. |
| Step 8: | Similarity measurement using Euclidean distance |
| Step 9: | Display the minimum distance image with index |

## V. PERFORMANCE EVALUATION

The performance evaluation of the CBIR system is done by using recall and precision. These two measurements are chosen from information retrieval namely, recall and precision. Recall measures the ability of the system to retrieve all the images that are relevant, while precision measures the ability of the system retrieve only the images that are relevant. They are defined as follows,

**Precision(P)** is the ratio of the relevant images to the total number of images retrieved.
$$P = n/n_1.$$
**Recall(R)** is the percentage of relevant images from all possible relevant images
$$R = n/n_2 .$$
Where,

   $n$ : Number of relevant images retrieved.
   $n_1$ : Total number of images retrieved.
   $n_2$ : Total number of relevant images in the database.

The proposed method has been implemented using MATLAB. This method is applied and tested for COREL database which contains 1000 images of 10 different classes each of 100 images. The test database contains 50 images of ten different classes like African people, Beach, Building, Bus, Dinosaurs, Elephant, Rose, Horse, Mountain and Food each of five images respectively.

Table 2. Precision and Recall using Histogram centroid and Histogram method

| Image Class | Using Histogram Centroid | | Using Histogram | |
|---|---|---|---|---|
| | Precision | Recall | Precision | Recall |
| Dinosaurs | 1 | 0.105 | 0.4 | 0.042 |
| Rose | 0.9 | 0.094 | 0.1 | 0.0105 |
| African people | 0.2 | 0.021 | 0.5 | 0.052 |
| Horse | 0.1 | 0.010 | 0.1 | 0.0105 |
| Average | 0.55 | 0.0575 | 0.275 | 0.02875 |

Table 2 above shows the performance details of HPCIR and histogram method for four image class they are Dinosaurs, Rose, African people and Horse. For the better performance, precision should be more and recall should be less.

From the table Histogram Centroid method gives 55% of precision and 5.75% of recall and Histogram method gives 27.5% of precision and 2.875% of recall. Figure 3 below shows the graph of precision and recall with its average of retrieving the image from the image set of 1000 using Histogram method. Figure 4 shows the precision and recall graph of the proposed HPCIR method. The HPCIR had improved with its Precision and Recall reaching to 100% for few test set like dinosaurs with lot improved with the average value in retrieval and also it shows the effectiveness of the proposed work.
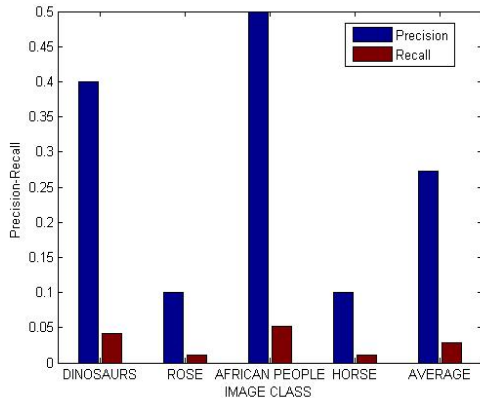
Fig 3: Precision and Recall for different image class using Histogram based Image Retrieval
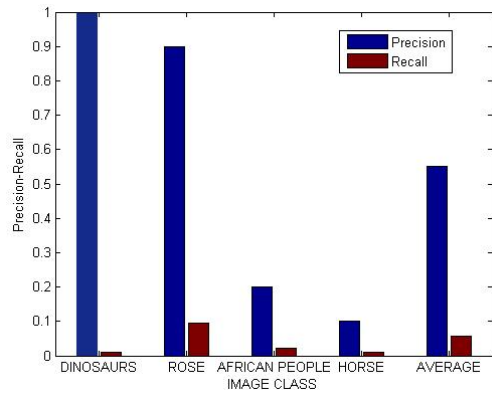


Fig 4: Precision and Recall for different image class using HPCIR

Few of the top 10 retrievals of the COREL database for the image sets using HPCIR are illustrated from Figure 6 to Figure 9. Query image of Dinosaurs, Rose, African people and Horse image class as shown in Figure5. For Dinosaurs image class ten Dinosaurs images are retrieved. Similarly, nine images from Rose class, four images from African people image class and nine images from Horse image class are retrieved. The retrieval results are good for Dinosaurs, Rose and Horse image classes.
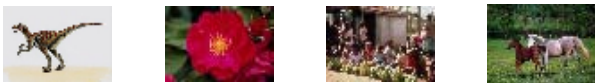


Fig 5.Query Image from Dinosaurs, Rose, African people and Horse Image Class.



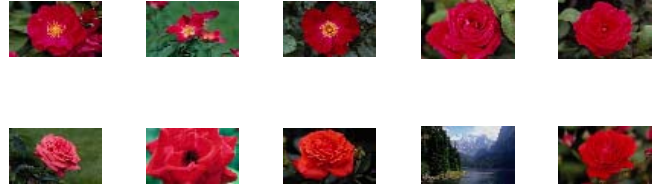Fig 6. Retrieved Images from Dinosaurs class Using HPCIR



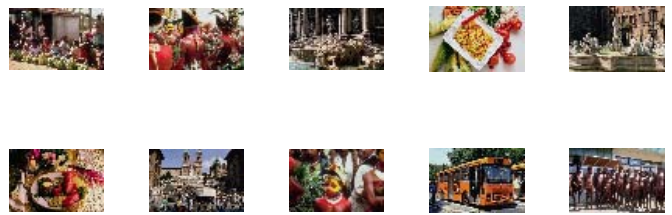Fig 7. Retrieved Images from Rose Class using HPCIR



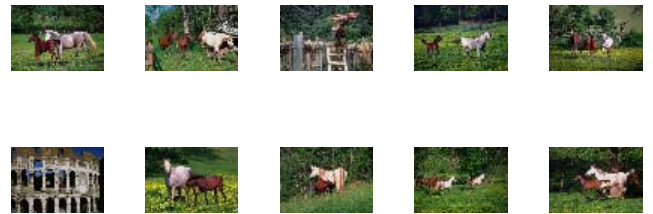Fig 8. Retrieved Images from African People class using HPCIR



Fig 9. Retrievd Images from Horse class using HPCIR

## VI. CONCLUSION

Histogram Positional Centroid for Image Retrieval would justify the speed in computations that result in vast application in crime prevention, internet applications, fashion, remote sensing, publishing, medicine and architecture etc. The proposed HPCIR method provides acceptable results compared with histogram method. The use of adding location descriptor and average colour makes the computation simpler and more effective in retrieval of images using less feature.

REFERENCES

[1] P Aigrain, H Zhang and D Petkovic, Content-Based Representation and Retrieval of Visual Media: A State of the Art Review, *Multimedia Tools and Applications,* vol. 3, pp. 179-202, 1996.

[2]   S Aksoy and R Haralick, Graph-Theoretic Clustering for Image Grouping and Retrieval, *IEEE Proceedings of Computer Vision and Pattern Recognition,* pp. 63-68, 1999.

[3]   A. Manno-Kovacs, "Content based image retrieval using salient orientation histograms," *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA,2016,   pp.2480-2484.doi:   10.1109/ICIP.   2016. 7532805 Measure, *Image Understanding,* vol. 75, nos. 1 2, pp. 46-58, 1999.

[4]   A. K. Jain, J. E. Lee, R. Jin and N. Gregg, "Content-based image retrieval: An application to tattoo images," *2009 16th IEEE International Conference on Image Processing (ICIP)*,Cairo,2009,pp.2745-2748.

[5]   C. Patvardhan, A. K. Verma and C. V. Lakshmi, "Robust content based image retrieval based on multi-resolution wavelet features and edge histogram," *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, Shimla, 2013, pp. 447-452.

[6]   J. Ahmad, M. Sajjad, I. Mehmood and S. W. Baik, "SSH: Salient Structures Histogram for Content Based Image Retrieval," *Network-Based Information Systems (NBiS), 2015 18th International Conference on*, Taipei, 2015, pp. 212-217.

[7]   A. Yoshitaka and T. Ichikawa, "A survey on content-based retrieval for multimedia databases," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 11, no. 1, pp. 81-93, Jan/Feb 1999.