

# Comparing CBIR for RGB Images Using Color Histogram Quantization with Euclidean and Bhattacharyya Distance

Asha Gowda Karegowda  
Associate Professor, Dept. Of MCA,  
Siddaganga Institute of Technology,  
Tumkur, Karnataka, India  
ashagsit@gmail.com

Shilpa Bhat  
Student, Dept. of MCA,  
Siddaganga Institute of Technology,  
Tumkur, Karnataka, India  
bhatshilpa29@gmail.com

Dhanalakshmi N  
Student, Dept. of MCA,  
Siddaganga Institute of Technology,  
Tumkur, Karnataka, India  
akshmi.kavya.n@gmail.co

**Abstract**— Content-based image retrieval (CBIR) is a technique that uses image attributes such as color, texture, shape to search for images which are similar to the input query image. Color histograms are widely used for the Content-Based Image Retrieval (CBIR). In this paper we compare image retrieval technique that uses color histogram for image indexing for RGB images. The color histogram of query image with those stored in the database is compared Euclidean distance and Bhattacharyya distance. In order to produce color histogram, color quantization is applied by varying the number of bins. Performance is measured using two measures namely accuracy and precision. The performance of CBIR using Bhattacharyya distance for color histogram showed improved results when compared to Euclidean distance.

**Keywords**— color-based image retrieval, Euclidean distance, Bhattacharyya distance, color histogram, quantization.

## I. INTRODUCTION

Classical CBIR Techniques are based on two types of visual features: Global and Local features. Image index by their visual content such as Color, Texture, Shape, Edge, Size, Spatial Layout etc [1] are retrieved by these low-level features are also termed as Global features, where as local features are based on key points or salient patches which involve domain knowledge on complex inference procedures. The color is a powerful descriptor since it facilitates the identification and extraction of objects from a scene and humans can detect thousands of shades and intensities of color, compared to about two dozen shades of gray [2,3]. The color histogram is a method for describing the color content of an image, it counts the number of occurrences of each color in an image. The paper is organized as follows. Section 2 provides the basic methodology adopted for retrieval of images based on color histogram using two distance measures: Euclidean distance and Bhattacharyya distance. The experimental results and the data are discussed in Section 3, followed by the conclusions in Section 4.

## II. COLOR HISTOGRAM

A color histogram is a representation of the distribution of colors in an image. There are several distance measures commonly used for finding the similarity between two color histogram. There are two types of color histograms, Global Color Histograms [GCHs] and Local Color Histograms

[LCHs]. GCHs represent one whole image with a single color histogram and an image will be encoded with its color histogram. The distance between two images will determine by the distance between their color histograms. While the LCHs divide an image into fixed blocks and takes the color histogram of each of those blocks [4,5]. While comparing two images distance is calculated using their histograms between a region in one image and a region in same location in the other image. The distance between the two images will be determined by the sum of all these distances. Thus when comparing GCHs one may get inconsistent result in terms of similarity of images when compare to LCH. Color quantization is a process that reduces the number of distinct colors used in an image, usually with the intention that the new image should be as visually similar as possible to the original image [6]. In this paper two famous histogram distance measures namely Euclidean distance and Bhattacharyya distance have been applied [7] for global histogram. Euclidean distance is used to measure the dissimilarity between the input query image and the dataset images. Less the dissimilarity between the query image and dataset image, those dataset images are relevant images. Bhattacharyya distance is used to compare the similarity between two color histogram, and returns the value between 0 and 1. The value nearer to 1 means the compared images are similar. Histogram-based image retrieval requires some form of quantization in view of the fact that the raw color images result in large dimensionality in the histogram representation. In this paper authors have experimented quantization by varying the number of bins from 5 to 50. The main drawback of histograms is that the representation is dependent of the color of the image being studied, and do not consider the other features like shape and texture.

## III. METHODOLOGY

The first method adopted for testing the performance of CBIR using color-histogram approach using histogram difference using Euclidean distance is described as follows.

- A. Read images in database and extract RGB format pixel information from images.

- B. Create bin ( quantization value  $n = 5$  to  $15$ ) normalized histograms using two methods for each of the RGB image for each image read from database .
- C. Read in a query image and extract RGB format pixel information and create histograms for the query image.
- D. Compute a Euclidean distance by comparing the query image histograms to that of each image in the database.
- E. Sort images in database in order of ascending Euclidean distance to query image and return the top  $k$  images as result.

The first method adopted for testing the performance of CBIR using color-histogram approach using histogram comparison using Bhattacharyya distance is described as follows:

- A. Read images in database and extract RGB format pixel information from images.
- B. Create bin ( quantization value  $n = 5$  to  $15$ ) normalized histograms using two methods for each of the RGB image for each image read from database .
- C. Read in a query image and extract RGB format pixel information and create histograms for the query image.
- D. Compare the similarity of query image histograms to that of each image in the database using Bhattacharyya distance.
- E. Sort images in database in order of descending Bhattacharyya distance to query image and return the top  $k$  images as result.

#### IV. EXPERIMENTAL RESULTS

The database used for the experiment contains a total of 44 images downloaded from Corel dataset. The original image database comprises 1000 images Divided into 10 classes. Authors have a total of 44 images with 3 classes query images namely Red bus, Horse and Beach each 11 in number have been included in addition other images of elephant and yellow, green and blue bus. The performance of CBIR with  $k = 10$  and  $n = 5$  for Beach query image with histogram distance computed using Euclidean and Bhattacharyya distance are shown in figure 1 and figure 2 respectively. Similarly the performance of CBIR with  $k = 10$  and  $n = 5$  for Horse and Red bus query images with histogram distance computed using Euclidean and Bhattacharyya distance are shown in figure 3-4 and figure 5-6 respectively. The value of  $k$  has been varied in the range of 5 to 10. The quantization value  $n$  has been varied in the range of 5 to 50. Precision and accuracy have been used as performance measures. Figure 11 and Figure 12 show's the comparative graph for precision by varying the quantization value of  $n$  and  $k$  value using Euclidean distance and Bhattacharyya distance respectively. In addition Figures 7-10 depict the comparisons accuracy of CBIR for different value of  $k$  and quantization values  $n$  using Euclidean distance and Bhattacharyya distance. Table 1 shows the best performance of CBIR for different values of  $k$  with best quantization value using Euclidean distance and Bhattacharyya distance. The results of Table 1 and figure 10 and figure 11 clearly have shown the enhanced performance of

proposed system using Bhattacharyya distance and quantization value when compared to Euclidean distance.

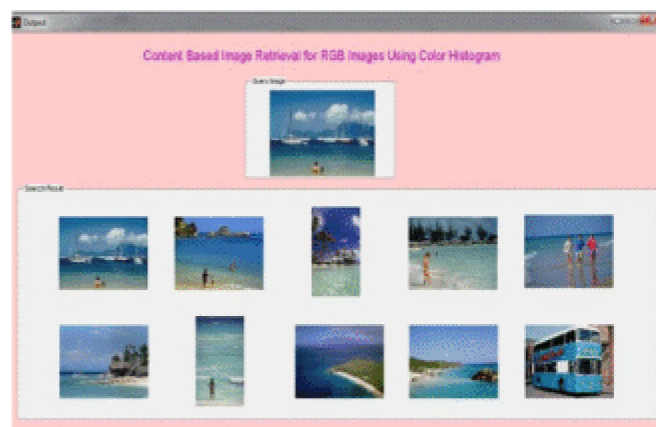


Fig1 Retrieval results for beach query image with  $k = 10$  and Bhattacharyya distance

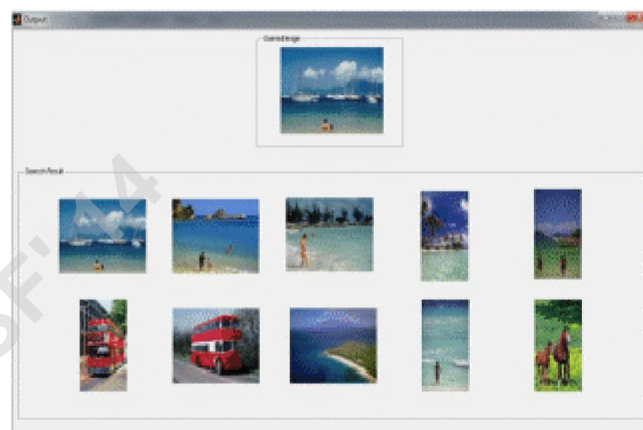


Fig 2 Retrieval results for beach query image with  $k = 10$  and Euclidian distance

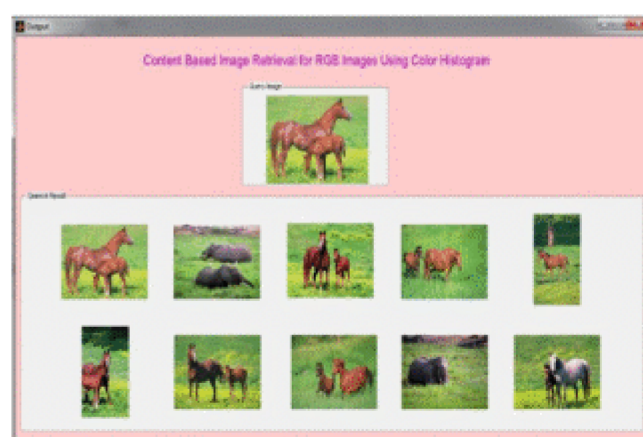


Fig3 Retrieval results for horse query image with  $k = 10$  and Bhattacharyya distance

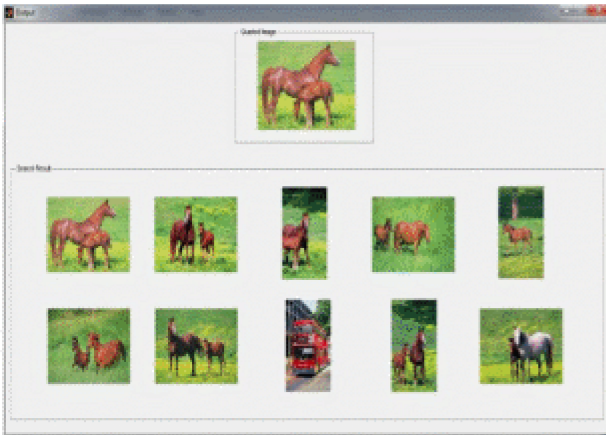


Fig 4 Retrieval results for horse query image with k = 10 and Euclidian distance

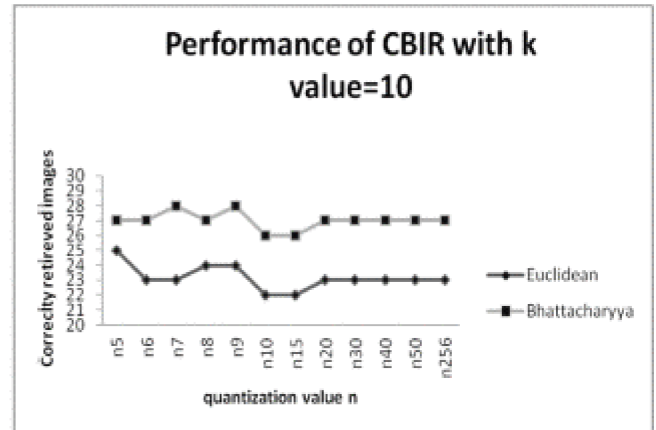


Fig 7 Comparing accuracy of RGB quantization with k = 10 for: (a) Euclidean distance, (b) Bhattacharyya distance



Fig 5 Retrieval results for red bus query image with k = 10 and Bhattacharyya distance

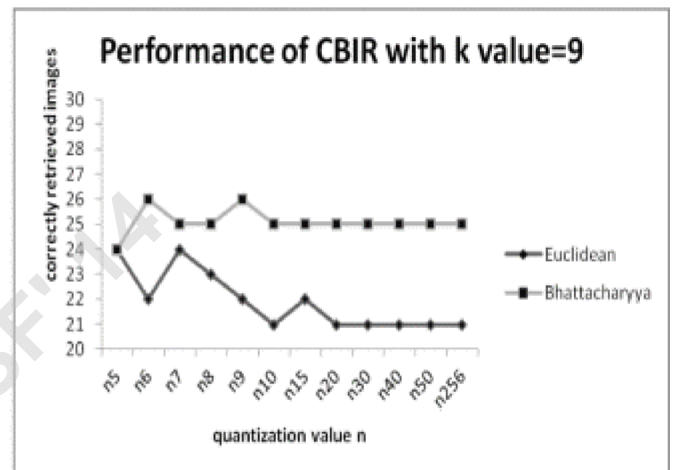


Fig 8 Comparing accuracy of RGB quantization with k = 9 (a) Euclidean distance, (b) Bhattacharyya distance

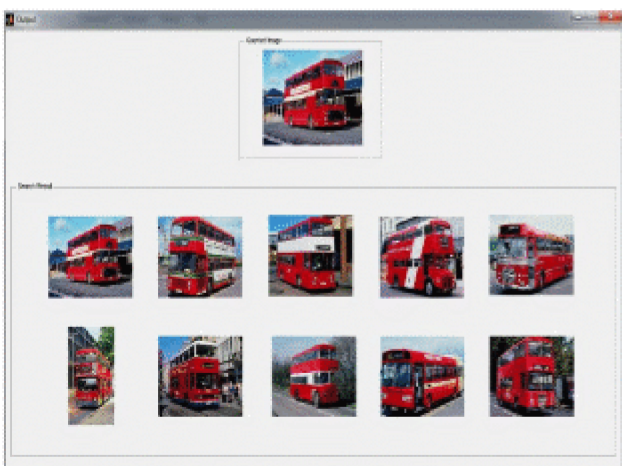


Fig 6 Retrieval results for Red bus query image with k = 10 and Euclidian distance

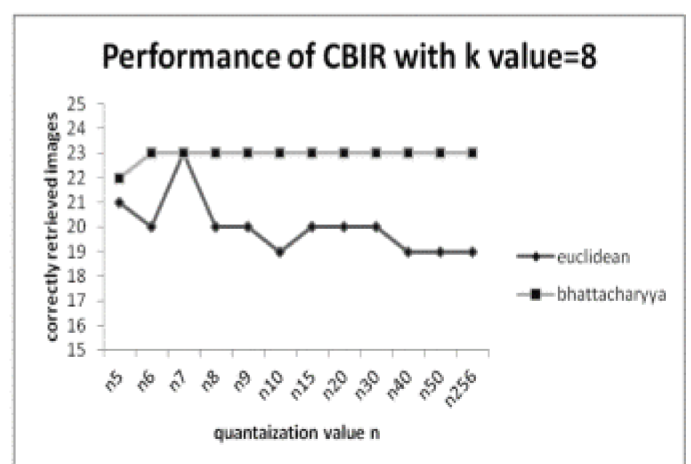


Fig 9 Comparing accuracy of RGB quantization with k=8: (a) Euclidean distance, (b) Bhattacharyya distance

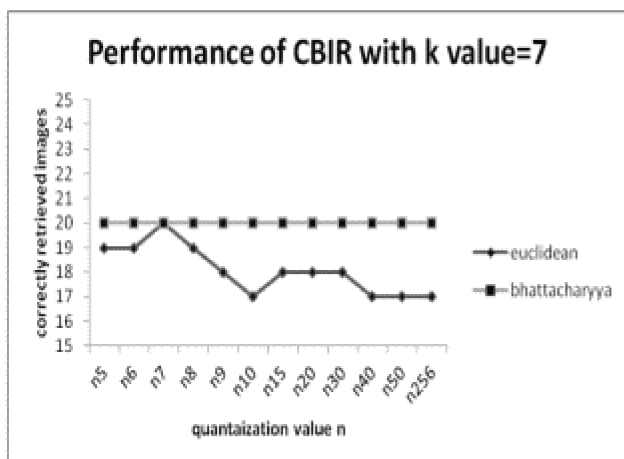


Fig 10 Comparing accuracy of RGB quantization with k = 7 (a) Euclidean distance, (b) Bhattacharyya distance

Table 1. Comparing of the accuracy according to the RGB quantization for different values of k with (a) Euclidean distance and (b) Bhattacharyya distance

value	Euclidean	Bhattacharyya	Euclidean	Bhattacharyya
10	5	7,9	83.3333	93.3333
9	5,7	6,9	88.8888	96.2962
8	5	6,7,8,9,10,15,20,30,40,50,256	87.5	95.8333
7	5	5,6,7,8,9,10,15,20,30,40,50,256	90.4676	95.2380

V. CONCLUSIONS

The color is a powerful descriptor for CBIR since it facilitates the identification and extraction of objects from a scene and humans can easily detect thousands of shades and intensities of color. This paper presents the use of color histogram using Euclidean distance and Bhattacharyya distance by varying the quantization value for RGC color space images. The performance of the proposed system proved to better with smaller number of bins and Bhattacharyya distance.

REFERENCES

- [1] Gudivada, V.N., Raghavan.V.V, "Content based Image Retrieval Systems", *EEE Computer*, Vol.28(9), pp.18-22, 1995.
- [2] Broek, E. L. van den (2005). *Human-Centered Content-Based Image Retrieval*. PhD-thesis Nijmegen Institute for Cognition and Information (NICI), Radboud University Nijmegen, The Netherlands – Nijmegen.
- [3] R. C. Gonzales and R. E. Woods. *Digital image processing*. Prentice-Hall, Inc., New Jersey, 2nd edition, 2002.
- [4] Shengjiu Wang, " A Robust CBIR Approach using Local Color Histograms", Department of computer Science, University of Alberta, Edmonton, Alberta, Canada, Oct 2001.
- [5] Asha Gowda Karegowda , Pushpa KR, Chitra M. "Color Histogram based Image Retrieval – A Survey ", *International Journal of Advanced Research in Computer Science*, Volume 4, Number 11, Nov-Dec 2013 .
- [6] Hafner, J., Sawhney, H.S., "Efficient color histogram indexing for quadratic form distance functions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7), pp. 729-736, 1995.].
- [7] Rahman M.M., Bhattacharya M.P., Desai B.C., "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback", *IEEE Trans. Inform. Technol. Biomed.*, Vol. 11(1), pp.58-69, 2007.

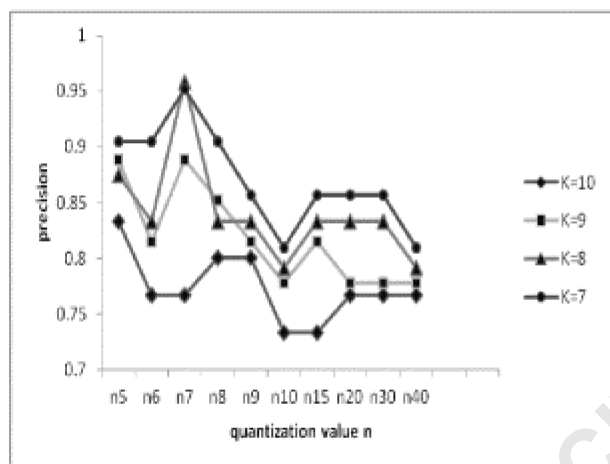


Fig 11. Calculation of the precision according to the RGB quantization for different values of k with Euclidean distance

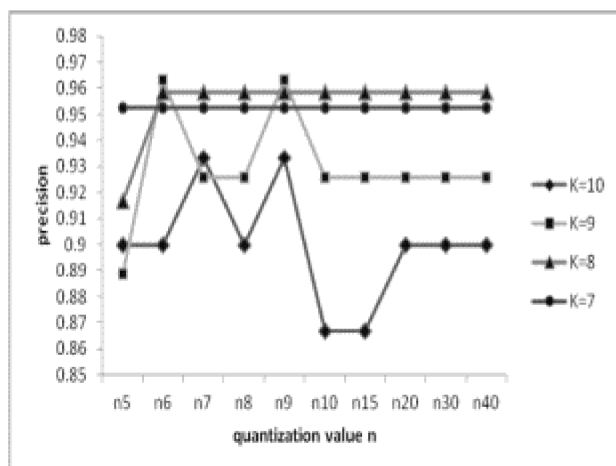


Fig12. Calculation of the precision according to the RGB quantization for different values of k with Bhattacharyya distance