# Survey on Performance of Color Histogram Using Statistical Features in CBIR

**Harkirat Kaur[1], Baljit Singh Khehra[2]**
[1, 2] Department of CSE
[1, 2] BBSBEC, Fatehgarh Sahib

***Abstract-*** *content based image retrieval (CBIR) is the process of searching similar images from an image database based on the visual contents of the input query image. The authors have presented a CBIR technique using color based feature. Since a color image, consists of three basic color components, i.e. red, green and blue, so in this work, we have given the same importance on all three color components during image retrieving process. In the presented CBIR technique, initially we have constructed three probability histograms for each color component and subsequently the histograms are divided into several numbers of significant bins and from each bin, we have computed several statistical values like standard deviation, skewness and kurtosis, with this a new factor of HMMD(hue min-max difference) is also computed. The computed statistical values are used as extracted features of the image data. The processing cost of the presented CBIR technique is significantly low. The technique has been tested on standard image databases and satisfactory results have been achieved.*

***Keywords-*** Content based image retrieval; probability histogram, standard daviation.

## I. INTRODUCTION

With the advancement in internet and multimedia technologies, a huge amount of multimedia data in the form of audio, video and images has been used in many fields like medical treatment ,Satellite data, video and still images repositories, digital forensics and surveillance system. This has created an ongoing demand of systems that can store and retrieve multimedia data in an effective way. Many multimedia information storage and retrieval systems have been developed till now for catering these demands. The most common retrieval systems are Text Based Image Retrieval (TBIR) systems, where the search is based on automatic or manual annotation of images. A conventional TBIR searches the database for the similar text surrounding the imageas given in the query string. The commonly used TBIR system is Google Images. The text based systems are fast as the string matching is computationally less time consuming process. However, it is sometimes difficult to express the whole visual content of images in words and TBIR may end up in producing irrelevant results. In addition annotation of images is not always correct and consumes a lot of time. For finding the alternative way of searching and overcoming the limitations imposed by TBIR systems more intuitive and user friendly content based image retrieval systems (CBIR) were developed. A CBIR system uses visual contents of the images described in the form of low level features like color, texture, shape and spatial locations to represent the images in the databases. The system retrieves similar images when an example image or sketch is presented as input to the system.

## II. PREVIOUS WORK

CBIR performs retrieval process based on the visual contents of the image data. The straightforward image to image searching mechanism is not considered in CBIR. Such approach is not the practically feasible to implement it in any real time applications because image data are comparatively huge in size. So in CBIR, we require suitable feature extraction techniques from the image data so that meaningful and relevant images can be retrieved based on those extracted image features. A number of CBIR techniques were developed based on considering the significant feature like color [6], texture [7] and shape [8-10]. Color feature is widely used in CBIR techniques since it is one of the most prominent low level features and it is also invariant to rotation, scaling, and other spatial transformations on the images. In general the histogram matching based CBIR techniques is relatively simple and faster. Swain et al. [11] method was based on color histogram and similarity measure was done using histogram intersection distance metric between the histograms of images. Another histogram based CBIR was proposed by F. Malik et.al. [12] where, they converted the color image into grayscale and the obtained grayscale was further preprocessed by Laplacian filter. The histogram of the processed image was used in CBIR. Wang et al. [13] proposed a method for image retrieval by concatenation of color features like most dominant colors of region, texture features and shape features that are based on pseudo Zernike moments. G. H. Liu et al. [14] used color difference histogram in CBIR. H. B. Kekre et al. [15] also devised CBIR techniques using color histograms and some statistical parameters. Murala et a1. [16] proposed CBIR is based on the features like color histogram and Gabor wavelet transform. In [17], the authors have decomposed image into 16 non overlapping color blocks and from each color blocks, they have computed various statistical values as a component of a feature vector. In [18], the authors have suggested a CBIR based on color moment and Gabor texture feature. M. Singha et al. [19] proposed two methods where in first method purely based on color histogram in spatial domain while the second method is based on wavelet domain. Ahistogram was constructed from the color image after color quantization process. In CBIR, the feature vector was computed from the histogram of the color image. In this paper,we have presented CBIR techniques using color histogram.The straight forward histogram comparison is not suitable in any real time scenario. So we have computed several statistical values from the histogram. The retrieving was done based on the extracted statistical values. The scheme is suitable and easy to implement in any real time applications.

## III. RETRIEVAL TECHNIQUE

In CBIR, indexing and retrieval of images is based on low level visual features of images such as color, texture and shape. Basically CBIR is performed in two steps: indexing and searching. In indexing, the low level features are extracted from image database and stored in the form of feature vector to create a new database which is known as feature database. In searching, initially the same feature extraction algorithm is employed in the query image to extract the feature vector and subsequently the sim ilarity is measured with the feature database using some distance measure like Euclidean distance.The best result is obtained those having minimum similarity distance.Performance of any CBIR techniques depends on the extraction of suitable features of the images. CBIR will give efficient and effective result if we consider more features of the images, but as a result overall computational cost will be high. So from a practical point of view, the dimension of the feature vector of the image should be selected appropriately so that the performance may not be affected impressively. Therefore, in CBIR, selection of the suitable dimension of the feature vector is challenging task. In this paper some statistical features of the images have been measured. In some the cases, the statistical features have produced significant results in the classification of the datasets. Some commonly used statistical parameters [20] are mean, standard deviation, skewness and kurtosis etc. These statistical parameters can be computed directly from an image histogram. Initially, the original histogram is converted into a form of a normalized histogram where X-axis denotes the intensity level, $ri$ and Y-axis denotes the estimated probability p($r_i$) of level $r_i$ The mean ($\pi$) for the range of intensity value [LB, UB] is computed as:

$$\mu = \sum_{i=LB}^{UB} r_i \, p(r_i) \qquad \dots (1)$$

The standard deviations (σ), skewness ( γ ), and kurtosis (k) for the range of intensity value [LB, UB] are obtained as follows:

$$\sigma = \sqrt{\sum_{i=LB}^{UB} (r_i - \mu)^2 \, p(r_i)} \qquad \dots (2)$$

$$\gamma = \frac{1}{\sigma^3} \sum_{i=LB}^{UB} (r_i - \mu)^3 \, p(r_i) \qquad \dots (3)$$

$$k = \frac{1}{\sigma^2} \sum_{i=UB}^{UB} (r_i - \mu)^2 \, p(r_i) \qquad \dots (4)$$

The standard deviation shows the contrast of the image in the particular significant bin of the histogram of the image. It is used to measure the distribution of the intensity values about mean in each bin of the histogram. If the value of the standard deviation in a particular block of the image is less the value of the other block in the image, it represents that high contrast in that particular block of the image than the other. The skewness is the measure of the skewed intensity values of the image in each block of the image about mean of that block. If the value of the skewness is negative in the particular region of the image, then it represents that most of intensity values lie on the right side of the mean than the left side and tail of the intensity values longer and lie on the left side of the mean of that region of the image. If the value is positive, then most of the intensity values lie in the left side of the mean in a region and tail of the intensity values more skewed in the right side of the mean. If skewness is zero then it represents that distribution of the intensity values about mean is equal. The kurtosis is used to calculate the peak of the distribution of the intensity values of the image about mean in a significant bin of the histogram. If Kurtosis has high value, then it represents the sharp peak distribution of the intensity values about mean and has longer and fat tail and kurtosis with low value represents flat distribution of intensity values with short and thin tail.Take the image input colour RGB and decompose into its three colour components, i. e. Red(R), Green (G) and Blue B) respectively where each colour component has L intensity levels. For each colour component, compute the probability histogram as follows

$$p(r) \; \left\{\frac{number\ of\ pixel\ in\ r}{width * height}\right\} V[0, L-1] \quad \dots (5)$$

Where p(r) represents the relative frequency or probability of r-th intensity value and the range of intensity value is [0, L -1] .Compute for each colour component, the standard deviation, skewness and kurtosis from $Bin_cV$ [0, n -1] where 'C' represents R, G and B colour components respectively. Construct feature vectors for the query image and the database images using the feature extraction technique. Compute the distance between feature vector of the query image and the feature vectors of the database images for the similarity measurement. Sort the distance in non-decreasing order and select top N images having minim distances. For HMMD (Hue Min-Max Difference),The RGB colour model algorithm used in the present work is not sufficient for better colour description. Therefore the HMMD (Hue Min Max Difference) colour space, is can be used in order to efficiently describe the colours in an image. The hue has the same meaning as in the HSV space, and max and min are the maximum and minimum among the R, G, and B values, respectively. The diff component is defined as the difference between max and min. Only three of the four components are sufficient to describe the HMMD space. This colour space can be depicted using the double cone structure as shown in the figure. In the CBIR core experiments for image retrieval, it is observed that the HMMD colour space is very effective and compared favourably with the HSV colour space. Note that the HMMD colour space is a slight twist on the HSI colour space, where the diff component is scaled by the intensity value.

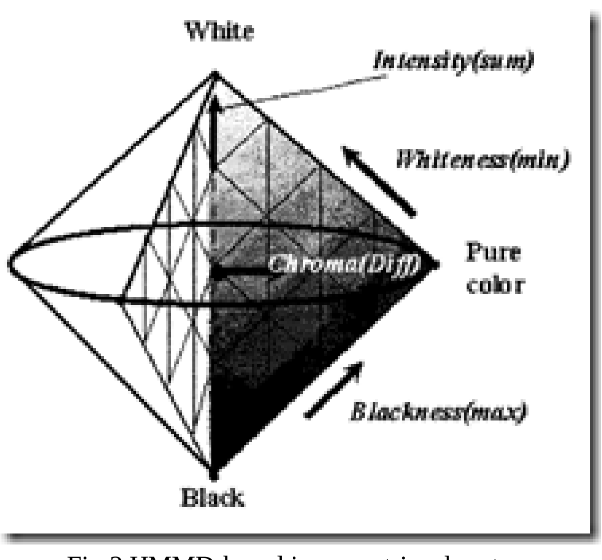


Fig.2.HMMD-based image retrieval system

## V. CONCLUSION

The proposed method is simple and effective during the retrieval process. It is based on the color histogram which is partitioned into several numbers of non-uniform bins that contain most significant information of the image. Features can be extracted from each significant bins of the histogram for retrieving of the relevant color images from the database .The statistical parameters are calculated directly from the image histogram as a result, it reduces the processing cost.The presented method gives the 100% precision value for horse and dinosaur categories images where for Building and mountain images gives the worst result due to the presence of more details features. However, the performance of this method for other images is acceptable and it is suitable for any real time applications.

## REFFERENCES

[1] M. Kokare, B. N. Chatterji, and P. K. Biswas. "A survey on current content based image retrieval methods. " IETE Journal of Research,vol. 48 no. 3-4, pp. 261-271, 2002.

[2] L. Ying, D. Zhang, Lu. Guojun, Ma. Wei-Ying. "A survey of content based image retrieval with high-level semantics," Pattern Recognition,vol. 40, no. 1, pp. 262-282,2007.

[3] A. W. M. Smeulders, M. Worring, S. Santini and A. Gupta, and R. Jain,"Content-based image retrieval at the end of the early years," Pattern Analysis and Machine Intelligence, IEEE Transactions , vol. 22, no. 12, pp. 1349-1380,2000

[4] A. Vailaya, A. T. Figueiredo, A. K. Jain, and H. 1. Zhang, "Image classification for content-based indexing," Image Processing, IEEE Transactions, vol. 10, no. I, pp. 117-130, 200 I.

[5] Tong S. and Chang E., "Support vector machine active learning for image retrieval," In Proceedings of the ACM International Conference on Multimedia, Ottawa, Canada, pp. 107-118,2001.

[6] N. Shrivastava and V. Tyagi, "An efficient technique for retrieval of color images in large databases," Computers and Electrical Engineering, 2014.

[7] S. D. cheng, X. U. Lan, and L. Y. Han, "Image retrieval using both color and texture features," The Journal of China universities of posts and telecommunications, vol. 14, pp. 94-99,2007.

[8] R. Krishnamoorthy and S. S. Devi, "Image retrieval using edge based shape similarity with multiresolution enhanced orthogonal polynomials model," Digital Signal Processing, vo1.23, no. 2, pp. 555-568, 2013.

[9] N. Alajlana, I.E., Rube, M.S. Kamel and G. Freeman, "Shape retrieval using triangle-area representation and dynamic space warping," Pattern Recognition, vol. 40, no. 7, pp. 1911-1920,2007.

[10] M. A. Z. Chahooki and N. M. Charkari, "Shape retrieval based on manifold learning by fusion of dissimilarity measures," Image Processing, lET, vol. 6, no. 4, pp. 327-336,2012.

[11] M. J. Swain and D.H. Ballard, "Color indexing," International Journal of Computer Vision, vol. 7, no. 1, pp. 11-32, 1991.

[12] Malik F., and Baharum B., "Feature analysis of quantized histogram color features for content-based image retrieval based on laplacian filter," International Conference on System Engineering and Modeling, Singapore, 2012.

[13] X. Y. Wang, Y. 1. Yong, and H. Y. Yang, "An effective image retrieval scheme using color, texture and shape features," Computer Standards & Interfaces, vol. 33, no. 1, pp. 59-68,2011.

[14] G. H. Liu, and 1. Y. Yang, "Content-based image retrieval using color difference histogram," Pattern Recognition, vol. 46, no. I, pp. 188-198, 2013.

[15] Kekre H. B. and Sonawane, K., "Use of equalized histogram CG on statistical parameters in bins approach for CBIR," In Advances in Technology and Engineering (lCATE), International Conference, IEEE,pp. 1-6,2013

[16] Murala S., Gonde A. B. and Maheshwari, R. P., "Color and texture features for image indexing and retrieval," In Advance Computing Conference(IACC), IEEE , pp. 1411-1416,2009.

[17] M. Imran, R. Hashim, and N. E. A. Khalid, "Color histogram and first order statistics for content based image retrieval," In Recent Advances on Soft Computing and Data Mining, Springer International Publishing,pp. 153-162,2014.

[18] S. M. Singh and K. Hemachandran, "Content-based image retrieval using color moment and gabor texture feature," International Journal of Computer Science Issues(IJCSI), vol. 9, no. 1, pp. 1694-0814,2012.

[19] Manimala S. and Hemachandran, K., "Content based image retrieval using color and texture," Signal & Image Processing, an International Journal (SIPIJ), vol. 3, no. I, pp. 39-57, 2012.

[20] R. C. Gonzalez, R. E. Woods and S.L. Eddins, "Digital Image Processing," 2nd Edition Pearson Prentice Hall Book, 2004.

[21] Li. Jia, and 1. Z. Wang, "Automatic Linguistic Indexing of pictures by a statistical modeling approach," Pattern Analysis and Machine Intelligence, IEEE Transactions, vol. 25, no. 9, pp. 1075-1088, 2003