International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-2 Issue-1, October 2012

Content based Image Retrieval using Color Space Approaches

Gaurav Jaswal, Amit Kaul, Rajan Parmar

Abstract— ***The decreasing costs of consumer electronic devices such as digital cameras and digital camcorders, along with the ease of transportation facilitated by the Internet, has lead to a phenomenal rise in the amount of multimedia data. With this rapid development of multimedia technologies, the problem of how to retrieve a specified image from large amount of image databases becomes an important issue. In this paper we have developed a CBIR system based on the color features in RGB and HSV color space. Global color histogram (GCH) which lacks spatial information about the image colors has been compared with LCH. Algorithms were tested on two Databases one comprising of 500 JPEG images and another comprising of 120 JPEG images of national flags of different countries. The LCH approach has been found to be better and more accurate than GCH approach.***

Index Terms—GCH, LCH, RGB & HSV.

I. INTRODUCTION

Human vision is the most advanced of all our senses and as such we gather majority of information from the real world by visual sense. The twentieth century has witnessed unparalleled growth in the number, availability and importance of images in all walks of life. As the diversity and size of digital image collections have grown exponentially, efficient image retrieval is becoming increasingly important. Large image databases are difficult to browse with traditional text searches because the task of user based annotation become very time consuming, as the text often fails to convey the rich structure of images. A content-based retrieval system solves this problem where retrieval is based on the automating matching of feature of query image with that of image database through some image-image similarity evaluation [1]. Therefore images will be indexed according to their own visual content such as color, texture, shape or any other feature or a combination of set of visual features. In this paper, our research is limited to color feature of the image. There are several techniques available for the color feature extraction like color histogram, color layout, clustering and color correlogram etc [2]. The aim of the paper is to study various color based approaches and implement color histogram approach at global and local level.

Manuscript published on 30 October 2012.
***** Correspondence Author (s)
Gaurav Jaswal*, ECE, Carrer Point University, Hamirpur, H.P, India
Amit Kaul, EED, National Institute of Technology, Hamirpur, India
Rajan Parmar, ECS, HPPCL, Shimla, H.P, India

II. LITERATURE REVIEW

Currently the most popular search engines for images have resulted in the development of algorithms to augment and replace tag based image retrieval with content based image retrieval. These algorithms compare the actual content of the images rather than text which has been annotated previously by a human being. Once the specified feature has been extracted from the image, there are also a number of options for carrying out the actual comparison between images [3,4]. Generally similarity between two images is based on a computation involving the Euclidean distance or histogram intersection between the respective extracted features of two images. The three most common characteristics upon which images are compared in content based image retrieval algorithms are color, shape and texture [5]. Utilizing shape information for automated image comparisons requires algorithms that perform some form of edge detection or image segmentation. However, the performance of the algorithm is not invariant on scale or translation manipulations of images. Information regarding the texture of images can be even harder to extract automatically during retrieval. Generally algorithms rely on the comparison of adjacent pixels to determine the contrast or similarity between pixels [6].

A. Color based Image Retrieval

The color feature is one of the most widely used visual features in image retrieval. It is relatively robust to background complication and independent of image size and orientation. In image retrieval, the color histogram is the most commonly used color feature representation. Statistically, it denotes the joint probability of the intensities of the three color channels. This paper attempts to explore and analyze such an algorithm that compares images based on their color content. Swain and Ballard proposed histogram intersection, an L1 metric, as the similarity measure for the color histogram [7]. To take into account the similarities between similar but not identical colors, Ioka and Niblack et al. introduced an L2-related metric in comparing the histograms. Furthermore, considering that most color histograms are very sparse and thus sensitive to noise, Stricker and Orengo proposed using the cumulated color histogram [8]. Their research results demonstrated the advantages of the proposed approach over the conventional color histogram approach. Besides the color histogram, several other color feature representations have been applied in image retrieval, including color moments and color sets. To overcome the quantization effects, as in the color histogram, Stricker and Orengo proposed the color moments approach [9].

The mathematical foundation of this approach is that any color distribution can be characterized by its moments. Furthermore, since most of the information is concentrated on the low-order moments, only first moment (mean), second and third central moments (variance and skewness) were extracted as the color feature representation. Weighted Euclidean distance was used to calculate the color similarity. To facilitate fast search over large-scale image collections, Smith and Chang proposed color sets as an approximation to the color histogram [10,11]. They first transformed the (RGB) color space into a perceptually uniform space, such as HSV, and then quantized the transformed color space into M bins. A color set is defined as a selection of colors from the quantized color space. Because color set feature vectors were binary, a binary search tree was constructed to allow a fast search [12].

B. Current content based image retrieval systems

There are several commercial image retrieval systems are available in market. IBM'S Query by Image Content (QBIC) was the first commercial CBIR system and probably the best known system [13,14]. It supports user to retrieve images by color, shape, and texture. It supports query based on example images, user constructed sketches and selected color and texture patterns. Virage is a content based image retrieval search engine developed at Virage Incorporation [15]. It performs image retrieval on images and videos on the basis of color, shape, texture, and color layout. It goes one step ahead than IBM's QBIC because it supports arbitrary combination of the above queries. Photo book developed at Massachusetts Institute of Technology (MIT). It allows user to retrieve images by color, shape, texture feature [15,16]. It provides a set of matching algorithms, including Euclidean, vector space angle, histogram and wavelet tree distances as distance metric. Visual SE K and Web SEEK, both were developed at Columbia University [17]. Visual SEEK allow user to form queries by specifying color regions and spatial locations. Web SEEK is an image catalog and search tool for the web. Retrieval Ware was developed by Excalibur Technology Corp[17,18]. It uses color, shape, texture, brightness, color layout and aspect ratio of image and combination of these as query feature. Netra is a prototype image retrieval system developed in the UCSB Alexandria Digital Library project [18]. The main contribution of Netra system are its Gabor filter based texture analysis, neural net based image construction. MARS is an inter disciplinary research effort involving multiple research communities: computer vision, DBMS, and information retrieval [19,20]. Blob-world developed at UC-Barkeley. Synapse is an implementation of retrieval by appearance using whole image matching [21].

III. ALGORITHM AND METHODOLOGY

The outline of the method in our testing of color-histogram approach to content based image retrieval is as follows:

1. Read images in database and extract RGB format pixel information from images.
2. Create normalized histograms for each of the RGB components of each image read from database. Thus, each image will have 3 histograms associated with it.
3. Read in a query image and extract RGB format pixel information
4. Create histograms for each of the RGB components of the query image.
5. Compute a Euclidean distance by comparing the query image histograms to that of each image in the database.
6. Sort images in database in order of ascending Euclidean distance to query image and return as.
7. Repeat the first six steps same for HSV Color model too.

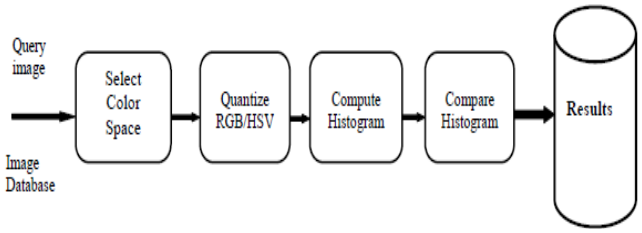

Fig.1Block Diagram of the Methodology

A. Color Histogram

Color Histogram is one of the widely used technique for color feature extraction in color based image retrieval. Color Histogram is a method for describing the color content of image, constructed by counting the number of pixel of each color. There are two traditional techniques for color-based image retrieval: Global Color Histograms (GCH) that represent images with single histograms; and Local Color Histograms (LCH) that divide images into fixed blocks and, for each block, obtain its color histogram. Global Color Histograms do not capture the content of images adequately, whereas Local Color Histograms contain more information and also enable the color distances among regions between images to be compared [39].

B. Color Space

Colors are defined on a selected color space such as RGB, LAB, LUV, HSV (HSL) and the hue-min-max-difference (HMMD). In order to use a good color space for a specific application, color conversion is needed between color spaces. In this paper, first of all we used RGB color space to represent color information but this is perceptually uniform in color based image retrieval. To improve perceptual uniformity we preferred HSV color space over RGB color space and compute both of the histograms. HSV are defined mathematically by transformations between the r, g, and b color coordinates in the RGB model and the h, s, and v coordinates in the HSV models.

C. Distance Measure

There are several distance formulas for measuring the similarity of color histograms. In this paper we used Histogram Euclidean distance to compare the histograms. Let h and g represent two color histograms. The Euclidean distance between the color histograms h and g can be computed as:

$$d^2(h,g) = \sum_A \sum_B \sum_C \left[h(a,b,c) - g(a,b,c) \right]^2$$

IV. IMAGE COLLECTION AND EXPERIMENT SET-UP

The retrieval performance is examined using a database of 500 JPEG images clicked by 3 mega pixel camera and for 120 JPEG images downloaded from the internet. Our examinations were performed on a "Pentium (R) Dual CPU T2310 @ 1.2 GHZ" based PC under the Vista operating system.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The programs implementing the feature extraction and comparison were written in MATLAB (version 7.6.0.324). In this thesis the color feature was extracted using GCH and then the same approach was compared with the LCH method with the use of RGB as well as HSV color space models. Our results showed that the "LCH" method performed better than "GCH" method. Our query set consisted of different query images; each query had a subjective answer set with 6 images. Finally, a Graphical User Interface (GUI) has been built, which took a query image as an input and retrieves the best six matching images from the database and displays them.

V. Results and Discussion

A. The GCH Approach (Database 1)

First of all, we used RGB color space. We took a query image and this method retrieved 3 best images from the database 1. It is shown below in the Fig. 1.

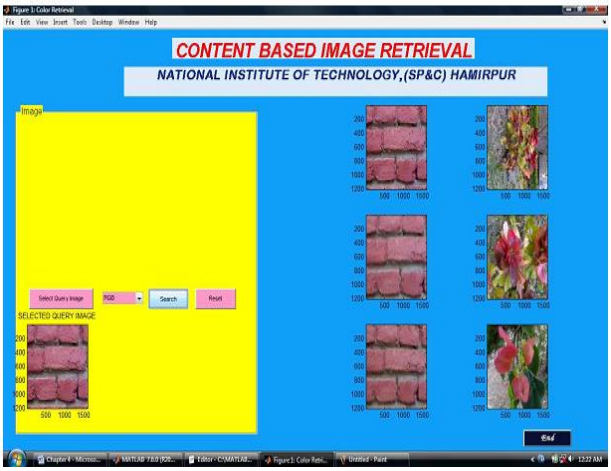

Fig.2 CBIR using GCH (RGB color space)

In Fig. 2, the approach GCH was used with HSV color space. In this experiment, the same query image was taken and the system retrieved 4 best images from the database1

.

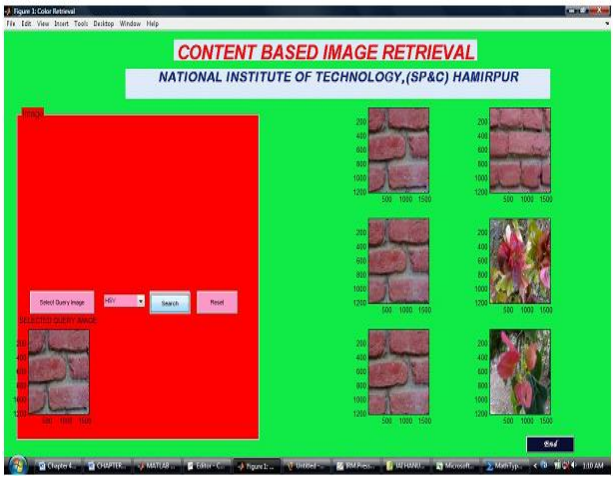

Fig.3 CBIR using GCH (HSV color space)

B. The LCH Approach (Database 1)

In this method the same query image was taken as we used in GCH and it retrieved the six best images from the database1 because it contains the spatial information of the images. This is shown in Fig. 3. It has found that LCH retrieves the maximum number of 6 images.

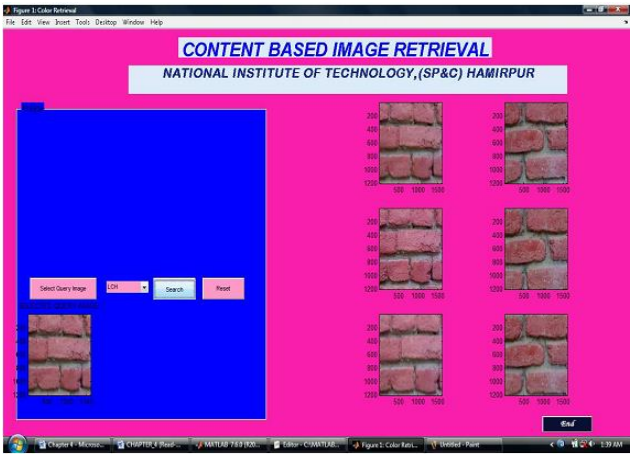

Fig.4 CBIR using LCH

Similarly, in the case of LCH method, GUI displayed four best images from the database 2 which showed that the results were found to be better than the previous approach. It is shown above in the Figure4.

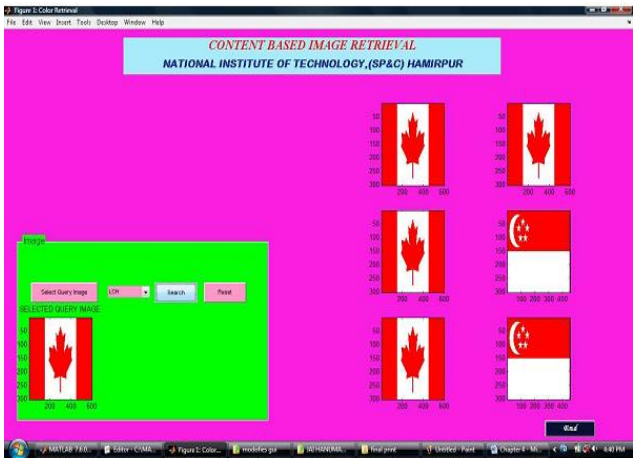

Fig.5 CBIR using LCH

The database 2 contained different images of countries flags. This type of CBIR system can be used for the purpose of home entertainment.

C. Retrieval Accuracy

The usage of Color Histogram algorithm requires the evaluation of matching performance. Precision and Recall graphs are a standard way to evaluate retrieval results for information retrieval systems. Let R be the set of relevant images in the database, A the answer set of retrieved images, and R (A) the set of relevant images in set A. Thus we can define two quantities.

(1) Recall is the fraction of relevant images in the database that have been retrieved in response to a query.

$$\text{recall} = \frac{|R(A)|}{|R|}$$

(2) Precision is the fraction of the retrieved images that are relevant to the query image.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$\text{precision} = \frac{|R(A)|}{|A|}$$

D. Comparison of two databases

The used algorithm is examined for two databases using two approaches and performance comparison is tabulated below.

Table1Database Comparisons

Algorithm Used	Precision	Recall
DATABASE 1		
GCH (RGB)	36%	31%
GCH (HSV)	53%	47%
LCH	88%	79%
DATABASE 2		
GCH (RGB)	20%	17%
GCH (HSV)	40%	36%
LCH	54%	49%

The average precision and average recall are computed by grouping the number of retrieved images sorted according to classification of database images with the query image.

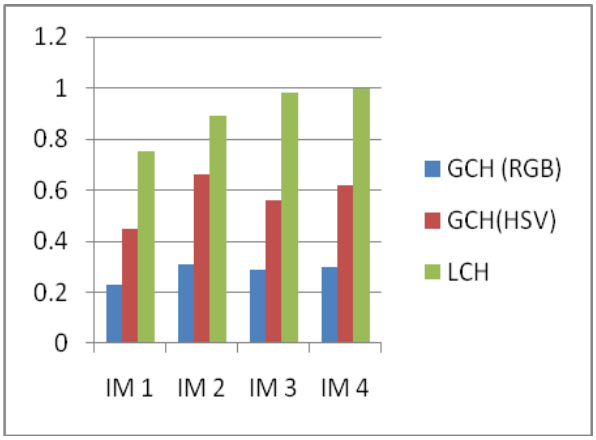

Fig. 5 Precision & Recall plotted against number of retrieve images for different image sets (Database 1)

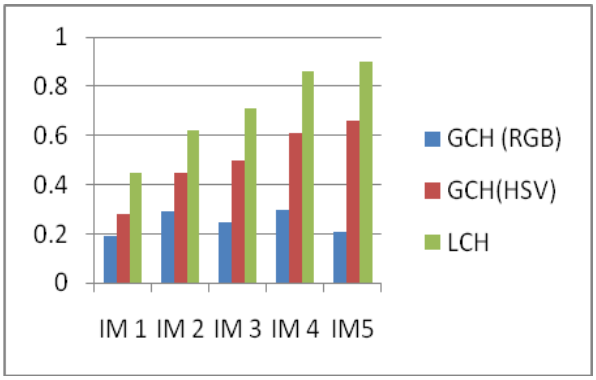

Fig. 6 Precision & Recall plotted against number of retrieve images for different image sets (Database 2)

VI. CONCLUSION

This paper was motivated by the desire to overcome the shortcomings of color-based image retrieval approaches. In this piece of work, we have implemented color space approaches to extract the color feature from the images. The retrieval accuracy is found to be 88% using LCH. When the same algorithm is applied on GCH methods, the accuracy came out to be 53% and 36% respectively. Thus we conclude that, LCH gives better results as comparison to GCH because GCH does not include spatial information.

VII. FUTURE WORK

CBIR is a vast research area and has many open questions and challenges. The ultimate aim is to reduce the gap between semantic information in the image and the extracted low-level features. In this paper, we provided an overview of content based image retrieval, investigated some techniques for color based image retrieval, and demonstrated the shortcomings of the GCH over LCH. The results obtained in this work can be improved further by using better features extractors such as Cumulative Color Histogram and Color Coorelogram.

REFERENCES

[1] J. R. Smith and S. F. Chang,“ Single Color Extraction and Image Query,” in Proc. IEEE International Conference on Image Processing, vol. 3, pp. 23-41, 1997.
[2] T. S. Chua, K. L. Tan, and B. C. Ooi, “Fast Signature based Color Spatial Image Retrieval,” in Proc. IEEE International Conference on Multimedia Computing and Systems, vol.1, pp. 362-369,1997.
[3] D. Hoilm, R. Sukhtankar, H. Schneiderman, and L.Huston, “Object based Image Retrieval using Statistical Structure of Images,” in Proc. CVPR, vol. 16, pp. 345-362, 2004.
[4] Y. Rui and T. S. Huang, “Image Retrieval: Current Techniques, Promising Directions, and Open Issues,” Journal of Visual Communication and Image Representation , vol.10, pp. 39-62, 1999.
[5] A. Jain and A. Vailaya, “Image Retrieval using Color and Shape,” Elsevier Science Ltd, vol. 29, pp. 1233- 1244, 1996
[6] R. Dutta, D. Joshi, J. Lee and J. Nang, “Image retrieval: Ideas, Influences and trends of New Age,” ACM Computing Surveys, vol.15, pp. 34-94, 2008.
[7] M. J. Swain and D. H. Ballard,“Color Indexing,” International Journal of Computer Vision, vol. 7, pp. 11-32, 1991.
[8] M. Stricker,and M. Orengo, “Similarity of Color Images,” SPIE, vol. 2420, pp. 381-392, 1995.
[9] S. Deb, and Y. Zhang, “An Overview of Content-based Image Retrieval Techniques”, in Proc. International Conference on Advanced Information Networking and Application, vol. 1, pp. 59-64, 2004.
[10] J.R. Smith and S. F. Chang,“A Fully Automated Content based Image Query Systems,” in Proc. ACM Multimedia, vol.2670, pp.87-98, 1996.
[11] W. Rasheed , G. Kang, J. Kang, J. Chun and J. Park, “Sum of Values of Local Histograms for Image Retrieval,” in Proc. International Conference on Networked Computing and Advanced Information Management, vol. 2, pp. 690- 694, 2008.
[12] B. Luo, X. Wang, and X. Tang, “A World Wide Web Based Image Search Engine using Text and Image Content Features,” in Proc. IS&T/SPIE Electronic Imaging, SPIE, vol. 5018, pp.142-148, 2003.
[13] R.Achanta, F. Estrada, P. Wils, and S. Susstrunk, “Salient Region Detection and Segmentation,” International Conference on Computer Vision Systems, vol. 5008, pp. 66-75, 2008.
[14] C. Carson, S. Belongie, H. Greenspan and J. Malik, “Image Segmentation using Expectation Maximization and its Application to Image Querying,” IEEE Transaction on Pattern Analysis and Mach. Intelligence, vol.24, pp. 1026-1038, 2002.
[15] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker, “Query by Image and Video Content: the QBIC System, ” Computer, vol. 28, pp. 23-32, 1995.
[16] J. R Smith and S. F. Chang, “Tools and Techniques for Image Color Retrieval,” in Proc. IEEE International. Conference Acoust, Speech and Signal Proc., vol.2670, pp. 3-25, 1996.
[17] J. Moustakes, K. Morius, and C. Stelios, “A Two Level CBIR Platform with Application to Brain MRI Retrieval,” IEEE International Conference on Multimedia and Technology, vol.24, pp. 818-856, 2005.
[18] M. Banergee, M. K. Kundu and P.K. Das, “Image Retrieval with Visually Prominent Features using Fuzzy Set Theoretic Evaluation,” ICVGIP India, vol.1611, pp. 67-86, 2004.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

[19] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain, "Content based Image Retrieval at the End of the Early Years," in Proc IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 1349-1380, 2000.
[20] S. F. Chang, A. Eleftheriadis and R. Mcclintock, "Next Generation Content Representation, Creation and Searching for New Media and Application in Education," in Proc. IEEE, vol.86, pp. 15-31, 1998.
[21] J. Sun Ding, J. CUI, "Spatial Feature Extraction and Image Retrieval based on Entropy," Systems Engineering and Electronics, vol. 28, pp. 791-794, 2006.