Asian Journal of Electrical Sciences
ISSN: 2249-6297 (P) Vol.3 No.2, 2014, pp.38-45

DOI: https://doi.org/10.51983/ajes-2014.3.2.1922

An Efficient Method for Color-Based Image Retrieval System

S.Selvam[1] and S.Thabasukannan[2]
[1]Research Scholar, Research and Development Center, Bharathiar University,
Coimbatore – 641 046, Tamil Nadu, India & Assistant Professor, Department of Computer Applications,
N.M.S.S.Vellaichamy Nadar College, Madurai, Tamil Nadu, India
[2]Principal, Pannai College of Engg. &Technology, Sivagangai – 630 561, Tamil Nadu, India
E-Mail: s.selvammscmphil@gmail.com,

Abstract - **Content-based image retrieval systems retrieve images from a database that are determined to be similar to a query image based only on features extracted from the images. This paper focuses on color-based image retrieval. We define methods to improve the efficiency and effectiveness of color-based retrieval. We have tested our system using a collection of color images and query images. Color histograms are used to extract and store the color content of the images. Our empirical results are very encouraging. The main aim of this paper is to reduce substantially the total color space without degrading retrieval performance. In addition, we are able to improve performance by conducting object retrieval based solely on color.**

I.INTRODUCTION

Image Retrieval is an extension of traditional information retrieval to include images [1]. Content-Based Image Retrieval (CBIR) is the process of retrieving images from a database on the basis of features that are extracted automatically from the images themselves [2]. In CBIR, a query is an image or portion of an image; relevant images are retrieved based on the similarity of the features of the query and the features of the individual images in the database. Possible features include color, quality, shape, positioning, or a combination thereof. This paper focuses on the use of a single image feature for retrieval purposes, namely color.

Color is a visual feature that one immediately perceives when looking at an image [1]. Thus, one can oftentimes readily identify and distinguish images based solely on their color content. Images that appear similar to the human eye are typically very similar in their color content. Hence, color would appear to be a good choice as a computerized feature by which to identify, compare, and retrieve images automatically.

However, although so-called *color-based image retrieval* has the advantage that it is easy to identify and extract the color content of any image, the representation of color is storage intensive and retrieval based on color is computationally intensive. Color representation is a high dimensional feature, with each image representing a combination of approximately 16 million colors. In addition, calculating the distance between images based on color is computationally complex. In this paper, we explore methods to improve the efficiency of color-based image retrieval without adversely affecting retrieval performance. We experiment with various methods to reduce the storage required of the color content and to improve retrieval efficiency.

II.IMAGE RETRIEVAL

A.Text-Based Retrieval

Images were stored along with string attributes – keywords prepared by an annotator that reflected in a relatively broad manner the content of the image. A user would search the database by specifying a Boolean combination of keywords describing the characteristics of images to be retrieved. The system would search the database looking for images whose string attributes matched the image query; images similar in *content* to the query would be retrieved.

B.Content-Based Retrieval

CBIR utilizes representations of features that are automatically extracted from the images themselves. Most of the current CBIR systems allow for querying-by-example, a technique wherein an image or part of an image is selected by the user as the query. The system extracts the features of the query image, searches through the database for images with similar features, and displays relevant images to the user in order of similarity to the query.

C.Color-Based Retrieval

It provides multiple measurements at a single pixel of the image, enabling classification to be done without the need for complex spatial decision-making. Color content is also independent of view and resolution and is easy to extract from an image and to manipulate.

III.COLOR-BASED IMAGE RETRIEVAL

Here we provide the background necessary for understanding how color can be used to support image representation and retrieval.

A.Color Representation

Here there are two models viz0020hardware-oriented or user-oriented. *Hardware-oriented models* are based on how optical devices such as color monitors and printers reproduce colors. *User-oriented models* correspond to the human perception of colors based on wavelength, amount of white light, and brightness.

B.RGB Color Representation

RGB is the most widely used representation of color in hardware. It is based on the physiology of the human eye, which recognizes colors based on the three color cones (Red, Green and Blue) present in the retina of the eye. The RGB space is represented as a solid cube with each of the three axes representing one of the colors, red, green, and blue. The colors black and white and the so-called *maximally saturated colors* red, green, blue, yellow, magenta, and cyan, represent the corners of the cube. The gray scale is from black to white – along diagonally opposite corners.

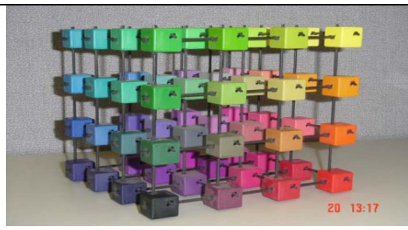


Fig. 1 RGB color cube

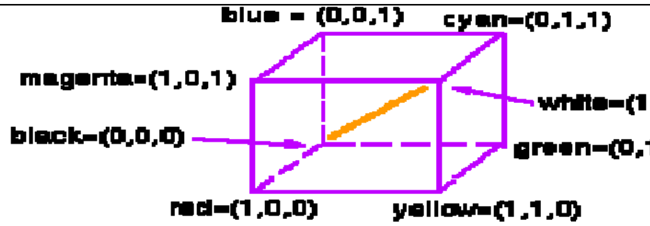


Fig. 2 Digitized RGB color cube

Every pixel is represented as a mixture of the three primaries red, green, and blue. The color to be displayed for each pixel is encoded using three bytes, one for each color with a scale ranging from 0-255. This model can be used to display $(2^8)^3$ distinct colors, that is, approximately 16 million colors.

C.Color Histograms

The color histogram of an image is obtained by counting the number of pixels in the image that corresponds to each discrete color. Thus, a color histogram is a vector whose length is equivalent to the number of distinct colors in the model and whose value for a given vector component is the number of pixels in the image as a whole containing the color represented by the component.

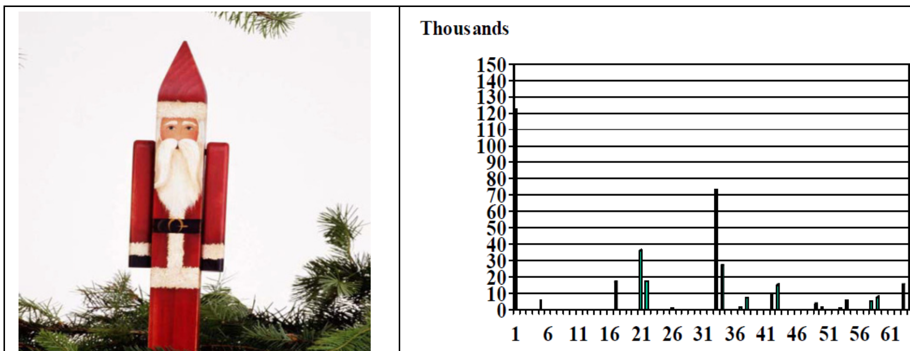


Fig. 3 An image and its 64-bin histogram

In this paper, we use histograms as the primary means of representing the color properties of images. We attempt to refine the generation of histograms, to minimize the size of histograms without reducing retrieval performance, and to reduce the computational intensity of comparing histograms. Our primary objective is to improve the efficiency of image retrieval without affecting retrieval performance.

IV.EXPERIMENTAL RETRIEVAL SYSTEM

Our system consists of a test set of color images, sample image queries, a known set of relevant images, methods for generating and displaying color histograms, and search mechanisms. To conduct retrieval experiments, a query image is chosen and the system returns all images in ranked order of perceived relevance to the query. We chose as our query image a Santa, an image that is dominantly red, white, and black. Our retrieval system uses a color histogram to represent the color content of an image. To determine the distance between two images, we elected to use the *histogram intersection* similarity measure. The intersection between the query image histogram *H(IQ)* and the histograms of the images in the database ***H(ID)*** is defined as

$$D_H(I_Q, I_D) = \frac{\sum_{j=1}^{n} min(H(I_Q, j), H(I_D, j))}{\sum_{j=1}^{n} (H(I_D, j)}.$$

From the equation above, it is obvious that colors not present in the query image do not contribute to the distance measure obtained by this formula. Thus, the greater the similarity value the greater the similarity of the two images.

Here we divide the RGB color cube into 64 smaller sub-cubes. This is accomplished by dividing each axis of the RGB cube described earlier into just 4 units, instead of 256 units. Each of the resulting 64 color sub-cubes is represented by the color whose pixel is at the center of the sub-cube. A picture is transformed into a 64-color representation by replacing the color of each pixel in the image with the color of the sub-cube in which the pixel's color resides. This image transformation allows us to represent images with color histograms containing 64 color bins.

In our experimental retrieval system, we represent all images in the database with such histograms. Histogram intersection is a simple and effective measure of the distance between histograms, and it has been shown that its performance in assessing the relevance of images based on color content is good for histograms of length 64.

Original Image *Transformed Image*

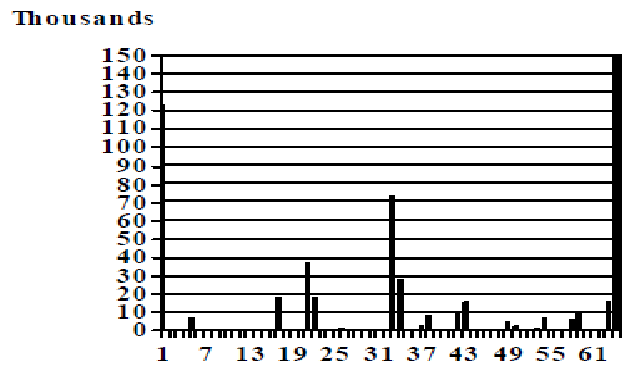
Fig. 4 A color image and its transformation to 64 colors

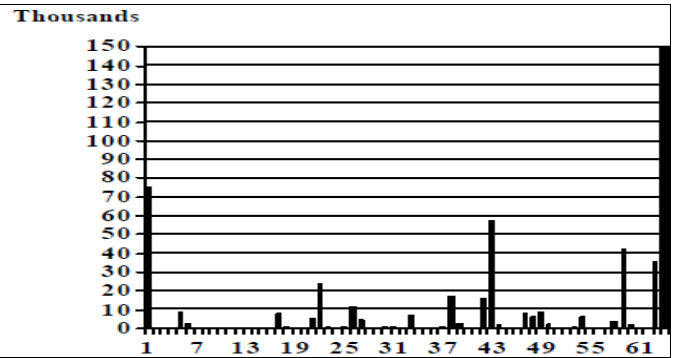
To conduct a retrieval experiment, our system compares a histogram representation of the Santa image of Figure 4 with the histograms of each image in the database. The resulting images are displayed in the order of perceived relevance with the query.

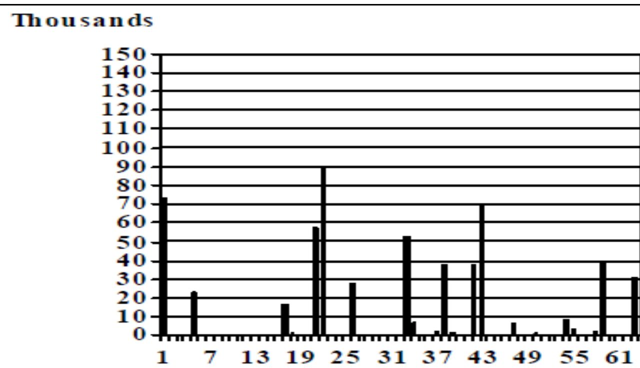
To generate a base case for retrieval, the case against which our experiments will be compared, we conducted a retrieval experiment using the 64-bin color histograms. The top 5 items returned by this search are displayed in Figure 5. Included in the results are the images themselves, the color histograms, and the histogram intersection similarity values. The query image is included as the first image in the list. These are considerably good results considering the similarity in color among all images in the database. Our experiments will assess the degree to which we can further reduce the length of the histograms or modify the search process to improve efficiency of the search process without sacrificing performance.

TABLE 1 RESULTS FOR A SAMPLE RETRIEVAL WITH 64-BIN

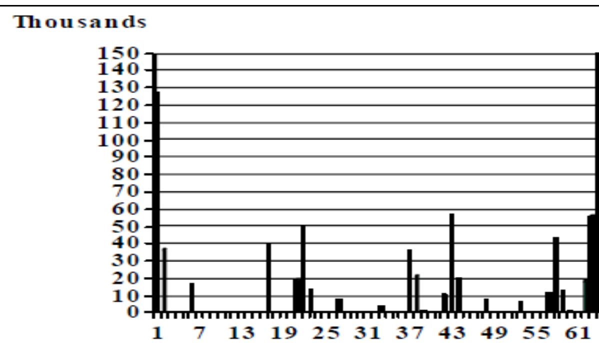
No. of Images	Relevant	Non-Relevant	Recall	Precision
5	3	2	20.0%	60.0%
10	5	5	33.3%	50.0%
15	6	9	40.0%	40.0%
20	7	13	46.7%	35.0%







0.72835



0.67550

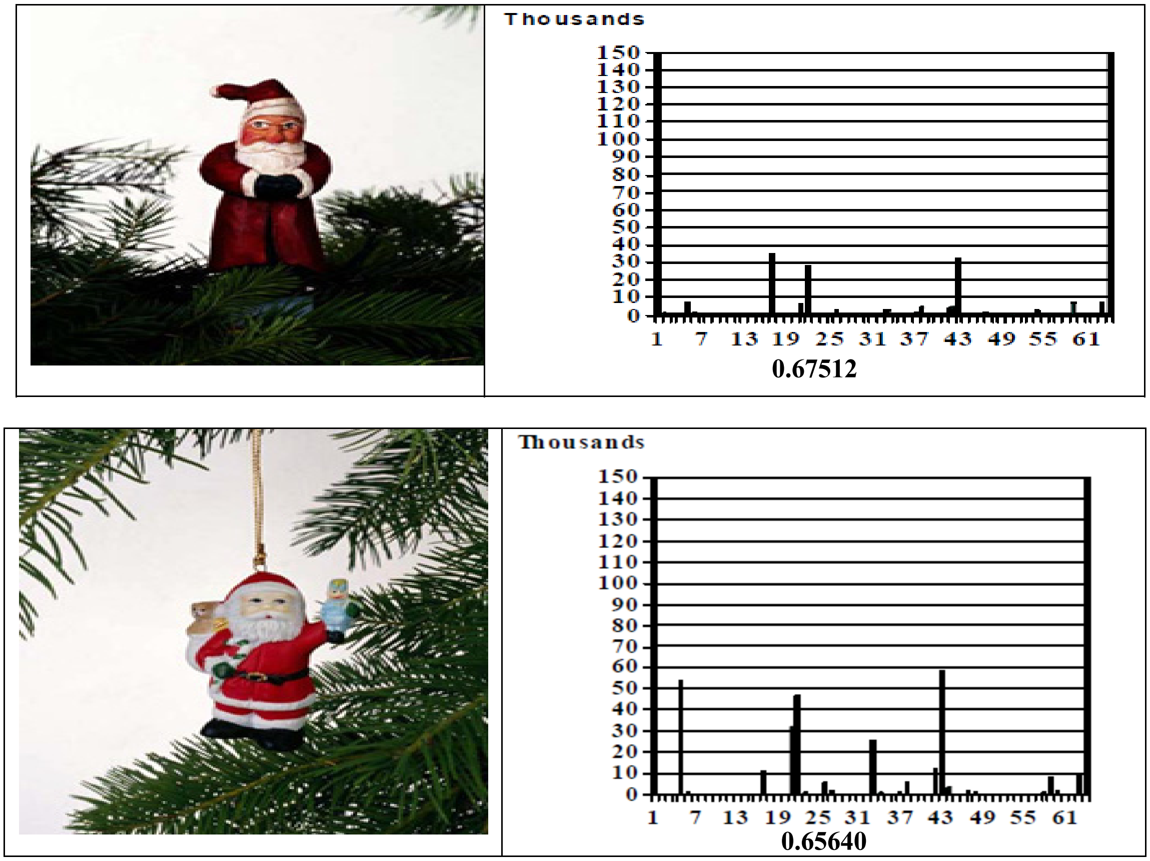


Fig. 5 Sample retrieval using 64 colors

V.EMPIRICAL STUDY

A.Retaining Highly Significant Colors

In figure 5 relatively few distinct colors actually contribute to an image's color content and hence the uniqueness of an image. Most colors in the histogram bins do not play a significant role in the comparison of the image with the query. Therefore, it would appear that such colors could be removed from the image representations without affecting retrieval effectiveness.

To test this paper, we analyzed the query vector and identified the dominant colors in its histogram. We identified the top *n* colors in the query, using values of *n* equal to 20, 15, 10, and 5. Figures 6 through 9 illustrate a comparison of a 64-color Santa image with each n-color image. The resulting images based only on the dominant colors are still very identifiable and lend support to our paper that only a few colors affect image retrieval.


Image with all colors *Image with 20 colors*

Fig. 6 Original image and its 20-color representation

Image with all colors *Image with 15 colors*

Fig. 7 Original image and its 15-color representation

Image with all colors *Image with 10 colors*

Fig. 8 Original image and its 10-color representation

Image with all colors *Image with 5 colors*

Fig. 9 Original image and its 5-color representation

We conducted four retrievals for the four values of n. In each retrieval experiment, we reduced the 64-bin query histogram and the histograms of each image to a length n by removing the color bins that represented the non-dominant colors. We then conducted retrieval experiment applying the histogram intersection similarity measure to the reduced histograms of length n. The precision/recall results are shown in Table 2. This validated our proposal that only a few dominant colors in the images contribute to its uniqueness and thus aided in comparison and retrieval. In fact, reduction of the histogram by a factor of 68% to 92% did not degrade retrieval performance.

This experiment confirms that very few colors in an image dominate and contribute toward the content of an image in a way that would be responsible for the recognition and uniqueness of the images. We can exploit this fact and reduce the computational cost of comparing the histograms and hence improve the efficiency of a search without adversely affecting retrieval performance.

TABLE 2 RESULTS FOR RETRIEVAL WITH DOMINANT N COLORS OF THE QUERY

RETRIEVAL USING 64 BIN AND TOP 20 COLORS

No. of Images	Relevant	Non-Relevant	Recall	Precision
5	1	4	6.7%	20.0%
10	5	5	33.3%	50.0%
15	5	10	33.3%	33.3%
20	7	13	46.7%	35.0%

RETRIEVAL USING 64 BIN AND TOP 15 COLORS

No. of Images	Relevant	Non-Relevant	Recall	Precision
5	1	4	6.7%	20.0%
10	5	5	33.3%	50.0%
15	5	10	33.3%	33.3%
20	7	13	46.7%	35.0%

RETRIEVAL USING 64 BIN AND TOP 10 COLORS

No. of Images	Relevant	Non-Relevant	Recall	Precision
5	1	4	6.7%	20.0%
10	4	6	26.7%	40.0%
15	5	10	33.3%	33.3%
20	7	13	46.7%	35.0%

RETRIEVAL USING 64 BIN AND TOP 5 COLORS

No. of Images	Relevant	Non-Relevant	Recall	Precision
5	1	4	6.7%	20.0%
10	5	5	33.3%	50.0%
15	7	8	46.7%	46.7%
20	8	12	53.3%	40.0%

B.Removing Non-Substantive Colors

To identify non-substantive colors, we generated the *centroid vector* of the histograms of all the images in the database. The centroid vector is a single histogram representing the entire collection; that is, a histogram generated by averaging the histograms of all the images in the collection. The color with the highest weight in the centroid vector, the one color that dominated all the images, was chosen for elimination. We performed retrieval experiments using the 64-bin color histograms and the reduced histograms. For these retrievals, we eliminated the non-substantive color when calculating the intersection between histograms. Results of these experiments conducted are shown in Table 3.

Retrieval performance improved for the 64-bin color representations and was maintained for the 20-bin color representation. Slight degradation in performance occurred as the number of colors was further reduced. These experiments support our paper that the dominant color did not contribute to retrieval effectiveness for larger histograms and could be eliminated from the decision process.

TABLE 3 RESULTS AFTER REMOVING THE DOMINANT NON-SUBSTANTIVE COLOR

Experiments	Relevant Images in top 20	Non-Relevant Images in top 20	Recall	Precision
Retrieval after e1iminating top color	8	12	53.3%	40.0%
Retrieval based on top 20 colors after eliminating top color	7	13	46.7%	35.0%
Retrieval based on top 10 colors after eliminating top color	5	15	33.3%	25%
Retrieval based on top 5 colors after eliminating top color	6	14	40.0%	30.0%

C.Object Retrieval

The retrieval process did not consider the color content of a particular object of interest within the query. It would appear that by focusing on the color content of a given object retrieval performance could be improved. To test this paper, we designed an experiment in which the query represented a sub image contained within a larger image in the database. We created a histogram for this sub image and compared it with rectangles of similar size and position within each of the other images in the database as a whole. The results are shown in Table 4. As may be noted, performance increased significantly.

TABLE 4 RESULTS FOR OBJECT RETRIEVAL

Experiments	**Relevant Images in top 20**	**Non-Relevant images in top 20**	**Recall**	**Precision**
Retrieval with query sub image compared to same sub-image of images in database	9	11	60.0%	45.0%

VI.CONCLUSIONS AND FUTURE WORK

The objective of this paper was to investigate how to improve the efficiency and effectiveness of CBIR based solely on color representation of the images. Improvement in efficiency was considered successful only if the performance of the system did not degrade as a result of our modifications to the retrieval system. By reducing the number of non-substantive colors and retaining only the primary colors in a query vector, color-based image retrieval performance was shown not to be impaired. Furthermore, by focusing only on sub-images containing the query object, retrieval effectiveness was shown to improve. Future work should consider applying the techniques we developed to larger databases with differing characteristics.

REFERENCES

[1] Thomas S. Huang, Yong Rui, and Shinh-Fu Chang, Image retrieval: Past, Present and Future, *International Symposium on Multimedia Information Processing*, 2007.

[2] C. C. Venters and M. Cooper, Content-based image retrieval, Technical Report, JISC Technology Application Program, 2009

[3] S.Selvam and Dr.S.Thabasu Kannan,"Design of an Effective Method for Image Retrieval", published IJIRAE, International Journal of Innovative Research in Advanced Engineering, Volume-1, March 2014, pp.51-56.

[4] S.Selvam and Dr.S.Thabasu Kannan," An Empirical Review on Image Retrieval System by using Relevance Feedback" proceeding of International Symposium on 'Research innovation for quality improvement in Higher Education' conducted by Bharathiar University, Coimbatore, October 2014 and published in "Research and Trends in Data mining and Image Processing Technologies and Applications", Bloomsbury publishing India, New York, Sydney,New Delhi, pp-1-11, October 2014, ISBN: 978-93-84052-11-9.

[5] P. G. B Enser, Query analysis in a visual information retrieval context, *Journal of Document and Text Management*, pg. 25 -52, 2013

[6] M. Myron 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 content: The QBIC system*", In IEEE Computer, pp. 23-31, Sept.2009

[7] Patrick M. Kelly, Michael Cannon and Donald R. Hush, Query by Image Example: The CANDID approach. In *Storage and Retrieval for Image and Video Databases III*, volume 2420, pp 238-248, SPIE, 2010

[8] Chad Carson, Serge Belongie, Hayit Greenspan, Jitendra Malik, *Blobworld: Image segmentation using Expectation-Maximization and its application to image querying*, Third International Conference on Visual Information Systems, 2012

[9] M. Das, E. M. Riseman, and B. A. Draper. Focus: Searching for multi-colored objects in a diverse image database, In *IEEE Conference on Computer Vision and Pattern Recognition*, pp 756-761, Jun 1997, *Proceedings of IEEE Intl. Conference on Image Processing*, pp. 568-571, 2012.

[10] V. E. Ogle and M. Stonebraker, *Chabot: retrieval from a relational database of images*, IEEE Computer, vol. 28, no. 9, pp. 40-8, Sept. 1995

[11] Michael Ortega, Yong Rui, Kaushik Chakrabarti, Sharad Mehrotra and Thomas S. Huang, *Supporting Similarity Queries*. In Proceeding of the ACM International Multimedia Conference, pp. 403-413, 2013.

[12] Petteri Kerminen, Moncef Gabbouj, *The Visual Goodness Evaluation of Colors Based Retrieval Processes*

[13] M. J. Swain and D. H. Ballard, Indexing via Color Histograms, ICCV'90, pp. 390- 393, 1990

[14] Tat-Seng Chua, Wai-Chee Low, and Chun-Xin Chu, Relevance feedback techniques for color-based image retrieval, *In Proceedings of Multi-Media Modeling'98*, IEEE Computer Society, pp 24-31, 2011.