International Journal of Electronics, Communication & Soft Computing Science and Engineering
ISSN: 2277-9477, Volume 2, Issue 2

Improving The Performance Of Content Based Image Retrieval System Using Color Distribution And Relevance Feedback

Sandip S. Patil, Atul V. Dusane

Abstract: **In this paper we use the color feature extraction technique that considers the image color distribution. For the color feature extraction process we use the clustering, where fixed number of clusters and variable number of clusters are formed. The proposed technique preserve the image color distribution and reduces the distortion that occurred during the feature extraction process by using binary quaternion moment preserving (BQMP) technique. With this system we devised efficient relevance feedback to improve the efficiency of the content based image retrieval system. Our Experimental results shows there is an improvement in proposed system over the prior binning technique.**

Keywords- **content based image retrieval, distance measure, color histogram, color feature extraction, relevance feedback, similarity measurement.**

I. INTRODUCTION

Content based image retrieval system overcome the lacunas of traditional text base image retrieval system such as heavy work load and strong subjectivity as need for efficient content based image retrieval [2][6] has need for efficient content based image retrieval has increases tremendously in many important area such as defense, biomedical, education, security, requires the correct and rapid searching among the large amount of databases. and due to this and become important and challenging research topic. Researcher follows the features like colors, texture, shape, for representation of image effectively and defines the proper similarity and measurement for matching the each feature. out of this features the color features is most commonly used in many applications[13][19] as it is tolerate the small change in the view point.

The color histogram is used to represent color distribution in a given image. The traditional color feature extraction method divides the total color space in to the fix number of set called as bins. Each bin representing some color and each of pixel is put in to the nearest colored bin and the size of bin is then the represented in terms of percentage of color in that image.

The measure disadvantages of the binning method that bin are static that is fix number of bin. So that this method does not consider the distribution of color in the image. Binning method fail's down when there is a color distribution in a image is mode .for example consider the image of group of flowers. Where particular color is present at different locations. The traditional method[20] just consider the percentage of particular color in the image without considering how it is distributed throughout the image. So that the traditional method is not achieve the balanced between expressiveness and compactness. In the new proposed color feature extraction method [21] use the fixed number of clusters and variable number of clusters and preserve the color distribution. If we consider the value of a pixel in an image as a random vector, the color distribution of this image will be equivalent to the probability distribution of this random vector. By probability theorems, a random vector can be characterized by its statistical moments [16]. In view of this, we apply the binary quaternion-moment-preserving (BQMP) thresholding technique proposed in [17] to the problem of color feature extraction. The BQMP thresholding technique can divide a data set into two subsets and , each of which has a representative, i.e., and , respectively. If we replace data points in with and those in with , the first 3 quaternion moments of are preserved in the resultant and . It is known that the higher the order of the moment, the less important that moment is. Thus, preserving moments up to the third moment would not distort the distribution of an image.

The resulting histograms from Fixed cluster or Variable Cluster will consist of different sets of colors. For calculating the distance between the two distribution we used, comparing histogram by clustering. To improve the efficiency of the system we also devised the relevance feedback mechanism. So that efficient results are return by the system. Relevance feedback (RF) is a query modification technique which attempts to capture the user's precise needs through iterative feedback and query refinement. In relevance feedback, whatever images are output when the system is executed first time, we can select the one relevant image from the displayed image and this image is given as an input to the system and the system is re-executed and we get the more correct result. The relevance feedback [1][13] mechanism makes it possible for CBIR systems to learn human concepts since users provide some positive and negative image labeling information, which helps systems to dynamically adapt and update the relevance of images to be retrieved.

II. WORKING MODEL OF SYSTEM

The following fig shows the working model of the proposed system.

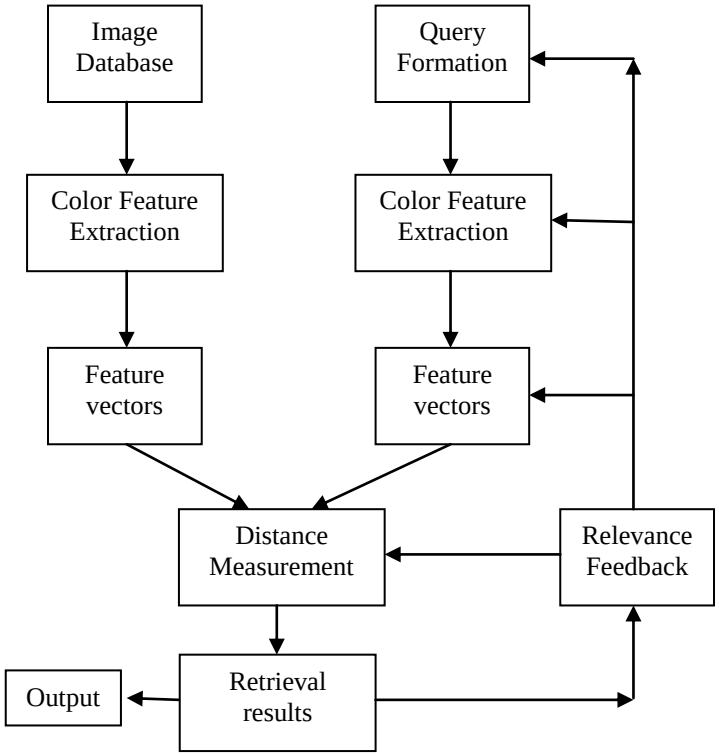

Figure1: Working model of system

III. COLOR FEATURE EXTRACTION

Let denote an image and α be a pixel in I . The color feature extraction is to define a function F : I → Q where Q= $r1,\ r2\ ,\ldots\ldots.r_n$ is the set of representative colors, such that F maps a pixel α to representative color r_i . The extracted color feature can be represented as a color histogram $H = h_{r1}, h_{r2}, \ldots\ldots\ldots, h_{rN}$, where $h_{r_i} = \sum_{\alpha \in I}^{Pr} \alpha | F\ \alpha\ = r_i$ represent the percentage of pixel in I which maps into r_i . If If we treat the value of a pixel in as a random vector, H represents a quantized probability distribution of I . The traditional extraction methods fix Q for all images disregarding their color distributions. There is need to find a proper function F for each image according to their color distribution and to define a proper distance measure for two histograms H_q and H_t obtained in this manner.

A. Binary Quaternion-Moment preserving-Thresholding technique: Based on the definition of the quaternion, the first three orders of quaternion moments are defined as follows:

$$m_1 = E\ q = E\ q_o\ + E\ q_1\ .i + E\ q_2\ .j + E\ q_3\ .k \qquad (1)$$

$$m_2 = E\ q^*.q\ = E(q_0{}^2 + q_1{}^2 + q_2{}^2 + q_3{}^2) \qquad (2)$$

$$m_3 = E\ q^*.q.q\ = E\ q_0{}^3 + q_0 q_1{}^2 + q_0 q_2{}^2 + q_0 q_3{}^2 + E\ q_1 q_0{}^2 + q_1{}^3 + q_1 q_2{}^2 + q_1 q_3{}^2\ .i + E\ q_2 q_0{}^2 + q_2 q_1{}^2 + q_2{}^3 + q_2 q_3{}^2\ .j + \mathrm{E}\ q_3 q_0{}^2 + q_3 q_1{}^2 + q_3 q_2{}^2 + q_3{}^3\ .\mathrm{k} \qquad (3)$$

Where E (.) represents the expectation. In practice, the expectation E(.) is replaced by the sample mean

B. Steps in extraction process

1. Input a dataset s.
2. Find a splitable sub cluster whose variance is maximum and then use the BQMP thresholding technique to split the sub cluster into two new sub clusters. If the sub cluster selected is unsplitable, mark it.
3. If there exist splitable sub clusters:
   i. In case of Fixed cluster, repeat step 2 until exactly N clusters are found
   ii. In case of variable cluster, repeat step 2 until the variance in sub cluster is below a variance threshold T_v.

The termination process for the fixed number of cluster and variable number of cluster different from each other. The process for fixed cluster terminate when specific number of pixel clusters have been extracted. Whereas the process for variable number of cluster terminate when number of pixel cluster are extracted till are sufficient to represent the image. This variance threshold value is predefined according to the previous result and experiment.

IV. DISTANCE MEASURE

As the fixed cluster and variable clusters gives the different sets of extracted colors $r1,\ r2\ ,\ldots\ldots.r_n$.

For the distance measurement we using the clustering based algorithm [21] it models h_r in H_q and H_t and as a point in the clustering problem. The distance between two point p_u and p_v is defined as the Euclidean distance between their representative colors

$$dist\ p_u, p_v\ =\ r_u - r_v \qquad (4)$$

Procedure for histogram comparison using clustering are as follow:

1. Map each h_r in H_q and H_t to a point

P = $(\vec{r}, w, img)$ in the clustering problem, where w = and $h_{\vec{r}}$

$$Img = \begin{cases} q & , h_{\vec{r}} \in H_q \\ t & , h_{\vec{r}} \in H_t \end{cases}.$$

2. Constraint on the cluster that the diameter of each cluster is not larger than threshold T_d.
3. Returns the distance between the clusters $H_q and\ H_t$.

V. Relevance Feedback Method

The relevance feedback mechanism makes it possible for CBIR systems to learn human concepts since users provide some positive and negative image labeling information, which helps systems to dynamically adapt and update the relevance of images to be retrieved. User query and feedback models from a user interface perspective Requiring a minimal amount of effort of a user is key in relevance feedback; if too much involvement is demanded, the user will be reluctant to use the system. Usually a user is searching for a particular category of items (category search) but the user can also be looking for something very specific (target search). A user searching for the latter will be harder to satisfy and the search process can then take many feedback rounds. A good user interface can help reduce the amount of interaction needed by making it easy for the user to submit a query and to give feedback.
In the proposed CBIR system, the relevance feedback processes are performed as
following steps:

i. A user request to CBIR system by submitting a query image.
ii. Firstly, the system will be extracted the color feature by considering the image color distribution. By using the feature extraction strategy of fixed cardinality and variable cardinality different sets of extracted colors $r_1, r_2, r_3, r_4, r_5, \ldots\ldots\ldots$, are resulted and these are represented by the histograms.
iii. After the feature extraction, all extracted feature will be form in feature vector.
iv. The query image then compare with images database by calculating the distance of dissimilarity between them.
v. Ranking all relevant images and sorting them in ascending order based on the distance of dissimilarity.
vi. Display resultant images with highest rank.
vii. If user did not satisfy the retrieval result that display by system, then user can feedback the relevant images to system.
viii. As before, system processes the feature extraction module.
ix. The system then ranking the images and display the top relevant images to user.

VI. Experimental Result

The CBIR system performance measurement is based on the Precision and Recall. The experimental result are conducted using Matlab7.5 on Intel core 2, 1.60GHz processor with 2GB memory. For this experiment we used image database publish by corel corporation.

Precision is defined as the number of relevant images retrieved by a search devise by the total number of images retrieved by the search.

Recall is defined as the number of relevant images retrieved by search divided by total number of existing relevant image

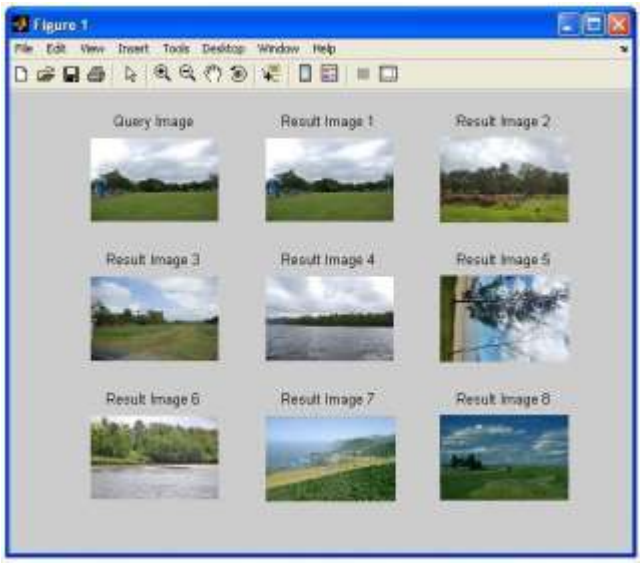

Figure2: Output of the queried image binning method.

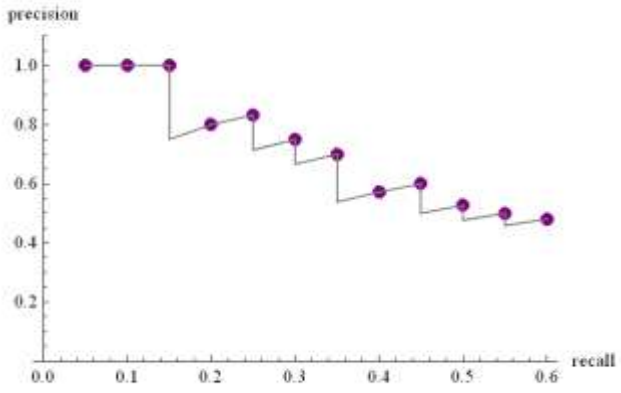

Figure 3: Precision Recall curve of prior method

The above figure 2 shows the output given by the binning method where we used the Euclidean distance for similarity comparison. The figure 4 shows the resultant output of the proposed method. The precision and recall values for the binning method with Euclidean distance are shown in Table 1. Whereas precision recall value for proposed feature extraction along with the relevance feedback mechanism are shown in Table 2. In both N1 is number of images retrieved and relevant. N2 is actual number of relevant images in the database.25 image are retrieved and displayed to the user. From those value the precision and recall is calculated.

Category	Image	N1	N2	Precision	Recall
Land & sky	Bo.jpg	12	20	48 %	60 %
Building	So.jpg	11	20	44 %	55 %
Sea	m1.jpg	10	18	40 %	56 %
mountain	S2.jpg	11	18	44 %	61 %

Table1: precision & recall values of binning method

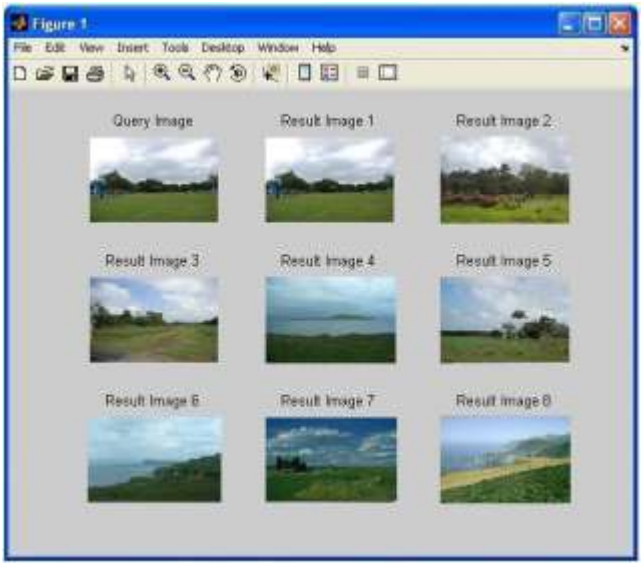

Figure 4: output of the queried image using the proposed method.

Category	Image	N1	N2	Precision	Recall
Land & sky	Bo.jpg	15	20	60 %	75 %
Building	So.jpg	14	20	56 %	70 %
Sea	m1.jpg	13	18	52 %	72 %
Mountain	S2.jpg	14	18	56 %	78 %

Table 2: precision and recall values of proposed color feature extraction and relevance feedback method

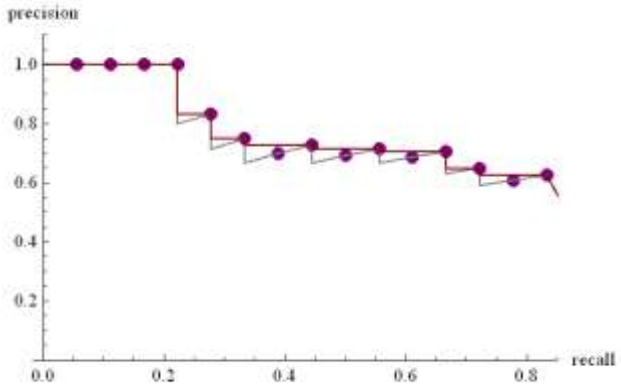

Figure 5: Precision & Recall curve of proposed method.

As per the analysis of our results we get the more accurate result as compared to the binning method. The relevance feedback technique increases the result of the content based image retrieval system more efficiently. Although the time requirement for the result with proposed feature extraction method with relevance feedback is more than traditional binning method. Because it requires more computation and used feedback but it preserves the image color distribution and also increase the result. Many other low level feature extraction techniques are proposed towards the Content Based Image retrieval problem that can be also used with proposed the system but that again increase the time of execution as the complexity of the overall process is increased due to more operations.

VII. CONCLUSION

In this paper, using the binary quaternion moment preserving thresholding technique, we extracted the color feature that preserves the image color distribution upto the third moment. By considering the clustering technique for feature extraction reduces the distortion during color feature extraction process. We focus on applying the new extraction methods to the global color feature extraction. However, as commonly used in practice, an image is usually divided into several subblocks and the regional color features are extracted. In this case, our extraction methods and distance measures will be even more preferable to the traditional binning methods because the color content in a subblock tends to be dominated by only a few colors. Our extraction methods are able to adaptively extract those dominated colors while the binning methods do not provide such adaptability. As the experimental result shows we achieved the improvement in the feature extraction process. Also our relevance feedback technique helps the user for the further

improving in the result as per requirement. The relevance feedback is effective and improves the retrieval performance of the content based image retrieval system significantly.

REFERENCES

[1] S. Antani, R. Kasturi, and R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing, and retrieval of images and video," *Pattern Recognit.*, vol. 35, no. 4, pp. 945–965, 2002.
[2] R. Veltkamp, "Content-based image retrieval system: A survey," University of Utrecht, Tech. Rep., 2002
[3] W. H. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *J. Classificat.*, vol. 1, pp. 1–24,1984.
[4] D. Defays, "An effecient alogrithm for a complete link method," *Comput. J.*, vol. 20, no. 4, pp. 364–366, 1977.
[5] 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," *IEEE Computer*, vol. 28, no. 9, pp. 23–32, Sep. 1995.
[6] Mourad Oussalah "Content Based Image Retrieval: Review of State of Art and Future Directions, IEEE Image processing Theory Tools & application 2008.
[7] H. Frigui, "Visualizing and browsing large image databases," in *Proc. Int. Conf. Information and Knowledge Engineering*, 2004, pp. 68–74. [8] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Nov. 1998.
[9] F. S. Hiller and G. J. Liberman, *Introduction to Mathematical Programming*.New York: McGraw-Hill, 1990.
[10] B. King, "Step-wise clustering procedures," *J. Amer. Statist. Assoc.*, vol. 69, pp. 86–101, 1967.
[11] V. Klee and G. Minty, "How good is the simplex algorithm," in *Inequalities*, 1972, vol. 3, pp. 159–175.
[12] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Query feedback for interactive image retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 644–655, May 2004.
[13] F. Long, H. Zhang, and D. D. Feng, "Fundamentals of content-based image retrieval," in *Multimedia Information Retrieval and Management Technological Fundamentals and Applications*. New York: Springer-Verlag, 2003.
[14] M. K. Mandal, T. Aboulnasr, and S. Panchanathan, "Image indexing using moments and wavelets," *IEEE Trans. Consum. Electron.*, vol. 42, no. 3, pp. 557–565, Aug. 1996.
[15] M. Oge and F. Borko, "Muse: A content-based image search and retrieval system using relevance feedback," *Multimedia Tools Appl.*, vol. 17, pp. 21–50, 2002.
[16] A. Papoulis and S. U. Pillair, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 2002.
[17] S.-C. Pei and C.-M. Cheng, "Color image processing by using binary quaternion moment-preserving thresholding technique," *IEEE Trans.Image Process.*, vol. 8, no. 5, pp. 614–628, May 1999.
[18] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. IEEE Int. Conf. Computer Vision*, 1998, p. 59.
[19] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *J. Vis. Commun. Image Represent.*, vol. 10, no. 1, Mar. 1999.
[20] A. Smeulders, M.Worring, S. Santini, A. Gupta, and R. Jain, "Contentbased image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
[21] W. Chen, W. Liu, M. Chen, "Adaptive Color Feature Extraction Based on Image Color Distributions" IEEE Transaction on image processing, vol 19, No. 8, August 2010.
[22] J. R. Smith and C. Li, "Image classification and querying using composite region templates," *Comput. Vis. Image Understand.*, vol. 75, pp. 165–174, 1999.
[23] M. Stricker and M. Orengo, "Similarity of color images," in *Proc. Storage and Retrieval for Image and Video Databases (SPIE)*, 1995, pp. 381–392.
[24] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, pp. 11–32, 1991.

AUTHOR'S PROFILE

Prof. Sandip S. Patil
Associate Professor in department of Computer Engineering at S.S.B.T. College of Engineering and Technology, Babhori Jalgaon. (M.S.) having 12 years of research experience. His area of interests are machine learning and soft computing .
Ph:9420347893
E-mail:san_78004@yahoo.co.in

Atul V. Dusane
Persuing M.E. at SSBT College of engineering & Technology, jalgaon
Ph:7588816387
E-mail:atulp_1987@yahoo.co.in