Image Similarity Retrieval for FMIQ by CCV and Histogram Refinement

Satyawati M.Patil, V.K.Patil

Abstract— **Fast Multiresolution Image Querying or Content based image retrieval or Content Based Visual Information Retrieval or 'Query by Image Content' is one of the most interesting and important issues in digital image processing field. The aim of this paper that has been focused here is the multi-resolution image querying based on Color Coherence Vector and histogram refinement algorithm along with maximum effectiveness of color characteristics and mathematical model. Generally, the two methods for image retrieval have been implemented are color histogram refinement(CHR) and Color Coherence Vector (CCV) so as to generate the feature vectors on estimation basis. However, the query image can be easily retrieved from database despite the storage of several images with an increment in speed of retrieval since CCV plays a crucial role in it. Color feature extraction technique on basis of CCV assist in flawless retrieval of an image with maximum precision and accuracy. Therefore, the CHR method generates the result of about 80% while CCV method ultimately generates the result of about 89% in terms ofprecision in image retrieval efficiency.**
***Keywords:*Fast multiresolution image querying, CCV, digital image signatures format.**

I.INTRODUCTION

Modern era of digital image processing has totally revolutionized the standards and ubiquity of images since they exhibit the cardinality of communication through the assortments of ever increasing demand for the library of several digital images. With an ever- growing library of processed digital images, the necessity of image retrieval has become the most crucial issue in the aspect of digital image processing. Fast multiresolution image querying or retrieval basically deals with the searching of of desired images from the datasets on the basis of variety of color, texture, shape and sketch characteristics[1]. This paper highlights an extreme novel ideology of fast and flawless image retrieval from the magnificient database on the basis of Color Coherence Vector(CCV) generation along with the histogram refinement under the attributes of the color characteristics and mathematical model for multiresolution approach. Many applications require simple methods for comparing pairs of images based on their overall appearance. For example, a user may wish to retrieve all images similar to a given image from a large database of images. Color histograms are a popular solution to this problem, the histogram describes the gray-level or color distribution for a given image, they are computationally efficient, but generally insensitive to small changes in camera position. Color histograms also have some limitations. A color histogram provides no spatial information; it merely describes which colors are present in the image, and in what quantities. In addition, color histograms are sensitive to both compression artifacts and changes in overall image brightness[8].However, due to maximum confinement of histogram features and the concerned strategies involved in theCCV computional mathematical model, image retrieval seems to possess extravagant potential in searching desired images with maximum precision and proficiency. To build an image querying based on color as well as generic shape information is the major objective of fast multiresolution image querying approach. Besides this ,user must be able to execute a query by example (i.e. specify a sample image to be used as query input). In another point of view user must also be able to execute a query by sketch (i.e. draw a sample sketch to be used as query input).The system must provide facility to the user to be able to use multi-resolution images. This technique supports both color and grayscale images which is an ever-increasing demand and objective in the aspect of image retrieval and processing. Image querying uses all forms of databases for fast execution of queries so that accuracy level can be enhanced to maximum extent. Another crucial objective can be considered that the user must be able to see multiple results in ascending order of comparison results. However, this mechanism of image querying should also provide easy database management. i.e. addition and deletion of images to database. Due to this, task accomplishment can be handled in the most lucid manner.

II. RELATED WORK

The concept of Content Based Image Retrieval (CBIR) was introduced by T. Kato in 1992.He initiated his experiments into automatic retrieval of images from a database on the basis of color and shape features. Furthermore, the database search and querying has received attention of several groups. Methods for reduction of the size of the characteristic vectors and increasing the speed of recovery in great databases were developed by Swain—Ballard (1991) and by Stricker--Orengo (1995). Efficient index structures were developed by Petrakis--Faloutsos (1995) and White--Jain (1996, 1997). Pre-filter techniques that reduce the amount of images in database for similarity computation was accomplished by Niblack et al. (1993). Several CBIR systems currently exist, and are being constantly developed. Another scheme, as reported by P. S. Hiremath, S. Shivashankar, J. Pujari,(2006) explains a similar kind of approach based on color texture analysis by using different color spaces. The work reported by Tian Yumin, Mei Lixia, "Image Retrieval Based on Multiple Features Using

***Satyawati M.Patil**, Department of Electronics and Telecommunication,D.N.Patel College of Engineering(Shahada),North Maharashtra,University,Jalgaon,India,9765157576 (satyawatipatil13@gmail.com)*
***V.K.Patil**,Department of Electronics and Telecommunication,D.N.Patel College of Engineering(Shahada),North Maharashtra University,Jalgaon,India.*

ISSN: 2278 – 909X

Wavelet," introduces a scheme based on wavelet which includes both color feature and texture feature circular region energy in low frequency band of wavelet transform of an image is used as color feature of the image and the synthesize energy in high frequency bands of multi-scale wavelet transforms is used as texture feature.

III. PROPOSED METHODOLOGY

A.System Model

The significant mechanism that has been emphasized here is the fast multiresolution image querying using CCV along with the combination of refined color histogram in accordance to the similarity measure mathematical model. Consequently, the digital signature matrices/ manhattan/Euclidean distance values are estimated and confined values are stored in database. Since CCV plays an active role in increasing the speed of image retrieval, the query images are instantly retrieved. Besides these, the inceptional stages of registration process includes image acquisition, RGB to HSV conversion, segmentation, CCV computation and generation of Color coherency vectors .System strategy gets implemented in several steps.Figure1 shows the block diagram of proposed system which highlights the confined image retrieval using both techniques.

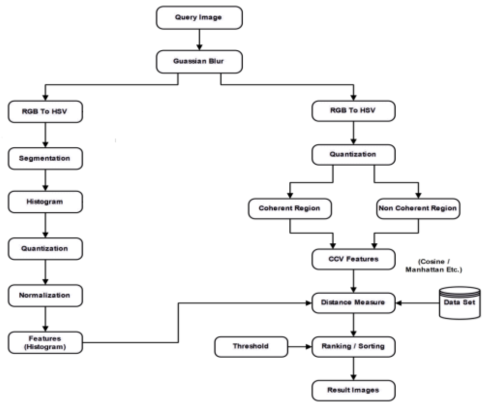


Fig.1 Block Diagram of Proposed System

B. Methodology

The exact mechanism in the proposed system includes following algorithms for color histogram and CCV systems in order to extract the features at required stages.

- Color histogram Refinement Technique

Color is one of the most outstanding features of the image, it is the most important human visual content and it is very easy to calculate. It is widely concerned by many researchers because it does not effected by natural rotation, scaling and translation of a image. The color distributions in the images can be represented using the Color Histograms. The histogram of the query image and the database images are compared for the retrieval. Color Histogram only describes global color distribution of an image. Color Histogram is a commonly used feature in image retrieval. It is very popular because Color histograms are computationally trivial to compute. Small changes in camera view point does not does effect in the histogram.

In color histogram, all images are scaled to contain the same number of pixels. The color space of the image are discretized in such a way that there are distinct (discretized) colors. Color histogram H is a vector h1, h2,..., hni, in which each bucket hj contains the number of pixels of color j in the image. Typically images are represented in the RGB color space, and a few of the most significant bits are used from each color channel. For example, Zhang uses the 2 most significant bits of each color channel, for a total of n = 64 buckets in the histogram. For a given image I, the color histogram HI is a compact summary of the image. A database of images can be queried to the most similar image to I, and can return the image I*0* with the most similar color histogram HI*0* . Typically color histograms are compared using the sum of squared differences (L2-distance) or the sum of absolute value of differences (L1-distance). Differences are weighted evenly across different color buckets for simplicity. The sequential strategy can be modeled in form of algorithm stated below.

- Proposed Color Histogram Refinement algorithm

Step1: Acquire the query image and blur it.
Step2: Convert it from RGB to HSV color space.
Step3: Determine HSV histogram features before segmentation.
Step4:Apply segmentation to HSV color image so as to divide it into number of blocks.
Step5:Quantise each pixel in HSV space to set histogram bins.
Step6:Obtain the normalized histogram by dividing it with the total number of pixels.
Step 7:Compute the histogram features and store them as color feature vector.
Step8:Repeat the steps 1 to 7 for all the images present in the database.
Step9: Compute Euclidean distance by comparing the query image histograms to that of each image in
the database.
Step 10: Sort images in database in order of ascending Euclidean distance to query image and return as
result.

- Color Coherence Vector Technique

The fundamental concept of our approach is based on CCV method, although we present some new techniques mostly based on suitable segmentation of different stages of CCV while improving the performance of each phase by modifications in computational stages for indexing and retrieval process from database. Intuitively, we define a color's coherence as the degree to which pixels of that color are members of large similarly-colored regions. This notion of coherence allows us to make fine distinctions that cannot be made with simple color histograms. For a given discredited color , some of the pixels with that color will be coherent and some will be incoherent. Let us call the number of coherent pixels of the jth discrete color αj and the number of incoherent pixels βj. Clearly, the total number of pixels with that color is $\alpha j + \beta j$, and so a color histogram would summarize an image as

$$<\alpha_1+\beta_1,\ldots,\alpha_n+\beta_n> \quad \textbf{(i)}$$

Instead, for each color we compute the pair (αj, βj) which we will call the coherence pair for the jth color. The color coherence pair's vector for the image consists of

$$<(\alpha_1,\beta_1),\ldots,(\alpha_n,\beta_n)> \quad \textbf{(ii)}$$

Classification of coherence is determined by a fixed value τ. Each pixel is checked whether coherent or not. A pixel is coherent if its surrounding pixels have the same values to form a large contiguous region.

ISSN: 2278 – 909X

- Proposed CCV algorithm

Step 1: Blur the acquired input query image by guassian filter.
Step 2: Convert RGB color components to HSV color space.
Step3: Discretize hsv color space into n distinct colors.
Step4:Determine connected components of each discretized color and classify the pixels as either coherent or incoherent.
Step5: Compute image coherencyto generate coherency feature vector
Step6: Evaluate Eucleidean distance between the ccv generated vectors and the stored vectors present in the database.
Step 7: Compare the generated values with the threshold to retrieve the desired images from database.
Step8:Repeat steps 1to 7 for all query images.

- Similarity Measurement Algorithm

Similarity measurement occupies a crucial role in CBIR algorithms. These algorithms search image database to find images similar to a given query so that they should be able to evaluate the amount of similarities between the images. To measure the similarity, the direct Euclidean distance between the image in the database and the query image is given by,

E.D.= ∑(Vpi-Vqi)(Vpi-Vqi) (iii)

where, Vpi and Vqi are the feature vectors of image and query image with respect to size'n'.It operates by considering each vector as a point in an 'n' dimensional vector space and assist in calculation of physical distance between two vector points of the images.

- Matching function

In CCV,we compute two values (C and N) for each color;where 'C' is the number of coherent pixels and 'N' is the number of incoherent pixels.It is clear that the summation of all color's C and N = number of pixels. Suppose that we have to compare 2 images a, b[5]. Therefore, matching function can be elaborated as,

$$D(a,b) = \sum_{i=0}^{n} (|\alpha_{Ci}-\beta_{Ni}|)+(|\alpha_{Ni}-\beta_{Ni}|) \quad \textbf{(iv)}$$

C_i : number of coherent pixels in color i.
N_i : number of incoherent pixels in color i.

This algorithm implies the strategy of matching on color by color basis. Inceptionally, the number of colors in query image and database image are estimated by analyzing histogram .However, if the amount of specific color in both the images are comparable, then the images gets easily matched. Retrieval result is not based on a display of a single image but the combination of multiple similarly colored images since content based image retrieval is not based on the exact matching. Suppose that 'q' is the query image while q' is the database image, then similarity measure can be evaluated by following steps:
Step1:.Determine color histogram vectors and CCV vectors of the database images.
Step2:Compute color histogram vectors and CCV vectors of the query image.
Step3:Calculate Euclidean distance between two feature vectors.
Step4:If d<=≤threshold, then images gets sorted or ranked and can be matched accordingly.

IV. Performance Evaluation

A. Parameters Estimation

Generally, there are two parameters namely precision and recall which helps to evaluate the performance of image retrieval system[1]. For a query q, the data set of images in the database that are relevant to the query q is denoted as $R(q)$, and the retrieval result of the query q is denoted as $Q(q)$.It has been reported that the histogram exhibits the best performance through the estimation of following parameters.

(a)Precision: The precision of the retrieval is defined as the fraction of the retrieved images that are indeed relevant for the query. It measures the ability of the system to retrieve only the models that are relevant. It can be defined as,

$$\textbf{Precision} = \frac{\textbf{Number of relevant images retrieved}}{\textbf{Total number of images retrieved}} \quad \textbf{(v)}$$

(b) Recall: The recall can be defined as the fraction of relevant images that is returned by the query.It measures the ability of the system to retrieve all the models that are relevant.

$$\textbf{Recall} = \frac{\textbf{Number of relevant images retrieved}}{\textbf{Total number of relevant images}} \quad \textbf{(vi)}$$

(c) Accuracy: An accuracy of the system can be defined as the half the fraction sum of precision and recall values, which can be expressed as,

Accuracy= (Precision + Recall) /2 (vii)

(d) Match Percentile =The match percentile MP of a given image is expressed as:

$$\textbf{MP} = (\textbf{N} - \textbf{R}) / (\textbf{N} - \textbf{1}) \quad \textbf{(viii)}$$

where N is the number of images in the database and R is the rank of the returned image.

All these parameters are calculated for each image and the results are averaged.

B. Experimental Results

Our system has been implemented to execute on the platform of MATLAB R2010 'a' version along with the procession of several modules related to digital image processing. It utilizes several tools of MATLAB concerned with the image acquisition,pre-processing,segmentation,quantization,segmentation , thresholding and so on. This system comprises the database of about 732 images including six different categories with the assortments of several images. All the images are stored in jpg format of different dimension size.Thus, according to the variation in database types, the system display variations in the accuracy retrieval performances. However, the modules and tools of the system can be incremented as per the required level. Table 1 and Table 2 indicates an ideology of computed comparison results of color histogram refinement and CCV mechanisms.

Table1. Experimental Retrieval Results of Color Histogram Refinement (CHR) Technique

Query Image Categories	Color Histogram Refinement Technique			
	Precision Rate [%]	Recall Rate [%]	Accuracy Rate [%]	Match Percentile [%]
British Army	83.2	72.2	77.7	71.8
Mountain	94.3	60.4	77.3	68.6
Bear	91.5	80.1	85.9	80.2
Deer	58.5	55.6	56.5	52.4

ISSN: 2278 – 909X

Eagle	**54.2**	**51.3**	**52.5**	**49.2**
Firework	**65.8**	**55.2**	**60.4**	**57.1**
AVG	**74.5**	**62.4**	**68.3**	**63.2**

Table 2. Experimental Retrieval Results of Color Coherence Vector (CCV) Technique

Query Image Categories	Color Coherence Vector Technique			
	Precision Rate [%]	Recall Rate [%]	Accuracy Rate [%]	Match Percentile [%]
British Army	**98.5**	**93.7**	**96.1**	**92.9**
Mountain	**99.9**	**85.9**	**92.9**	**89.8**
Bear	**96.8**	**83.5**	**90.1**	**83.6**
Deer	**60.4**	**80.6**	**70.5**	**75.4**
Eagle	**95.3**	**40.8**	**68.0**	**72.5**
Firework	**97.6**	**86.9**	**92.2**	**78.9**
AVG	**91.4**	**78.5**	**84.9**	**81.6**

The tabulated values are averaged for each and every image present in the category of query images so as to estimate precision and recall values. Precision and recall measures the accuracy confined match percentile of image retrieval efficiency of the overall system.

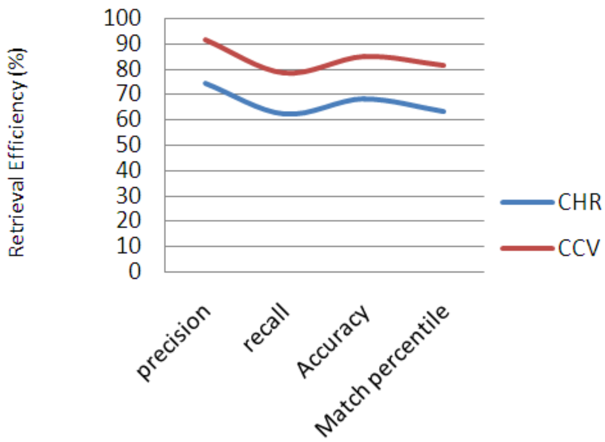


Fig 2.Comparative Color Histogram Refinement and Color Coherent Vector methods of the Proposed FMIQ for various retrieval efficiency measurement parameters

V. Conclusion

With the advent of several search engines, image searching has almost become an easier task. But, the existing search engines utilizes textbased image indexing and retrieval mechanisms which are not sufficient to the upheaval requirement. Therefore, fast multiresolution image querying provides an extravagant impact to complement existing mechanisms so as to allow the system to retrieve more similar images from the source images and hence provide effectively precise results. The major objective of this paper signifies the strategy of improvement in the accuracy and speed of image retrieval in order to handle predicaments of large image databases while optimizing several applications with more efficiency.

REFERENCES

[1] K.Valli Madhav1, R.Tamilkod, .Bala Dinakar, K.JayaSudha, "An Innovative Technique for Content Based Image Retrieval Using Color and Texture Features," International Journal of Innovative Research in Computer and Communication Engineering,Vol. 1, Issue 5, July 2013.

[2] G.Raghupati, R.S.Anandand M.L.Dewal, "color and texture features for content based image retrieval using motif", second international conference on multimedia and content based image retrieval, july21-23,2010.

[3] Young Deok Chun, Nam Chul Kim, Member, IEEE, and Ick Hoon Jang, Member, IEEE, "Content-Based Image Retrieval Using Multiresolution Color and Texture Features", IEEE Transactions On Multimedia, Vol. 10, No. 6, October 2008.

[4] Michael S. Lew, Nicu Sebe, Chabane Djeraba, and Ramesh Jain, "Content-based multimedia information retrieval: State of the artand challenges," ACM Trans. Multimedia Comput. Commun. Appl., vol. 2, no. 1, pp. 1-19, 2006.

[5] Irena Valova, Boris Rachev and Michael Vassilakopoulos, "Optimization of the Algorithm for Image Retrieval by Color Features",International Conference on Computer Systems and Technologies- CompSysTech" , pp 1-4, 2006.

[6] Shamik Sural, Gang Qian and Sakti Pramanik, "Segmentation and Histogram Generation Using the HSV Color Space for ImageRetrieval", International Conference on Image processing, Vol. 2, pp. 589-592, 2002.

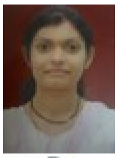
[7] Aleksandra Mojsilovic, Jianying Hu and Emina Soljanin, "Extraction of Perceptually Important Colors and Similarity Measurement for Image Matching, Retrieval, and Analysis", IEEE Transactions on Image Processing, Vol. 11, No. 11, pp 1238-1248, November 2002.

[8] Greg Pass, Ramin Zabih and Justin Miller, "Comparing Images Using Color Coherence Vectors", Proceedings of the fourth ACM international conference on Multimedia,pp.65-73,1996.

[9] Vellaikal and C. C. J. Kuo, Content Based Image Retrieval using Multiresolution Histogram Representation" " , SPIE - Digital Image Storage and Archiving Systems, Vol. 2606, pp. 312-323, 1995.

BIOGRAPHY

ISSN: 2278 – 909X



Satyawati M. Patil is a P.G. student .She has completed her B.E.(E&TC) and currently pursuing M.E. from D.N. Patel College of Engineering under North Maharashtra University, Jalgaon. She is mainly interested in Digital Image Processing..

V.K.Patil is working as a professor, Head of Department (H.O.D.) of E &TC Department in D.N.Patel College of Engineering ,Shahada..He graduated from S.S.G.M.College of Engineering, Shegaon under Amravati University and accomplished M.E. in Electronics from Jawaharlal Nehru Engineering College under Dr.Babasaheb Ambedkar Maratwada University ,Aurangabad. His special field of interest include Digital Image Processing, Power Electronics and Digital Communication.

ISSN: 2278 – 909X