|| Volume 3 || Issue 10 || October 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Content Based Image Retrieval using Colour Histogram Area Co-efficients

Vijayakumar Bhandi[1], Sumithra Devi. K. A.[2]
Dept of CSE, Jain University, Bangalore, India[1]
Dept of ISE, Dayananda Sagar Academy of Technology & Management, Bangalore, India [2]
vijay.bhandi@gmail.com[1]

***Abstract*— In this paper, we propose a novel method for improving the precision of Content based image retrieval [CBIR] by segmenting image into sub regions in a new way and use colour histogram with area coefficients. A new method is proposed to divide an image into 5 simple rectangular sub regions and assign weights according to their area. A bigger central rectangular area is chosen and assigned more weight as the centre region contains important information about the image. Colour histogram feature is computed for each region separately and the resultant features are clubbed into a fusion vector according to the region area coefficient. Additionally, we use univariate chi-squared statistical test for features selection to select best 35% of the features for image similarity measurement. Chi-squared distances are used to measure the feature similarity for image retrieval. The proposed method has been validated through experiments conducted on standard Corel1K image dataset. It is observed that the proposed CBIR method improves average precision of the image retrieval by 6.33% as compared against baseline CBIR using colour histogram feature over entire image.**

Keywords: *Content based image retrieval, Colour histogram, Feature Vector.*

I INTRODUCTION

Content based image retrieval [CBIR] deals with retrieval of matching images from a database of images using low level features of the image. CBIR is used extensively in various real world applications like medical diagnosis, face detection, retail catalogues, digital gallery etc. There are various low-level features used in CBIR, the important ones are colour, texture and shape features. Colour is the most dominant and distinguished visual feature of a digital image. The colour feature is relatively robust to background complication and independent of image size and orientation. Hence colour is most widely used visual feature in CBIR. Colour histogram is the basic colour content representation and it captures global colour distribution of an image. Colour histogram describes the statistical colour distributions by quantizing the colour space. Colour histograms are easy to compute, but they result in large feature vectors which are difficult to index. The colour histogram feature does not show the colour layout information of an image and results in an increased rate of false positives. This is critical for the image retrieval from large image databases, where many images might have similar colour histograms. Various approaches have been proposed to add the colour layout information with colour histogram to overcome this issue. The commonly used technique is to divide the image into different regions and compute colour features separately for each region. Many ways have been proposed to partition an image into sub regions for this purpose.

In this work, we have proposed a new way of dividing image into simple sub regions. We partition image into 4 border rectangular regions of same size and 1 big centre rectangular region. The reason being, the centre part of the image contains the most information in the image. We have chosen rectangular shapes for ease of computation. We have assigned weightage to colour feature of each region based on region area, we call them as area coefficients. This will help us to give more weightage to the centre region of the image and assign lower significance to the border regions.

Colour histogram calculated over different regions of an image produce a quite large dimension. This high dimensionality problem will result into high computational cost in distance calculation for similarity retrieval and inefficiency in indexing and search. To overcome this, we have used univariate chi-squared statistical test features selection method. This will enable us to select best compact feature set for image features comparison without additional computing cost.

This paper is organized as follows: Section II describes the related work. Section III explains the proposed methodology. Section IV details about the experimental results and Section V presents the conclusion of this research work.

II RELATED WORK

Retrieving images similar to a query image from a given set of images has always been an interesting research problem. Traditional text-based image retrieval methods used manually annotated keywords for searching the relevant images. This is labour-intensive, time consuming and expensive method. Also, the rich semantics of the image are difficult to describe and human perception affects it [1] [2]. This led to development of content based image retrieval [CBIR] method, which relies on low level features that can be automatically extracted from the image. A typical CBIR system works on query by example method, where low level features of query image are compared against database image features to extract top N similar images [1] [2]. Colour, texture and shape features are most widely used and accepted low level visual features in CBIR [1] [2].

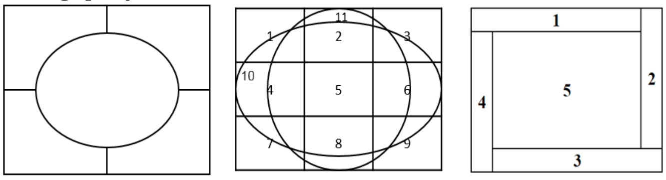
Colour feature is considered as the most distinguishing and dominant low level visual feature in CBIR [2]. Humans perceive colour as a combination of three basic colours: Red, Green and Blue. These form the basic colour space RGB, which

|| Volume 3 || Issue 10 || October 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

is widely used in image processing. By separating the luminance from the chromatic information, we can obtain more colour spaces. There are various popular colour spaces used in image processing, important ones being RGB, HSV, LAB, and CIE. Among all colour spaces, the Hue Saturation Value [HSV] is considered perceptual. This means colours can be matched in a way that is consistent with human perception. Hence HSV colour space is most suitable for colour image analysis [12] [13]. HSV colour space is used in our work.

The colour histogram feature is simple to calculate and provides good discriminating power in image retrieval. In cases where the image collection is large, the global colour feature tends to give many false positives. Hence general thought is to extend the global feature to local regions. A general practice to divide the image into multiple blocks and extract colour feature from each of the block [4] [5]. Malinga, Bongani et al. [6] partitioned image into a fixed number of equal blocks and represented them in terms of their local (block-wise) HSV histogram but no weights or coefficients were used to give importance to any of the blocks.

Stricker and Dimai [3] proposed a method where they divided an image into 5 partially overlapping regions as in Fig. 1(a) and computed colour moments for each region separately. They gave more weight to the centre eclipse region than other regions. The reason being the most important information of an image is captured in centre of the image. Stricker and Dimai [3] used arbitrary weights between 0 and 5 for the sub regions which can be specified during query time.



1(a) Stricker and Dimai [3] 1(b) Juling et al. [7] 1(c) Proposed Method

Figure 1 Image partition methods from earlier methods and current proposed method

III PROPOSED WORK

The proposed CBIR method consists of an offline database indexing module and online query module as outlined in Fig. 2.

A. Offline database indexing

In this module, we process all images in the database to compute colour histogram feature and create a fusion features database. This is one-time offline exercise. Every image is converted to HSV colour space and partitioned into 5 regions as depicted in Fig. 1(c). Colour histogram feature [8 Hue bins, 12 Saturation bins, and 3 Value bins] is calculated and normalized for each region separately. Colour histogram feature calculated for each region produces a vector of 288 values. Hence for each image we get a fusion vector of 1440 values as we have divided image into 5 regions. The calculated colour histogram features are stored as a fusion vector in features database.

B. Online query module

In this online module, the retrieval of matching images corresponding to input query image is performed. The query image is converted to HSV colour space and divided into 5 regions as in Fig. 1(c), and then colour histogram feature [H-8, S-12, V-3 bins] is calculated and normalized for each region. The resultant 1440 features fusion vector is subjected to features selection to select best suitable features for image retrieval. We use univariate chi-squared statistical test selection method to select best features from the set of fusion vectors. Our experiments have shown that 35% of the features (504 features) provides better image retrieval performance. These features are compared against the same set of 35% best features from features database using chi-squared similarity measurement method. Based on the chi-squared distances, the retrieved images are ranked and top N images are returned.

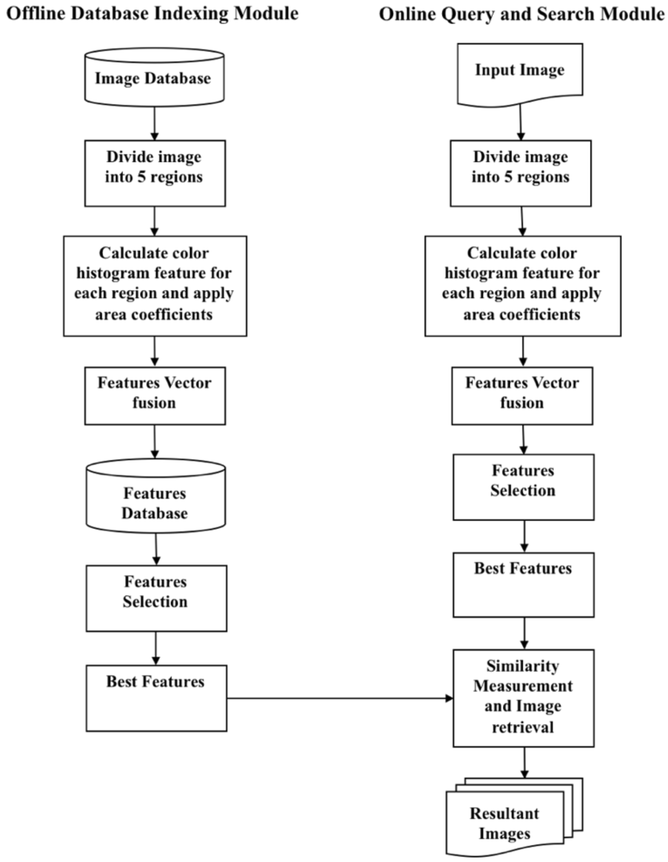


Fig. 2 Image partition methods from earlier methods and current proposed method

C. Performance Evaluation

The performance of the image retrieval is measured in terms of precision and recall rates as defined by equations 1 and 2.

Precision = Number of relevant images retrieved / Total number of images retrieved

(1)

Recall = Number of relevant images retrieved / Total number of images in the image class

(2)

D. Implementation

The proposed CBIR method is implemented using Python 2.7 on Windows 7 (64 bit) operating system with 16GB RAM and Intel core i7 processor.

IV EXPERIMENTAL WORK

Proposed CBIR method is subjected to experimental study on standard Corel 1K image dataset [Wang's dataset]. The dataset consists of 1000 images which are grouped into 10 classes with 100 images in each class [9]. The images are of the size 256x384 or 384x256. Given a query image from a class, it is assumed that the user is searching for images from the same class, and therefore the remaining 99 images from the same class are considered relevant and the images from all other classes are considered irrelevant. The precision and recall rates are measured by varying the number of images retrieved.

The experiment is carried out in 3 steps as outlined in below sections. One image from each class is used as query image. Same query images have been used in all 3 experiments.

A. Global Colour Histogram feature only [Method 1]

We create a CBIR model by using colour histogram feature on the entire image. We use this as a baseline model to measure performance of proposed method over it. Since it is difficult to compare against earlier methods [3] [7] due to difference in image dataset and system configurations used.

For the baseline CBIR, colour histogram feature [H-8, S-12, V-3 bins] is calculated in HSV colour space for each image in the dataset and stored in features database. Similarly, colour histogram feature for query image is calculated and compared against features database. Chi-squared distances method is used for features similarity measurement and matching images are retrieved. In this baseline CBIR, we get 64.47% average precision and 25.87% average recall across all classes. The class wise average precision is tabulated in table 1.

B. Color histogram for 5 regions without area coefficients [Method 2]

This step is carried out to check whether the features selection impacts the average precision of image retrieval in the proposed CBIR. The image is converted to HSV colour space and is split into 5 rectangular regions as in Fig. 1(c). Then colour histogram feature is calculated for each region and feature fusion vector of 1440 values is created. No weight coefficients are used during features fusion. Then we utilize univariate chi-squared statistical test features selection process to select best suitable 35% of features [504 features]. For few classes like 3 and 9, the average precision is quite lower than the baseline CBIR (method 1). This method gives us average precision of 57.05% and average recall 22.81%, which is lower than the baseline CBIR. The image retrieval results from this method are tabulated in below table 1.

E. Colour histogram for 5 regions with area coefficients [Method 3]

In this step, our proposed CBIR method using region area coefficients is subjected to experimentation. The image is converted to HSV colour space and is split into 5 rectangular regions as in Fig. 1(c). For each image region colour histogram feature is calculated and is multiplied with area coefficient to form a final fusion vector as in equation 3.

Feature fusion = (CH region 1 x a1) U (CH region 2 x a2) U (CH region 3 x a3) U (CH region 4 x a4) U (CH region 5 x a5) (3)

Where a1, a2, a3, a4 and a5 are area coefficients of each image region. CH is the colour histogram feature for each region.

The area coefficients are calculated as in equation 4:

Area coefficient = Area of the region/Total area of the image (4)

The area 5 is the centre big rectangle and will get more weightage due to its large area.

Image retrieval results from our proposed approach are tabulated in table 1. Using proposed CBIR method we get 70.79% average precision and 28.6% average recall rate. The results show that the new proposed CBIR method improves the average precision by 6.33% and average recall by 2.73% as compared to baseline CBIR. As depicted in Fig. 3, the proposed CBIR method outperforms baseline CBIR by improving precision across classes and number of images retrieved.

TABLE I: AVERAGE PRECISION RATES FOR IMAGE RETRIEVAL

Image Classes	**Average Precision Rates from various methods**			
	[Method 1] Baseline CBIR using colour histogram on entire image	**[Method 2] CBIR with colour histogram, 5 regions, 35% features but no area coeff.**	**[Method 3] Proposed CBIR method using area coeff.**	**Improvement in precision [Method 3 vs Method 1]**
Class 1 [Africa]	65.13 %	68.33%	71.80%	6.67%
Class 2 [Beach]	47.20 %	56.20%	55.47%	8.27%
Class 3 [Monuments]	47.40 %	23.67%	50.53%	3.13%
Class 4 [Buses]	72.87 %	49.47%	86.27%	13.40%
Class 5 [Dinosaurs]	99.27 %	98.67%	99.47%	0.20%
Class 6 [Elephants]	37.53 %	39.13%	40.33%	2.80%
Class 7 [Flowers]	73.07 %	60.33%	88.67%	15.60%
Class 8 [Horses]	92.73 %	89.67%	93.27%	0.53%
Class 9 [Mountains]	52.07 %	37.87%	54.00%	1.93%
Class 10 [Food]	57.40 %	47.20%	68.13%	10.73%

|| Volume 3 || Issue 10 || October 2018 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

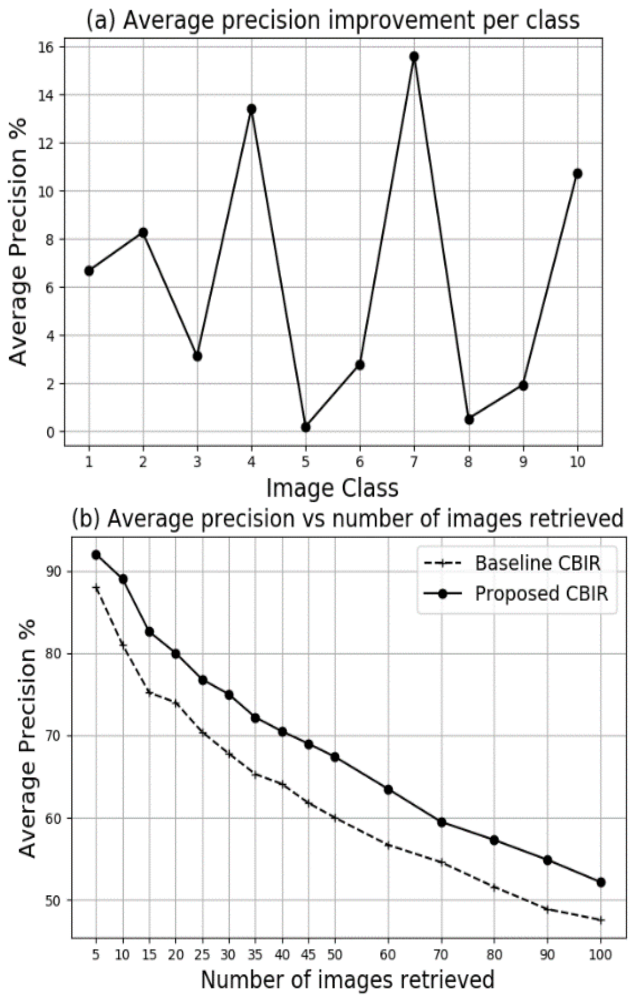


Figure. 3 (a) Average precision improvement per class (b) Average precision vs number of images retrieved for baseline CBIR and proposed CBIR.

V CONCLUSION

In this work, we presented a novel method to improve the precision of content based image retrieval using color histogram with area coefficients. We proposed a new image partitioning method to create 5 simple rectangles with a bigger center rectangle in order to give more importance to the center area of the image. Also, we proposed simple area coefficients to be used as weights for assigning required importance to image partitions. From experimental results we conclude that our proposed CBIR method (70.79% average precision and 28.6% average recall rate) outperforms baseline CBIR (64.47% average precision and 25.87% average recall rate) which uses color histogram feature over the entire image. Our proposed CBIR method improves the average precision by 6.33% and average recall by 2.73% as compared to baseline CBIR. The improvement in average precision is seen across all image classes and number of images retrieved. Results from method 2 shows that feature selection does not impact the average precision positively and proposed CBIR method is solely responsible for improvement in precision rates.

Future work: The current work was limited to using color histogram feature alone, other widely used features like color, texture, and shape features can be clubbed with color histogram. The optimal areas of the sub regions can be determined to improve the precision further

REFERENCES

[1] Yong Rui and Thomas S. Huang, “Image Retrieval: Current Techniques, Promising Directions and Open Issues”, Journal of Visual Communication and Image Representation, 1999.

[2] Wei-Ying Ma and HongJiang Zhang, “Content-Based Image Indexing and Retrieval”, Handbook of Multimedia Computing, CRC Press, 1999.

[3] Markus Andreas Stricker, Alexander Dimai, “Color indexing with weak spatial constraints”, Proc. SPIE 2670, Storage and Retrieval for Still Image and Video Databases IV, 1996.

[4] T. S. Chua, K.-L. Tan and B. C. Ooi, “Fast signature-based color-spatial image retrieval”, Proc. Of IEEE ICMCS, Ottawa, Canada, pp.362-369, 1997.

[5] C. Faloutsos, M. Flickner, W. Niblack, D. Petkovic, W. Equitz and R. Barber, “Efficient and effective querying by image content, Technical Report”, IBM Research Report, 1993.

[6] Malinga, B., Raicu, D., & Furst, J.D, “Local vs. Global Histogram-Based Color Image Clustering”, Technical report, University of Depaul, 2006.

[7] L. Junling, Z. HongWei, K. Degang and C. Chongxu.: Image retrieval based on weighted blocks and color feature, 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), Jilin, 2011, pp. 921-924.

[8] Huan Liu and R. Setiono, “Chi2: feature selection and discretization of numeric attributes”, Proceedings of 7th IEEE ICTAI, Herndon, VA, 1995, pp. 388-391.

[9] James Z.Wang Research Group. http://wang.ist.psu.edu/docs/home.shtml

[10] Y. Yang and J. O. Pedersen.: A comparative study on feature selection in text categorization, Mach. Learn. Work. Then Conf., pp. 412–420, 1997.

[11] E. Rachmawati, M. L. Khodra and I. Supriana, “Histogram based color pattern identification of multiclass fruit using feature selection”, ICEEI, Denpasar, 2015, pp. 43-48.

[12] Xiu-Qi Li, Shu-Ching Chen, Mei-Ling Shyu, and Borko Furht, “An Effective Content-Based Visual Image Retrieval System”, Proc. of 26th IEEE COMPSAC, Oxford, 2002.

[13] H.D. Cheng, X. H. Jiang, Jingli Wang, “Color Image Segmentation Based on Homogram Thresholding and Region Merging”, Pattern Recognition, 2001.