

Content Based Image Retrieval Using Gray Scale Weighted Average Method

Kamlesh Kumar

School of Computer Science & Engineering
University of Electronic Science & Technology of China
Chengdu, China

Jian-Ping Li

School of Computer Science & Engineering
University of Electronic Science & Technology of China
Chengdu, China

Zain-ul-abidin

School of Computer Science & Engineering
University of Electronic Science & Technology of China
Chengdu, China

Riaz Ahmed Shaikh

School of Computer Science & Engineering
University of Electronic Science & Technology of China
Chengdu, China

Abstract—High feature vector dimension quietly remained a curse element for Content Based Image Retrieval (CBIR) system which eventually degrades its efficiency while indexing similar images from database. This paper proposes CBIR system using Gray Scale Weighted Average technique for reducing the feature vector dimension. The proposed method is more suitable for color and texture image feature analysis as compared to color weighted average method as illustrated in literature review. To prove the effectiveness of retrieval system, two standard benchmark dataset namely, Wang and Amsterdam Library of Texture Images (A LOT) for color and texture have been selected to evaluate the system retrieval accuracies as well as efficiencies generated by each method. For the purpose of image similarity, Euclidean distance has been employed which matches query image feature vector with image database feature vectors. The experimental results generated by two methods showed that overall performance of the proposed method is relatively better in terms of average precision, average recall and its average retrieval time.

Keywords—Color Weighted Average Method; Gray Scale Weighted Average Method; Feature Extraction; Precision; Recall; CBIR

I. INTRODUCTION

Due to over increasing growth in digital images on the web their storage management as well as retrieval is becoming a challenging task. Traditional Text Based Image Retrieval (TBIR) system is an inadequate method to retrieve similar images from the databases, because it is difficult to extract visual features with the aid of textual information. To address this problem the Content Based Image Retrieval System (CBIR) was introduced in early 1990, since then this area has been most extensively researched for images. The CBIR aims to retrieve semantically relevant images from database on the basis of low level visual image features such as color, shape, texture and object location. However, content based image retrieval system also suffers from either feature extraction complexity or by high dimensional feature matching [1]. Because of these factors image retrieval system outcome decreases in terms of accuracy and at the same time system behaves poorly in response time. Until now many image

retrieval methods have been proposed by many authors where they have presented multilevel feature extraction using color, texture and shape to improve the accuracy of image retrieval process. And also various indexing and machine learning techniques have been implemented to reduce the feature vector dimension for improving the efficiency of system retrieval [2]. For instance, Tree Based Index (R Tree), Hash Based Index (LSH), Visual Based Index (BOW) [3], Principle Component Analysis (PCA), Kernel Independent Component Analysis (KPCA) and Multidimensional Scaling (MDA) [4]. However, reducing the feature vector dimension after features extraction certainly losses discriminative visual features from the image, which eventually affects the accuracy and efficiency of the retrieval system.

The goal of this paper is to introduce one dimensional color feature vector for image retrieval using gray scale weighted average method. The similar method has been used in [7] that take weighted average of RGB color image. The proposed method uses gray scale intensity image instead of color image because it occupies less space in memory and retains important features related to the image. The rest of present paper is arranged as follow: The literature review has been presented in Section II. The methodology of proposed method has been explained in Section III. Section IV, demonstrates the experimental results and discussion. The efficiency of each system on two different color and texture dataset has been evaluated in Section V.

The conclusion has been given in Section VI and future work is elaborated in Section VII.

II. LITERATURE REVIEW

Since last two decades, various color based image feature extraction schemes have been suggested by many researchers for CBIR. A simple histogram method for color image retrieval was proposed by Swain *et al.* [5], which counts total number of occurrence of each distinct color in the image through multi-dimensional feature vectors. Similarly Kekre *et al.* [6] suggested color average technique. In this paper, total average of row and column from each red, green and blue have been calculated to form a feature vector. However it also

supported multidimensional structure for image retrieval. Moreover Ali *et al.* [7] described a novel color image retrieval using RGB color weighted average method. In that work, firstly feature vector of corresponding red, green and blue color have been added together that forms one dimensional structure. Secondly histogram is generated to find unique pixel values and then using probability distribution function weighted average feature vector is formed. The generated feature vector has been used for image to image similarity measure from the database. Color image retrieval using visual weighted block was demonstrated by Wu *et al.* [8]. In this paper, authors have divided an image into block and generated histogram for it. After that each pixel was given a saliency score for finding the local pixel variation as to form separate visual weighted blocks. The obtained sub-blocks are concatenated for color image similarity. In other paper An and Le [9], presented color feature extraction from salient region with spatial layout. They have used color contrast method for silent region and spatial layout for each dominant color, which was obtained through binary map. Color based image retrieval was presented by Varish and Kumar [10]. Here probability histograms for each RGB plane have been generated and from them significant bins are obtained. The statistical values are calculated from each bin to describe the feature vectors. Color histogram and color correlogram approach for image retrieval was given by Soni and Mathai [11]. In this paper, HSV and RGB color space have been used for local and global color histogram analysis. And color correlogram is used for image spatial layout. Finally both color feature were combined together to form a feature vector. Image retrieval using fusion of cell color histogram (CCH) and color coherence vector (CCV) was presented by Salami and Boucheham [12], in their method classification of coherent and non-coherent pixels have been performed using CCV and then CCH was implemented for color distinction.

III. GRAY SCALE WEIGHTED AVERAGE METHOD

In this paper, gray scale weighted average method for content based image retrieval has been proposed and also it is compared with colorful image weighted average method. The proposed algorithm works as similar to color weighted average method [7]; except it takes grey scale intensity image for image retrieval. The gray scale image produces different shades of pixels with eight bit value, which range from 0 – 255 as opposite to color image, which assigns eight bit value to each red, green and blue planes separately. Similarly each plane has distinct value from 0 – 256. The advantage of using gray scale intensity image is that it simplifies amount of information than three dimensional color images and also takes less space in the memory [13]. The key information related to image in grey scale is not lost such as edges, regions, blobs etc. In Fig. 1 (a) and (b), RGB color fruit's image is converted into gray scale image as shown below. However one can still distinguish red and green apples from gray scale image.

The formula for converting RGB color image to gray scale image is given in equation 1.

$$(n, m) = \alpha I_{colour(n,m,r)} + \beta I_{colour(n,m,g)} + \gamma I_{colour(n,m,b)} \quad (1)$$



Fig. 1. (a) RGB color image (b) Grey scale image

In this paper, grey scale intensity image is used for image feature analysis as its algorithm is described in Table I below.

TABLE I. ALGORITHM FOR GREY SCALE WEIGHTED AVERAGE METHOD

The proposed weighted average algorithm has been divided into six steps, which are explained below.

1. Convert RGB color image into Gray scale Image.
2. Generate image histogram and calculate the sum of the occurrence of all unique gray shades from this array.

$$\text{total} = \sum_{g=0}^N \text{his}(g) \quad (2)$$

3. Now apply probability distribution function for finding the occurrence of individual unique value in the gray scale image array.

$$\text{prob}(g) = \frac{\text{his}(g)}{\text{total}} \text{ where } 0 \leq g \leq N \quad (3)$$

4. Then calculate weighted average from unique values and their corresponding probabilities.

$$\text{weighted average} = \sum_{g=1}^N g \cdot \text{prob}(g) \quad (4)$$

5. Finally generate the feature vector value for gray scale image.

$$\text{Grayscale}_{\text{image}} \quad f_v = (g) \quad (5)$$

6. Compute Euclidean distance between the query image feature vector and image feature vector in database for retrieving most similar images. It can be calculated using following formula.

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^n (q_i - d_i)^2} \quad (6)$$

Here, his(g) denotes the histogram of gray scale image and N represents the total number of unique values in the gray scale image. The prob(g) shows the probability of each unique value in the gray image array and fv represents obtained feature vector. The q_i and d_i measures distance between query image and images in the databases.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental works for proposed method and color weighted average method have been generated using MATLAB. The image retrieval testing has been carried out on two standard benchmark datasets, one is Wang [14] database and another is ALOT [15]. These databases contain 500 and 700 images respectively; we have selected five different classes from each of them. Although pre-processing is implemented for resizing all database images into 162x162 resolutions. The performance of retrieval system for proposed and color average weighted method have been evaluated through average precision, average recall and its retrieval time.

precision: it is defined as how many like images are retrieved by the system as similar to the query image.

recall: it represents how many images are exactly recognized as similar to the query image.

retrieval time: it calculates total time taken by the system during image retrieval.

For image retrieval, threshold is used on each image dataset to retrieve top twenty most relevant images.

A. Image Retrieval Testing on Wang_Dataset

In this section, Wang data_set has been used for simulation. Here image retrieval testing has been demonstrated for proposed method and color average weighted method. Fig. 2 shows that a query of bus image is input to system using proposed method. The system has retrieved 12 relevant images out of 20 images. The system generated precision for proposed method is 60% and recall and retrieval time are 12%, 0.85 seconds respectively.



Fig. 2. Image retrieval using proposed method

The similar image in Fig. 3 is given as input to retrieval system using color average weighted method. The system has retrieved only 8 images as similar to the query while 12 images are irrelevant. The precision and recall obtained using color average weighted method are 40% and 8% respectively.

These are considerably lower as compared to proposed method.



Fig. 3. Image retrieval using color average weighted method

The color average weighted method retrieved images in 0.98 seconds. The performance of proposed method is quite high than color average weighted method on similar dataset.

B. Image Retrieval Testing on ALOT Data_Set

In this section, another experiment has been performed for further investigating the performance of proposed method. Here ALOT image dataset has been used to test retrieval query results. The obtained accuracy and efficiency of system has been compared with color average weighted method. In Fig. 4 below, coins texture image is input to the system using proposed method.

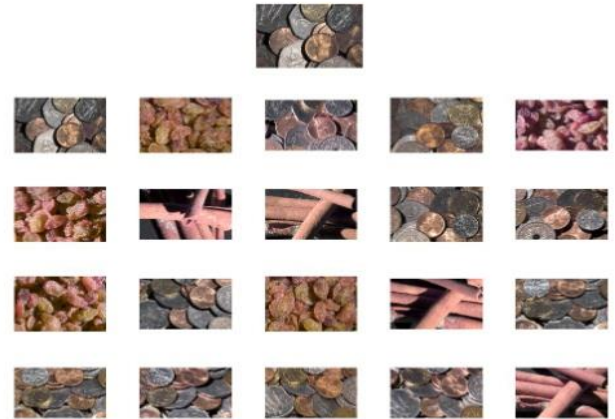


Fig. 4. Image retrieval using proposed method

The generated precision and recall rates using proposed method are 55%, 11% respectively i.e. retrieval system has retrieved 20 images out of them 11 are like matches as query image. The retrieval time it takes is 1.10 seconds.

Another experiment has been performed using color average weighted method. Fig. 5 depicts that similar image is input as reference to retrieval system. It has retrieved 10 images as similar to query image and rest of 10 are irrelevant. The system precision and recall rate are 50% and 10% respectively. It takes 1.26 seconds during image retrieval.



Fig. 5. Image retrieval using color average weighted method

V. RETRIEVAL SYSTEM PERFORMANCE ANALYSIS ON COLOR AND TEXTURE IMAGE DATASET

This section evaluates the performance of retrieval system using proposed method and color average weighted method. For this purpose, five different query image classes from each Wang and ALOT database are randomly selected for testing image retrieval system as shown in Fig. 6 and Fig. 7. It can be observed from Table II that average precision, average recall obtained by proposed method is quite higher than that of color average weighted method for Wang database images these are enriched with color features. To further validate the effectiveness of proposed method, it has been tested on ALOT database, which contains all images that have high concentration of textural features. The output generated by proposed method for texture images in terms of average precision, average recall are also shown increased in Table II as compared to color average weighted method. However retrieval time taken by proposed method for color and texture images database separately are 0.85 seconds and 1.17 seconds respectively. Similarly color average weighted method takes 0.96 seconds and 1.25 seconds for color and texture database images retrieval respectively. The advantage of proposed method is that it improved the accuracy without affecting retrieval system efficiency.

Fig. 8 and Fig. 9 are showing generated average precision and average recall for query results buses, dinosaurs, roses, horses and natures images. The proposed method has retrieved more like images as compared to color average weighted method, except for dinosaurs query image, where both have similar retrieval accuracies.

In Fig. 10 average retrieval time of system for Wang color images dataset is approximately equal for proposed and color average weighted method that is 0.85 seconds and 0.96 seconds respectively. The average precision and recall in Fig. 11 and Fig. 12 shows that accuracy of proposed method for ALOT textural images. This is quite improved for raisins and twined reed query images. Although for query fur, cinnamon sticks and coins images both retrieval systems have generated almost similar results. Fig.13 shows that computed retrieval time of system for ALOT textural image dataset by proposed method and color average weighted method, which are 1.17 seconds and 1.25 seconds respectively, which is almost similar.



Fig. 6. Wang color image data_set (image from [14])

TABLE II. PERFORMANCE ANALYSIS

Images	Performance	Method		
		Color Weighted Average Method	Proposed Method	Remarks
Wang Dataset	Average Retrieval Time (Seconds)	0.96	0.85	Almost Equal
	Average Precision (%)	57	70	Big Improvement
	Average Recall (%)	11	14	Big Improvement
ALOT Dataset	Average Retrieval Time (Seconds)	1.25	1.17	Almost Equal
	Average Precision (%)	53	61	Big Improvement
	Average Recall (%)	10	12	Big Improvement



Fig. 7. ALOT image data_set (image from [15])

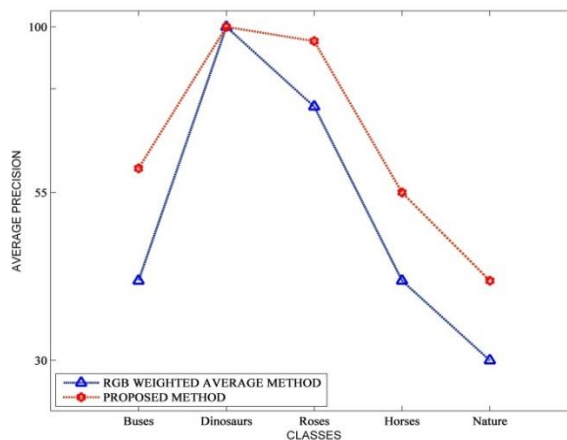


Fig. 8. Average precision of proposed and rgb weighted average method for wang data_set

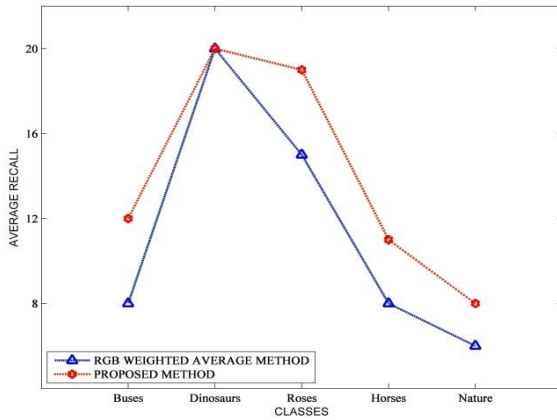


Fig. 9. Average recall of proposed and rgb weighted average method for wang data_set

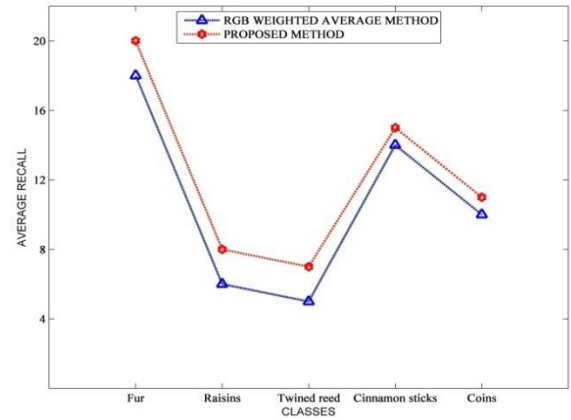


Fig. 12. Average recall of proposed and rgb weighted average method for alot data_set

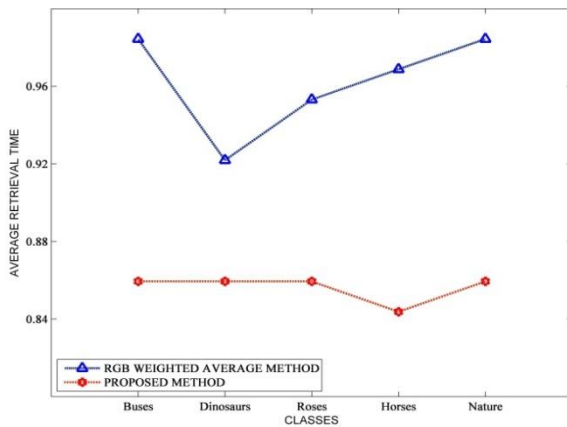


Fig. 10. Average retrieval time of proposed and rgb weighted average method for wang data_set

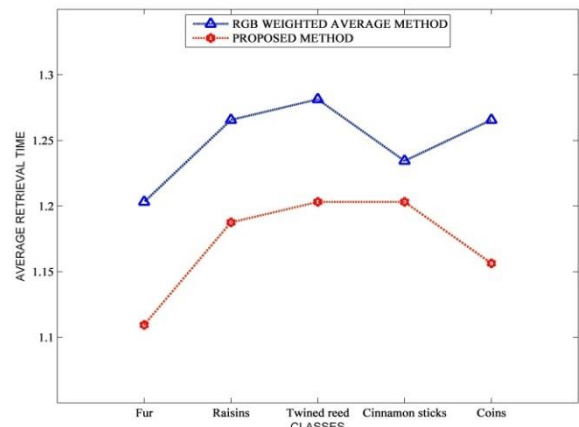


Fig. 13. Average retrieval time of proposed and rgb weighted average method for alot data_set

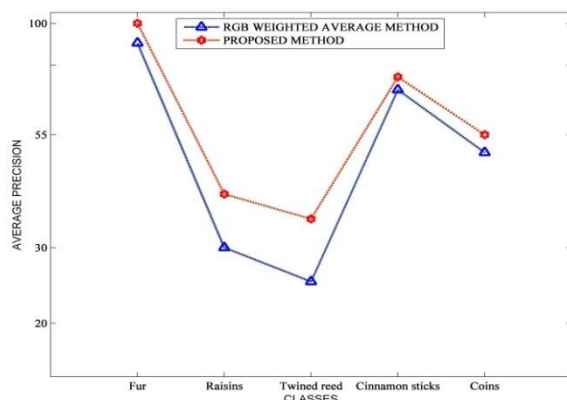


Fig. 11. Average precision of proposed and rgb weighted average method for alot data_set

From these above figures it can be concluded that proposed method is more efficient for color and texture image analysis than color average weighted method. It has improved the accuracy without affecting the efficiency of the CBIR system.

VI. CONCLUSION

Image feature extraction remains a big dilemma for researches in the context of image retrieval since decades, which directly affects the performance of retrieval system.

In this paper, gray scale weighted average method was proposed to reduce the feature vector dimension for increasing the overall throughput of image retrieval system. The performance of CBIR system using proposed approach is quite better than color weighted average method.

VII. FUTURE WORK

We proposed content based image retrieval using gray scale weighted average method for reducing the feature vector dimension as well as for improving the efficiency of retrieval system. The effectiveness of the proposed method was evaluated on two color and texture image datasets where obtained results showed the usefulness of our image retrieval system. Although proposed method works well on texture and color images but still this algorithm can be improved such that it could able to identify color and texture images from hybrid database and then produce refine results on the basis of user preferences. For this purpose, in future we plan to incorporate image clustering scheme.

ACKNOWLEDGMENTS

This paper was supported by the National Natural Science Foundation of China (Grant No.61370073). The National High-Technology Research & Development Program of China (Grant No.2007AA01Z423).

REFERENCES

- [1] Raimondo S., Gianluigi C., Isabella G., "Feature extraction for content-based image retrieval", Encyclopedia of Database Systems, Springer, USA, pp. 1115-1119, 2009.
- [2] LF. Ai, J.Q. Yun, Y.F HE, T. Guan " High dimensional indexing technologies for large scale content- based image retrieval": a review: Journal of Zhejiang University-Science C (Computers and Electronics), vol. 14(7), pp. 505-520, May, 2013.
- [3] B. Celia, I. Felci Rajam, "An efficient content based image retrieval framework using machine learning techniques" Data Engineering and Management, Springer Berlin Heidelberg, pp. 162-169, 2012.
- [4] S. Alireza, "An actual survey of dimensionality reduction", American Journal of Computational Mathematics, vol. 4, pp. 55-72, Dec. 2013.
- [5] M.J Swain and D. H., Ballard, "Color Indexing," International Journal of Computer Vision, pp. 11-32, 1991.
- [6] H.B. Kekre "Color Feature Extraction for CBIR," International Journal of Engineering Science and Technology, vol. 3(12), pp. 8357-8365, 2011.
- [7] M. Ali, L. Dong, Y. Liang, Z. Xu, L. He and N. Feng, "A color image retrieval system based on weighted average", International Conference on Signal Processing, Communication and Computing, pp. 184-189, Aug. 2014.
- [8] J. Wu, Z. Li, C. Ji, Y. Li, X. Xing, " Image retrieval using visual weighted blocks", International Conference on Dependable, Autonomic and Secure Computing, pp. 338-343, 2014.
- [9] Jaehyun An, S.H Lee, and N.I Cho, "Content based image retrieval using color features of silent regions", International Conference on Image Processing, pp. 3042-3046, 2014.
- [10] V. Naushad, A.K Pal, " Content based image retrieval using statistical features of color histogram", Internal Conference on Signal Processing, Communication and Networking, pp. 1-6, 2015.
- [11] S. Devyani, K.J. Mathai, " An efficient content based image retrieval system based on color space Approach using color histogram and color correlogram", International Conference on Communication Systems and Network Technologies, pp. 489-493, 2015.
- [12] M. Salami, B. Boucheham, "Content based image retrieval on cell color coherence vector", International Symposium on ISKO-Maghreb: Concepts and Tools for Knowledge Management, pp. 1-5, 2014.
- [13] Chris Solomon, Toby Breckon, "Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab, Book: Publisher Wiley Blackwell, pp. 11-12. 2011.
- [14] Wang data_set. A subset of corel database images [online]. <http://savvash.blogspot.com/2008/12/benchmark-databases-for-cbir.html>
- [15] Amsterdam Library of Digital Images [online]. http://aloi.science.uva.nl/public_aloi/