Journal of Theoretical and Applied Information Technology
15th December 2011. Vol. 34 No.1

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

MEAN AND STANDARD DEVIATION FEATURES OF COLOR HISTOGRAMUSING LAPLACIAN FILTER FOR CONTENT-BASED IMAGE RETRIEVAL

FAZAL-E-MALIK
Baharum Bin Baharudin
Department of computer and Information Sciences
UniversitiTeknologi PETRONAS
Malaysia
E-mail: fazal_i_malik@yahoo.com, baharbh@petronas.com.my

ABSTRACT

Due to the development and improvement in internet with high speed for the last few years andthe availability of a large digital image collection, efficient image retrieval systems are required. Retrievalcan be text-based and content-based. In content-based, which is also called Content-Based Image retrieval(CBIR), low level features of images are extracted such as color, texture and shape. These features areused in similarity measurement to retrieve relevant images from an image database. A CBIR algorithm isproposed which is based on the color histogram using Laplacian filter to reduce the noise and provides anenhanced image with more detail informations. Color histogram of the filtered image is divided into bins.Mean and standard deviation are calculated for pixels in each bin to get feature vector which is used forimage retrieval. After various experiments with user queries, results show the good retrieval of images byalgorithm.

Keywords: *Content-Based Image Retrieval (CBIR), Feature Vector, LaplacianFilter, Histogram.*

1. INTRODUCTION:

The huge collection of digital images are collected due to the improvement in the digital storage media,image capturing devices like scanners, web cameras, digital cameras and rapid development in internet. Thisleads to rapid and efficient retrieval of these images for visual information in different fields of life likemedical, medicine, art, architecture, education, crime preventions etc. to achieve this purpose many imageretrieval systems have been developed. In 1970's the first approach for searching of images in image collectionwas text-based in which manually annotated images are retrieving by key words. Examples are Google and Yahoo. But this approach has two drawbacks, first is to annotate huge number of images, requires a lot ofhuman labor and second is the different subjective perceptions of human for example Lilly flower can beannotated as water lilies, flowers in pond etc. Due to these disadvantages, in 1980's another approach emergedcalled content based image retrieval (CBIR)[1].

CBIR retrieves images by their visual contents such as color, shape and texture instead of annotated textmethod [2]. CBIR systems have been developed which include some commercial systems, some productionsystems, some research systems and some demonstration systems such as QBIC, ADL, BDLP, Virage,AltaVista, SIMPLIcity, etc., a detail survey can be found in [3].

Texture and color features are fused to retrieve the relevant images in CBIR from the image database. Byusing histogram technique, color features of image are computed. by using statistical measurements entropy,smoothness and uniformity in histogram gives texture feature[4].

In this paper a CBIR algorithm is proposed which is based on the color histogram using Laplacian filter toreduce the noise and provides an enhanced image with more detail information. Color histogram is divided intobins. Mean and standard deviation are calculated for pixels in each bin to get feature vector which is used forimage retrieval. The rest of the paper is organized as such that section 2 discusses related works, section 3describes the methodology in detail, section 4 evaluates the image retrieval experimental results and section 5concludes this paper.while for the local features genetic algorithm (GA) is used in HSV color space. Accuracyof this method is high as compared to the previous method [10].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

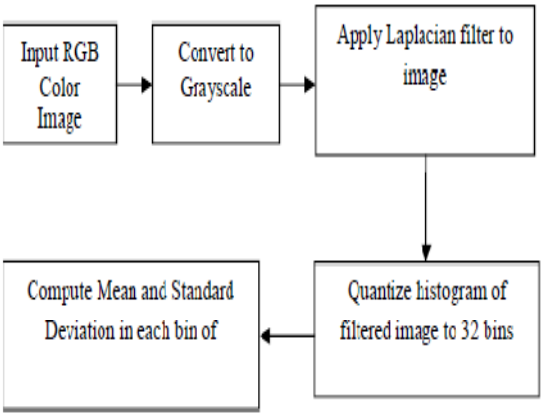

Figure 1. Block diagram for the feature extraction algorithm process of algorithm.

2. RELATED WORKS:

For the last decade many methods and algorithms have been developed for CBIR.A comprehensive review about CBIR of 200 references is given in [5]. This paper discusses the detail workingstatus of CBIR like image type, semantic role, semantic gap, and computation of feature extraction, similarity offeatures, image retrieval and relevant feedback for enhancement of systems.

Color histogram is very prominent technique for feature extraction. It is mostly used for CBIR. In this paperColor histogram based method is proposed, in which color and shape features are used. Also a new set offeatures such as size, mean, variance of objects are used for retrieval [6]. Analysis of features using color histogram, based on Edge extraction and Median filter to reduce noise andkeep the original edge information. Feature vector is created by taking average of pixels in each bin to retrieveimages [2].

Color and texture features are extracted by using color histogram and Gabor wavelet transform techniques [7].

Color, shape and texture features are considered for CBIR. Gabor filter is used to get regions of interest(ROIs). In each ROI texture is calculated by using Gabor features, color by using histogram and color moments,shape by using Zermikes moments [8].

Single region is better than whole image as a query example for retrieval of images and SVM is used forclassification [9].

The CBIR will be efficient and effective if the algorithm used is fast in computation of feature extraction andaccurate in the result after similarity calculation. In this paper a novel method is proposed in which color andtexture information are used for retrieval of images. Global and local color features are extracted. To extractglobal color features, histogram method is used in RGB color space while for the local features geneticalgorithm (GA) is used while for the local features genetic algorithm (GA) is used in HSV color space.

Accuracy of this method is high as compared to the previous method [10].

3. PROPOSED ALGORITHM:

Algorithm is based on color histogram. Statistical measurements Mean and standard deviation are computedin histogram. Before applying histogram, Laplacian filter is applied for noise removal. The block diagram ofalgorithm is show in Figure 1.

Method consists of the following steps

1. The input RGB image is acquired and converted into grayscale image.
2. Apply Laplacian filter to grayscale image.
   The Laplacian filter uses a mask w of 3*3 with -4 at the center. Let f is an histogram equalized image and gis the filtered image. During filtering some information are lost. To restore lost information, Laplacianimage is subtracting from original image histogram equalized image such that g1=f-g, to get g1 enhancedimage [11]. The filtration process is shown in Figure 2.
3. Quantize the filtered image into 32 bins.
   Quantization is a process in which the histogram is divided into levels or bins. As grayscale image consistsof 256 levels. Computations for the feature extraction in these 256 levels will be slow. To increase thespeed of computations, the histogram of image is reduced to 32 bins [6].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

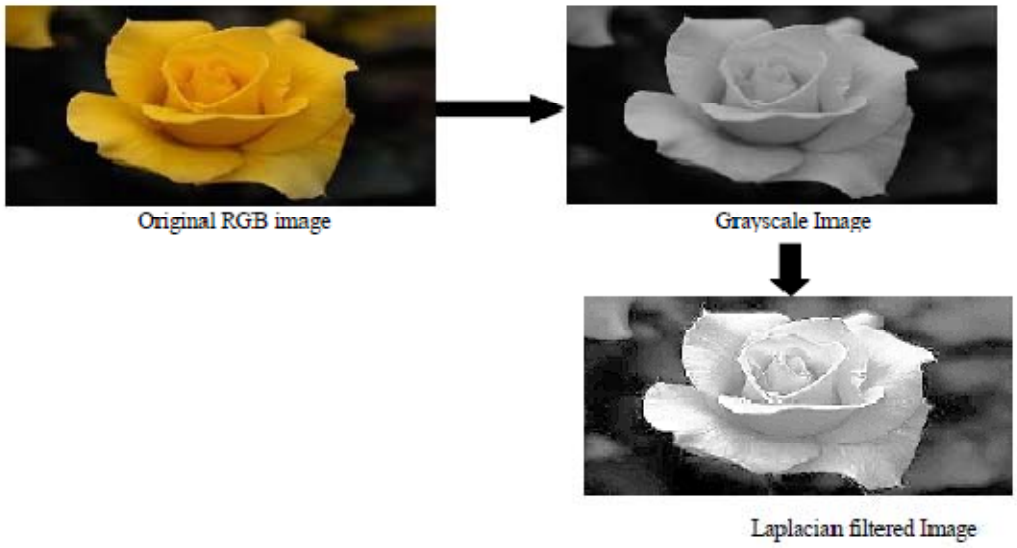

Figure 2.Filtration process of algorithm.

4. Calculate the mean and standard deviation of pixels in each bin.
   Each bin consists of a range of some pixelsvalues. These values in each bin can be used to calculate the mean of bin which represents the somebrightness of the image in that bin. If mean of a bin is high then it means that the image is bright in thatbin and if mean is low then it means that the image is dark in that bin. The standard deviation in each is alsocalculated by using the mean and pixel values of each bin. The standard deviation reveals something aboutthe contrast of image in particular bin. If standard deviation is high then it shows the high contrast of imagein a particular bin. If standard deviation is low then it will show the low contrast in image of a particularlevel of histogram [14].

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji} \quad (1)$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ji}-\mu_j)^2} \quad (2)$$

   As histogram is divided into 32 bins then it means that 32 features will be computed for mean and 32features for standard deviation. These features will be combined to get a feature vector of total 32+32=64features. Let μ_j is the mean and σ_j is the standard deviation of a particular bin j, where j=1, 2, 3…, 32 arebins. These features can be calculated by using the statistical measurements [15, 16] as under

   Where xji is the pixel values in bin and N is total number of pixels in each bin.
   A feature vector is given by

$$fv = \{\mu_1, \mu_2, \ldots, \mu_{32}, \sigma_1, \sigma_2, \ldots, \sigma_{32}\} \quad (3)$$

   This feature vector fv will be calculated for all images in the collection of images and will be stored in adatabase to be retrieved by query image. In the same way this feature vector fv will also be calculated forquery image by using same procedure.

5. A query image will be acquired from the user as an example to retrieve similar images from the database byusing the extracted features.
6. The similarity measurement will be performed for the matching of query image with database images. Forthis purpose the sum-of-absolute difference (SAD)[15] between the extracted query feature vector Fq anddatabase feature vector Fp will be calculated for all n=64 features, as given be equation.

$$\Delta S = \sum_{i=1}^{n} \left| (\Gamma_p(i) - \Gamma_q(i)) \right| \quad \text{where } i=1, 2\ldots, n \quad (4)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 3. Query result for Elephants

If ΔS=0 then both images are same and if small the images are relevant to the query image.
ΔS is the difference between query image and database image and this is calculated for all image indatabase and arrange in ascending order so that the smallest values will be on top which represent the mostrelevant or similar images and irrelevant will at bottom. The top most images are displayed to the userwhich is the required images.

4. RESULTS

The proposed algorithm is tested by the database of images provided by James Wang et al [12, 13] which isfreely available for researchers. The database consists of 1000 images having 10 categories of people, beach,building, bus, dinosaur, elephant, rose, horse, mountain, and food.
First all images are acquired one after another for feature extraction by algorithm and stored in database withfeatures vectors. The Laplacian filter is applied for removal of noise and to enhance the image, then histogram isapplied on filtered images for extraction of features. Various experiments were performed for the 10 categoriesin which the user was asked to select a query image and relevant images were displayed to the user. A set of 200queries were applied for all categories. Results obtained as shown in Table I.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4. Query result for roses.Figure 5. Query result for foods.

Figure 6. Query result for horses. Figure 7. Query result for people.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table I shows the average precision and recall of all categories and over all average of all categories. It canbeseen that the algorithm gives good results for dinosaurs.

S.No	Categories	Precision %	Recall %
1	People	36.67	72.20
2	Beaches	40.55	45.54
3	Buildings	19.44	42.58
4	Buses	27.22	58.01
5	Dinosaurs	59.44	82.14
6	Elephants	22.78	56.67
7	Roses	43.33	74.05
8	Horses	24.44	44.41
9	Mountains	30.55	49.79
10	Foods	22.78	55.83
Average		32.72	58.12

Figures 3-8 show the result of user queries. Each Figure consists of a query image and the retrieved images from the database by using the proposed algorithm. The top single image is the query image and below 9 are therelevant images. The results show that proposed algorithm has good retrieval accuracy.

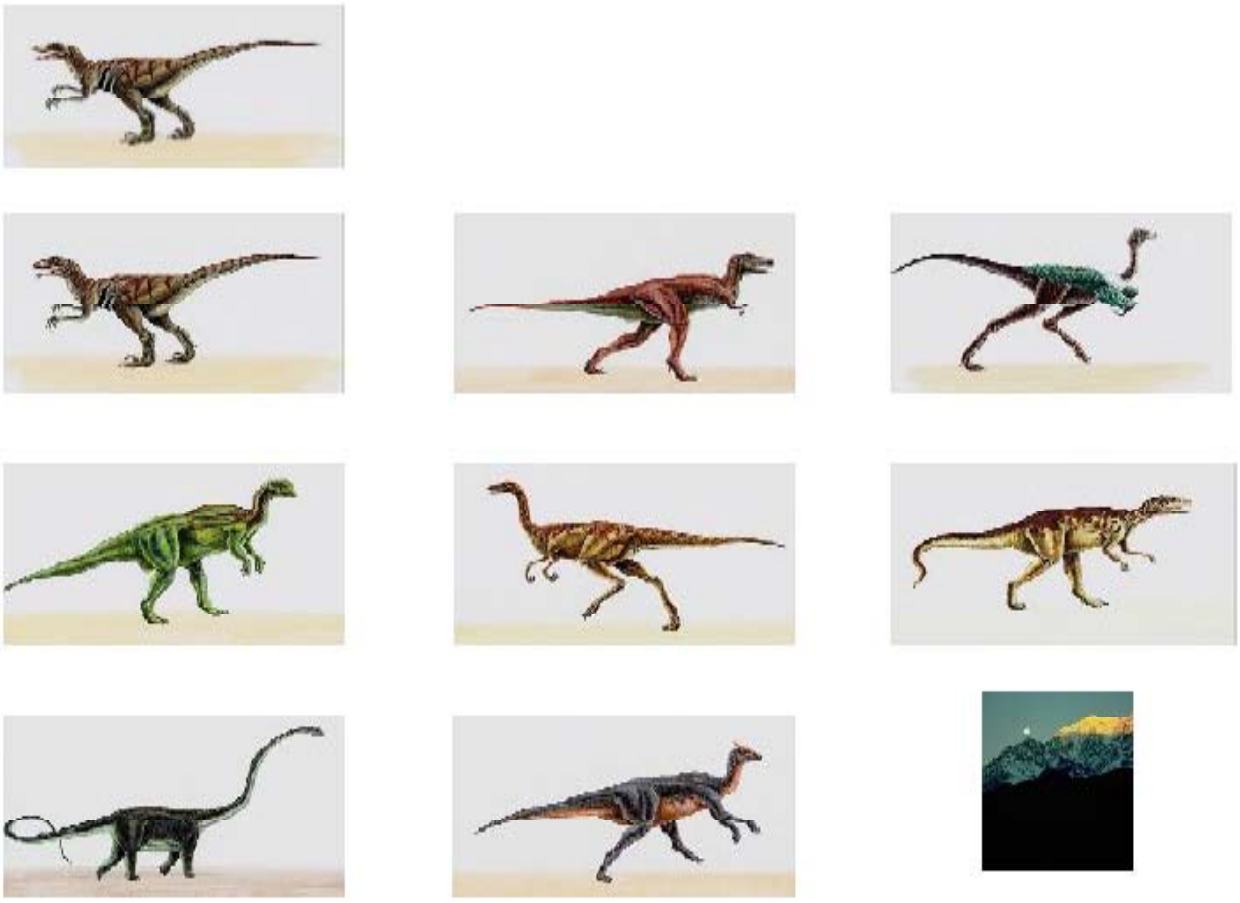

Figure 8. Query result for dinosaurs

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. CONCLUSION AND FUTURE WORK:

In this paper an algorithm is proposed which is based on color histogram and Laplacian filter to get moreenhanced image. The histogram is applied to filtered image to get features in image. The mean and standarddeviation of pixels in bins of histogram image are calculated to get feature vector. Mean represents brightnessand standard deviation represents contrast of image. Algorithm tested with various queries and good resultsobtained specially for roses and dinosaurs. Hence this approach provides good result for CBIR.

We will improve the accuracy further of this method in our future and also the computational speed offeatures extraction.

REFERENCES

[1] Liu, Y, D Zhang, G Lu, and W Ma. "A surveyof content-based image retrieval with high-level semantics."*Pattern Recognition* 40,no. 1 (January 2007): 262-282. doi:10.1016/j.patcog.2006.04.045.

[2] Zhao, Hui, Pankoo Kim, and Jongan Park. "Feature analysis based on Edge Extraction and Median Filtering for CBIR." *Work*(2009):245-249. doi:10.1109/UKSIM.2009.40.

[3] Veltkamp, Remco C, and MirelaTanase. "Content-Based Image Retrieval Systems: A Survey." *October* (2002).

[4] Thawari, P B, and N J Janwe. "CBIR BASED ON COLOR AND TEXTURE." International Journal of Information Technology andKnowledge Management January-June 2011, Volume 4, No. 1, pp. 129-132

[5] Smeulders, a.W.M., M. Worring, S. Santini, a. Gupta, and R. Jain. "Content-based image retrieval at the end of the early years."*IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 12 (2000): 1349-1380. doi:10.1109/34.895972.

[6] J. Park, Y An, G. Kang, W. Rasheed, S. Park, G. Kwon, "Defining a New Feature Set for Content-Based I amge Analysis UsingHistogram Refinement " 2008 Wiley Periodicals.Inc

[7] Murala, Subrahmanyam, Anil BalajiGonde, and R. P. Maheshwari. "Color and Texture Features for Image Indexing and Retrieval."*2009 IEEE International Advance Computing Conference*, no. March (March 2009): 1411-1416. doi:10.1109/IADCC.2009.4809223.

[8] Choraś, Ryszard S., Tomasz Andrysiak, and MichałChoraś. "Integrated color, texture and shape information for content-basedimage retrieval."*Pattern Analysis and Applications* 10, no. 4 (April 2007): 333-343. doi:10.1007/s10044-007-0071-0.

[9] D.N.F. AwangIskandar,Thom, James A, and S M MTahaghoghi. "Content-based Image Retrieval Using Image Regions as QueryExamples." *Database* 75 (2008): 38-48.

[10] Park, Sang-Sung, Young-Geun Shin, and Dong-Sik Jang. "A novel efficient technique for extracting valid feature information."*Expert Systems with Applications* 37, no. 3 (March 2010): 2654-2660. doi:10.1016/j.eswa.2009.08.013.http://linkinghub.elsevier.com/retrieve/pii/S0957417409007660.

[11] R.C.Gonzalez, R.E. Woods, S. L. Eddins, "Digital Image Processing using MATLAG", 2nd Edition, Pearrson Prentice Hall 2004,Book, page 100-103.

[12] J. Li and J. Z. Wang, "Automatic Linguistic Indexing of Pictures by a statistical modeling approach," Pattern Analysis and MachineIntelligence, IEEE Transactions on, vol. 25, pp. 1075-1088, 2003.

[13] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching Libraries," Advances in Visual forPicture Information Systems: 4th International Conference, VISUAL 2000, Lyon, France, November 2-4, 2000: Proceedings, 2000.

[14] Sergyan, Szabolcs. "Color histogram features based image classification in content-based image retrieval systems." *2008 6thInternational Symposium on Applied Machine Intelligence and Informatics* (January 2008): 221-224.doi:10.1109/SAMI.2008.4469170.

[15] Kodituwakku, S R, and S Selvarajah. "Comparison of Color Features for Image Retrieval." Indian Journal of Computer Science andEngineering Vol. 1 No. 3 207-211.

[16] Bannour, Hichem, LobnaHlaoua, and BechirAyeb. "Survey of the Adequate Descriptor for Content-Based Image Retrieval on theWeb : Global versus Local Features." *Recherche*(2009): 445-456.