Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 8, August 2015, pg.322 – 329

RESEARCH ARTICLE

Content Based Image Retrieval using Haar Wavelet to Extracted Color Histogram and Texture Features

Yusra .T. Mshari, Hameed A. Younis
Computer Science, University of Basrah, Iraq
Email: yusra.alzidii@gmail.com, hameedalkinani2004@yahoo.com

Abstract

The Content Based Image Retrieval has been an active research area. Given a collection of images, it is to retrieve the images based on a query image, which is specified by content. In this paper, wavelet transform has been used, which is proved to be a very useful tool for image processing in recent years. It allows a function which may be described in terms of details that range from broad to narrow. We use Haar wavelet transformation for feature extraction of the given image. The database image features are extracted by discrete wavelet transform and multiwavelet based features at different levels of decompositions. We have tested 1000 images with 10 different categories. Swain &Ballard Distance and Euclidean Distance are used as similarity measure. The experimental results are performed for Wang's database show the better results in terms of retrieve accuracy and computation complexity.

Keywords*:* Color Histogram, Image retrieval, discrete wavelet transform, multiwavelet.

1. Introduction

Content-Based Image Retrieval (CBIR) is the process of retrieving images from a database on the basis of features that are extracted automatically from the images themselves [1]. A CBIR method typically converts an image into a feature vector representation and matches with the images in the database to find out the most similar images. In the last few years, several research groups have been investigating content based image retrieval.

A popular approach is querying by example and computing relevance based on visual similarity using low-level image features like color histograms, textures and shapes. Text-based image retrieval can be traced back to the 1970‟s; images were represented by textual descriptions and subsequently retrieved using a text-based database management system [2].

Content-based image retrieval utilizes representations of features that are automatically extracted from the images themselves. Most of the current CBIR systems allow for querying-by example, a technique wherein an image (or part of an image) is selected by the user as the query. The system extracts the features of the query image, searches through the database for images with similar features, and displays relevant images to the user in order of similarity to the query [3].

Image Retrieval aims to provide an effective and efficient tool for managing large image databases. With the ever growing volume of digital image generated, stored, accessed and analyzed.

2. Algorithms of Proposed Method

In this paper, we proposed two algorithms for extracting features of given images. One for extracting color features using Haar discrete wavelet transform (DWT), and another for extracting texture features.

2.1 Algorithm of Color Histogram using Haar DWT

Algorithm (1) describes the main steps of color features of a given image by using Haar DWT.

Algorithm (1): Color Features Extraction

Step1. Read image.

Step2. Image resize to all image in database.

Step3. Decompose color image using Haar DWT at 1st level to get approximate coefficient and culpa detail coefficients.

Step4. Assign the weights 0.003 to approximate coefficients.

Step5. Convert the approximate coefficient image in to HSV plane.

Step6. Color quantization is carried out using color histogram by assigning 18 bins to hue, and 3 bins to saturation and 4 bins to value to give a quantized HSV space with 18+3+4=25 histogram bins.

Step7. Repeat step1 to step6 on an image in the database.

Step8. Calculate the similarity matrix of query image and the image present in the database.

Step9. Repeat the steps from 7 to 8 for all the images in the database.

Step10. Retrieve the images.

2. 2. Algorithm of Multiwavelet to Extract Texture

The algorithm of extracting texture features using multilevel is illustrated in Algorithm (2).

Algorithm (2): Texture Features Extraction

Step1. Read image.

Step 2.image resize to all image in database.

Step3. Decompose Color image using Haar Wavelet transformation at 1st level ((wavedec2) multilevel 2-D wavelet decomposition).

Step4. Take Energy for 2-D wavelet decomposition, returns Ea, which is the percentage of energy corresponding to the approximation, and vectors Eh, Ev, Ed, which contain the percentages of energy corresponding to the horizontal, vertical, and diagonal details, respectively.

Step5. Repeat the steps from 3 to 4 for three levels.
Step6. Repeat step1 to step 5 on an image in the database.
Step7. Calculate the similarity matrix of query image and the image present in the database.
Step8. Repeat the steps from 6 to 7 for all the images in the database.
Step9. Retrieve the images.

3. The Wavelet Transform

The Haar transform of x is given by the set of "difference" values d_i^l $(0 < l \le L, 0 \le i < 2^{l-1})$,and the "average" value for the last levela_0^L. In the frequency domain, the valuesa_i^l correspond to the output of a low pass filter, thus representing low-frequency information, whereas thed_i^lvalues correspond to the output of a high pass filter, thus representing high-frequency information.

In our case, the signal is a 2-D color image, where the "time" domain is the spatial location of pixels and the "frequency" domain is the color variation between adjacent pixels.

For an N x M image, the first transformation step decomposes the signal into four sub-images of size N/2 x M/2, representing the sub-bands in the frequency domain. The obtained sub-images are labeled as LL; LH ; H L; H H, where L and H represent low- and high-frequency information, respectively, and the first and the second position refer to the horizontal and the vertical direction, respectively. The second transformation level decomposes the LL sub-image, obtaining four images of size N/4 x M/4, and so on. Figure (1) shows the decomposition of the frequency domain at different scale levels:

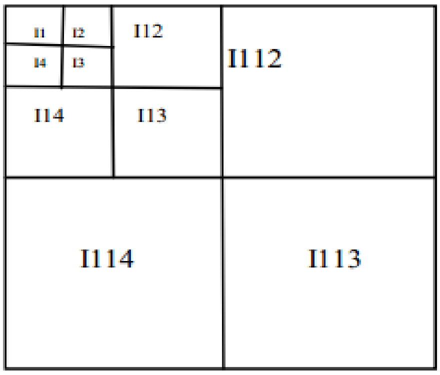

Figure (1): After applying Haar wavelet transformation for 3 iterations

Multiwavelets were defined using several wavelets with several scaling functions [4]. Multiwavelets have several advantages in comparison with scalar wavelet [5]. A scalar wavelet cannot possess all these properties at the same time. On the other hand, a multiwavelet system can simultaneously provide perfect representation while preserving length (Orthogonality), good performance at the boundaries (via linear phase symmetry), and a high order of approximation (vanishing moments) [6]. Thus multiwavelets offer the possibility of superior performance and high degree of freedom for image processing applications, compared with scalar wavelets. During a single level of decomposition using a scalar wavelet transform, the 2- D image data is replaced by four blocks corresponding to the sub bands representing either low pass or high pass in both dimensions. These sub bands are illustrated in Figure(2).

The multi-wavelets used here have two channels, so there will be two sets of scaling coefficients and two sets of wavelet coefficients. Since multiple iteration over the low pass data is desired, the scaling coefficients for the two channels are stored together.

Likewise, the wavelet coefficients for the two channels are also stored together.

L_1L_1	L_1L_2	L_1H_1	L_1H_2
L_2L_1	L_2L_2	L_2H_1	L_2H_2
H_1L_1	H_1L_2	H_1H_1	H_1H_2
H_2L_1	H_2L_2	H_2L_1	H_2L_2

Figure (2): Image decomposition after a single levelscaling for multiwavelets

4. Color Histogram and Texture Feature

A color histogram represents the distribution of colors in an image, through a set of bins, where each histogram bin corresponds to a color in the quantized color space. A color histogram for a given image is represented by a vector in Equation (1):

$$\boldsymbol{H} = \{\boldsymbol{H}[\mathbf{0}], \boldsymbol{H}[\mathbf{1}], \boldsymbol{H}[\mathbf{2}], \boldsymbol{H}[\mathbf{3}], \dots\dots\dots \boldsymbol{H}[\boldsymbol{i}], \dots\dots\dots, \boldsymbol{H}[\boldsymbol{n}]\} \dots.(\mathbf{1})$$

Where i is the color bin in the color histogram and H[i] represents the number of pixels of color i in the image, and n is the total number of bins used in color histogram. Typically, each pixel in an image will be assigned to a bin of a color histogram. Accordingly in the color histogram of an image, the value of each bin gives the number of pixels that has the same corresponding color. In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H' is given as in Equation (2):

$$\boldsymbol{H'} = \{\boldsymbol{H'}[\mathbf{0}], \boldsymbol{H'}[\mathbf{1}], \boldsymbol{H'}[\mathbf{2}], \dots\dots, \boldsymbol{H'}[\boldsymbol{i}], \dots., \boldsymbol{H'}[\boldsymbol{n}]\} \dots..(\mathbf{2})$$

Where$\boldsymbol{H'}[\boldsymbol{i}] = \frac{\boldsymbol{H}[\boldsymbol{i}]}{\boldsymbol{p}} \dots..(\mathbf{3}),$ p is the total number of pixels of an image [7].

Like color, the texture is a powerful low-level feature for image search and retrieval applications. Much work has been done on texture analysis, classification, and segmentation for the last four decade, still there is a lot of potential for the research. So far, there is no unique definition for texture; however, an encapsulating scientific definition as given in [8] can be stated as, “Texture is an attribute representing the spatial arrangement of the grey levels of the pixels in a region or image”. The common known texture descriptors are Wavelet Transform [9], Gabor-filter [10], co-occurrence matrices [11] and Tamura features [12]. We have used Wavelet Transform, which decomposes an image into orthogonal components, because of its better localization and computationally inexpensive properties [13 and 14].

4.1.Texture Image Retrieval Procedure

The multi wavelet is calculated of each image from the database. The multi wavelet decomposed in to 16 sub bands then calculate the energies of 16 sub bands and those energy values are arranged in a row vector.

For example:

Energy=[e1 e2 e3……e16] ;

Wheree1, e2, e3 are the energy values of each sub band. Calculate the energy of all decomposed images at the same scale, using Equation (4):

$$\boldsymbol{E = \frac{1}{MN}\sum_{i=1}^{m}\sum_{j=1}^{n}|X(i,j)|} \textbf{……(4)}$$

Where M and N are the dimensions of the image, and X is the intensity of the pixel located at row „i" and column „j" in the image map. These energy level values are stored to be used in the Swain &Ballard Distance or Euclidean distance algorithm.

5. Feature Similarities

In this section, we are using Swain &Ballard Distance and Euclidean distance to measure the similarity between Database images and query image.

Swain & Ballard Distance between two vectors I and J is given by Equation (5)

$$S(\boldsymbol{I},\boldsymbol{J}) = \frac{\sum_{i=1}^{N}\mathbf{min}(\boldsymbol{f}_{i(I),}\boldsymbol{f}_{i(J)}}{\sum_{i=1}^{N}\boldsymbol{f}_{i(J)}} \text{…….(5)}$$

Here,

fi(I): is the energy vector or color Histogram of Query mage

fi(J): is the energy vector or color Histogram of database images.

The Euclidean distance is given by Equation (6).

D= (sum ((ref-dbimg). ^2). ^0.5) ……(6)

Here,

ref: is the query image vector either energy or color Histogram vector. dbimg: is the database image vector.

In equation (6) the numerator signifies the difference and denominator normalizes the difference. Thus distance values will never exceed one, being equal to one whenever either of the attributes is zero. The performance is measured in terms of the average retrieval rate, which is defined as the average percentage number of patterns belonging to the same image as the query pattern in the top 9 matches.

6. Results

The proposed system is developed on Matlab tool for the **evaluation** of performance metrics. The obtained simulation results were processed on Dinosaur and Horse database images. The simulative results obtained are illustrated below in Figures (3) to Figure (6).

7. Conclusion

This paper presents image retrieval based on Discrete Wavelet transform and multiwavelet has been proposed. It is better accuracy and computation complexity is low. The computational steps are effectively reduced with the use of Discrete Wavelet transform and multiwavelet. As a result, there is a substation increase in the retrieval speed. The whole indexing time for the 1000 image database takes 2 minute 3 seconds.

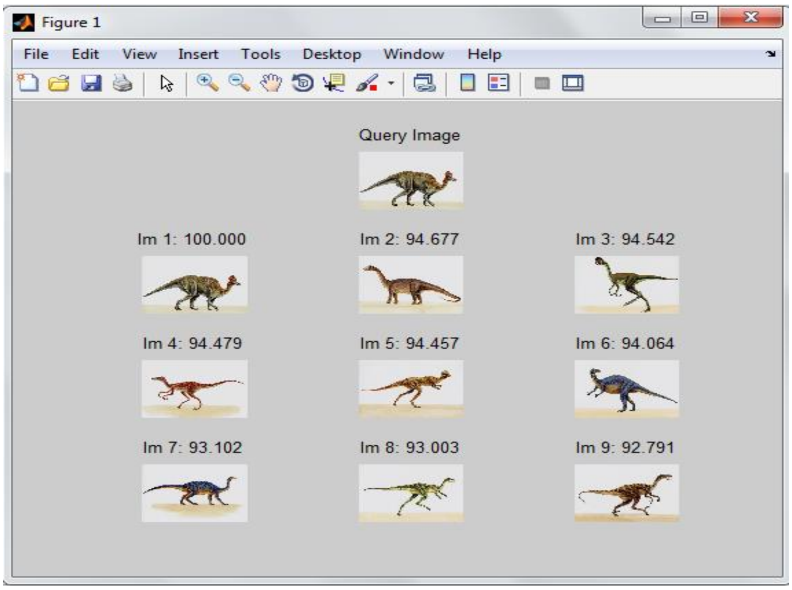

Figure (3): Dinosaur query image with its retrieved images using wavelet color histogram.

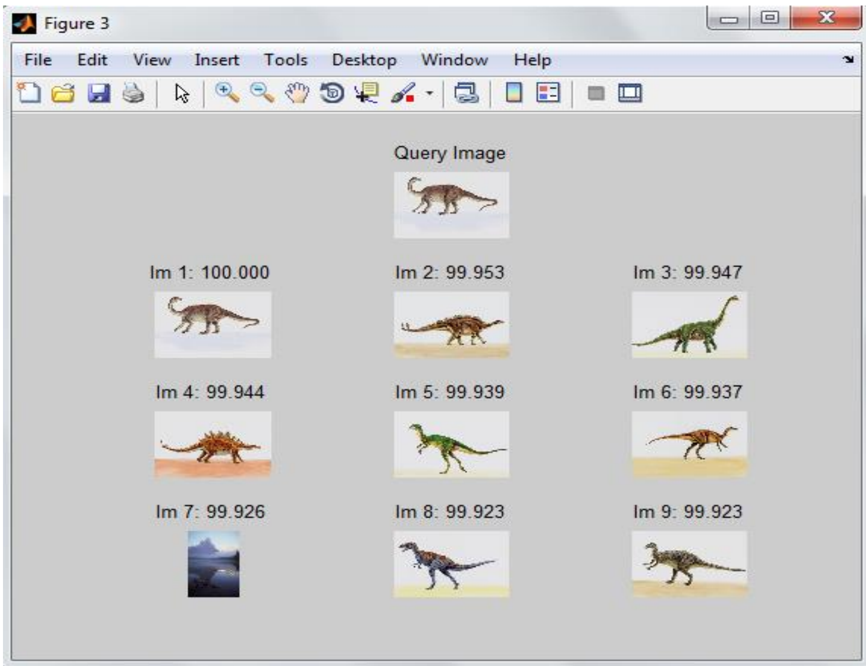

Figure (4): Dinosaur query image with its retrieved images using wavelet-based texture feature extraction

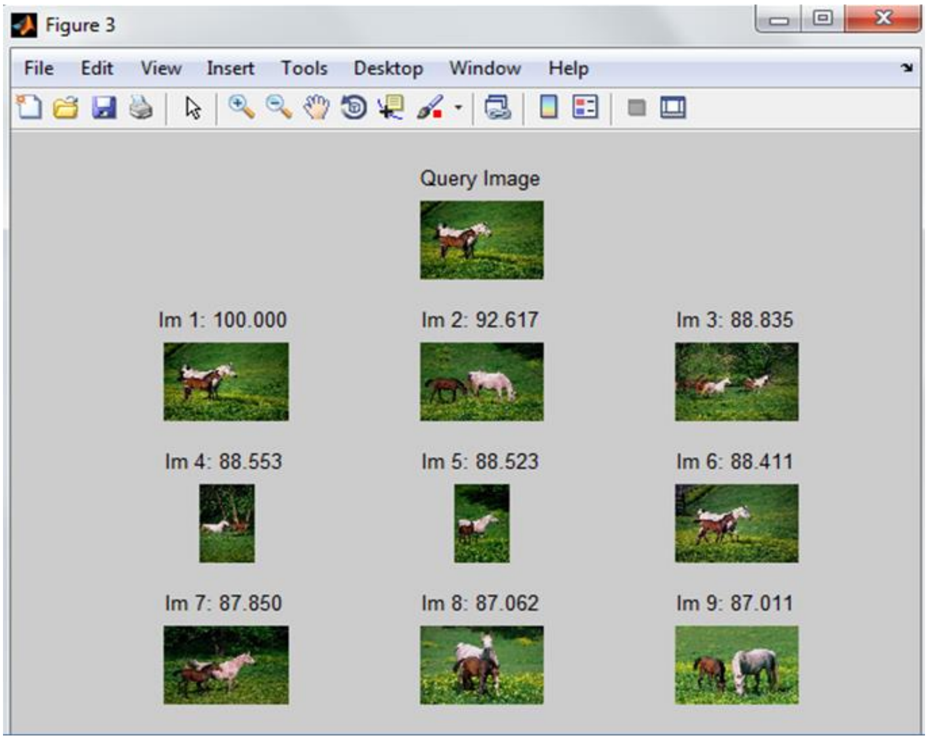

Figure (5): Horse query image with its retrieved images using wavelet color histogram.

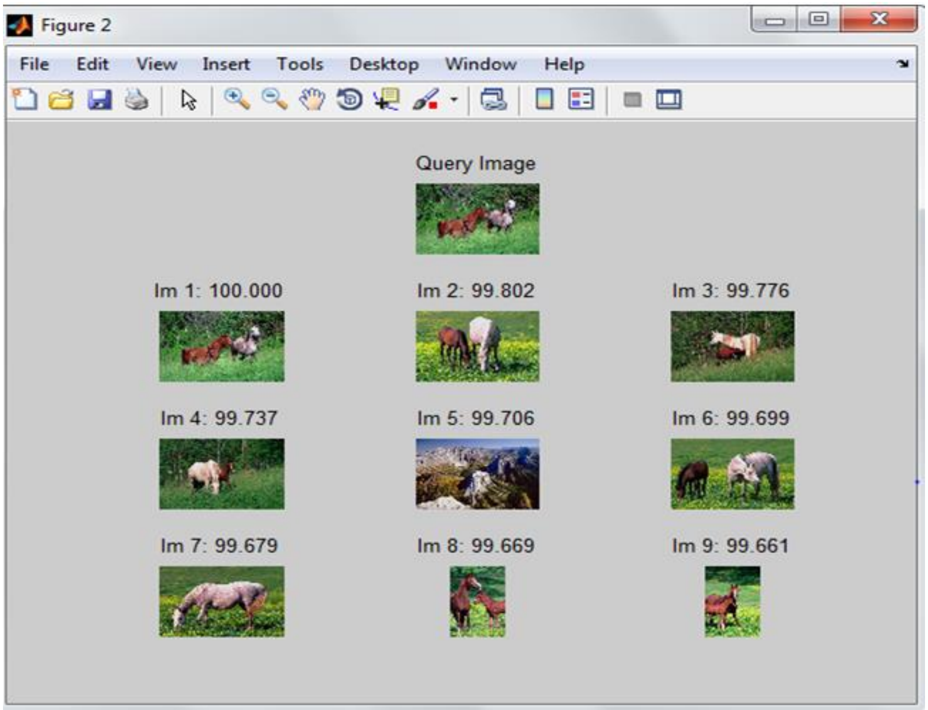

Figure (6): Horse query image with its retrieved images using wavelet-based texture feature extraction

References

[1] J Eakins, M E Graham, Content-based Image Retrieval, A report to the JISC TechnologyApplications Programme, 1999.

[2] Thomas S. Huang, Yong Rui, and Shinh-Fu Chang, Image retrieval: Past, Present and Future, International Symposium on MultimediaInformation Processing, 1997.

[3] M. Myron Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D.Petkovic, D.Steele andP.Yanker, “Query by image content: The QBIC system”, In IEEE Computer, pp. 23-31, Sept.1995

[4] S. Satini, R. Jain. Similarity Measures. IEEE Transactions on pattern analysis and machine Intelligence, Vol.21, No.9, pp 871-883, September 1999.

[5] Hiremath.P.S, Shivashankar . S. Wavelet based features for texture classification, GVIP Journal, Vol.6, Issue 3, pp 55-58, December 2006.

[6] B. S. Manjunath and W.Y. Ma. Texture Feature for Browsing and Retrieval of Image Data. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 8, 1996.

[7] J. R. Smith and S. F. Chang, “Tools and techniques for color image retrieval”, in: IST/SPIE-Storage and Retrieval for Image and Video Databases IV, San Jose, CA, 2670, 426-437, 1996.

[8] IEEE. IEEE standard glossary of image processing and pattern recognition terminology. IEEE Standard. 610.4-1990. 1990.

[9] J.R. Smith and S. Chang, “Transform Features for Texture Classification and Discrimination in Large Image Databases. Proceeding”, IEEE International Conference on Image Processing, Vol. 3, pp-407-411, 1994.

[10] B. Manjunath, P. Wu, S. Newsam and H. Shin, “A texture descriptor for browsing and similarity retrieval”, Journal of Signal Processing: Image Communication, vol. 16, pp- 33-43, 2000.

[11] R. Haralick, “Statistical and structural approaches to texture”, Proceedings of the IEEE, Vol. 67, pp. 786–804, 1979.

[12] H. Tamura, S. Mori and T. Yamawaki, “Textural features corresponding to visual perception”, IEEE Transactions. On Systems, Man and Cybern., Vol. 8, pp- 460-472, 1978.

[13] G. V. D. Wouwer, P. Scheunders and D. V. Dyck, “Statistical texture characterization from discrete wavelet representation”, IEEE Transactions on Image Processing, Vol.8, pp-592–598, 1999.

[14] S. Livens, P. Scheunders, G. V. D. Wouwer and D. V. Dyck, “Wavelets for texture analysis, an overview”, Proceedings of Sixth International Conference on Image Processing and Its Applications,Vol. 2, pp-581–585, 1997.

