استرجاع الصور المركبة باستخدام تقانة دمج الصفات المختارة

اياد عبدالقهار عبدالسلام
جامعة بغداد، كلية التربية للبنات، قسم الحاسبات
بغداد، العراق
ydsalam@yahoo.com

المستخلص

تستند معظم أنظمة استرجاع الصور المستندة إلى المحتوى (CBIR) على صفة واحدة مما يؤدي إلى نتائج استرجاع ضعيفة في معظم الحالات. ولذلك، فإن تقانة دمج الصفات هي واحدة من الحلول المعتمدة لتحسين أداء أنظمة استرجاع الصور لأن الصفات المختلفة يمكن أن تمثل خصائص مختلفة ومتكاملة للبيانات. تقنية دمج الصفات المقترحة تعمل على تقليل حجم تمثيل الصفات وتتميز باختيار السمات الأنسب التي تحدد الصفات الأكثر تمثيلا وتمييزا. أثناء عملية الاسترجاع، تتم مقارنة الميزات الخاصة بصور الاستعلام بتلك الموجودة في قاعدة بيانات الخصائص من أجل استرجاع الصور الاكثر تشابه وفقا لمقياس تشابه معد مسبقا. استخدمنا مجموعة بيانات (DTD) المخصصة للاختبار والمكونة من 5640 صورة مصنفة الى 47 مجموعة متشابهة، واثبتت النتائج أن التقانة المقترحة التي تعمل على دمج أربعة ميزات للصورة هي تقانة فعالة في تحقيق دقة أعلى في استرجاع الصور مقارنة مع النتائج التي تم الحصول عليها باستخدام الميزات الفردية اوعن طريق استخدام مزيج من الميزات.

Abstract

Most of Content Based Image Retrieval systems CBIR are based on a single feature at a time which led to poor retrieval results in most situation. Therefore, feature fusion technique is one of the solutions adopted for improving the performance of CBIR systems because different features can represent different and complementary characteristics of the data. The proposed fusion technique is characterize by a reduction of features fusion dimensions, through selecting the most appropriate features that decide the most representative and discriminative features. Retrieval becomes when features and descriptors of the query are compared to those in the features database in order to retrieve image according to its distance to the query. Describable Textures Dataset DTD texture images dataset used for testing, which is contains 5640 texture image, results demonstrate that the proposed technique working on the fuse of four image features, is an effective technique in achieving higher accuracy of image retrieval comparing with results that obtained by individual feature and by using combinations of features.

Keywords: Multimedia, CBIR, Fusion, Texture image, Feature Extraction, Feature Selection.

I. Introduction

The growing volume of digitally created images in the all life areas like medicine, daily the doctors in the hospital have to access large volumes of images, Journalists also need to look for the images by several conditions, and private life, the home users have a database which contains hundreds of images, etc., all of these fields and others demanding new methods for retrieving images.

Two adaptive approaches for image retrieval these are; firstly, Text-Based Image Retrieval (TBIR) approach which needs human classification and explanation of the image collection using keywords only as query to retrieve the similarity images, thus it is performed based on image annotation or metadata of image to retrieve the desired image [1]. Secondly, Content-Based Image Retrieval (CBIR) where it uses an example of an image as query, hence the similarity images are retrieved based on image content known as low-level features like color, shape, and texture [2, 3].

Many algorithms were proposed and developed by researchers for similarity image retrieval based on TBIR, the best real example of these algorithms that implemented in the search process through Yahoo, AltaVista, Google, etc., are relies deeply on the metadata associated with images, such as keywords and caption and represent the top of TBIR systems [4].

TBIR forced many problems because of the increasing amount of digital images which make the keywords annotation not applicable, and people assign different labels to the same images, so using the second approach CBIR as an alternative mechanism is necessary. With the development of computer technology, CBIR will be more important to overcome the problems of TBIR system mentioned above, there are various applications using CBIR, like crime investigation, face recognition, finger print or retina scanning for access privileges, medical diagnosis, and museum images [5].

II. FEATURES FUSION

The main and important component of CBIR is Feature Extraction (FE) process, which various features are extracted that represent the most important information which describe the image in understandable format like shape, color, and texture of any image. There are many descriptors can extracted from an image to represent these features. So, these feature descriptors will be collected in Feature Vectors (FVs) [6].

The backbone in accomplishing an efficient retrieval system using CBIR approach is to select the most proper features that represent the image as unique as possible using one of feature sufficient and discriminative selection method to increase the efficiency of retrieval. Most of CBIR systems depend on one of these features in each retrieval process, since each feature is describe the certain aspect of image content, it is very difficult to get satisfactory retrieval results, so designing and developing CBIR based on combination of appropriate relevant features to yield better retrieval performance be the ideal solution to achieve perfect retrieval by measuring the image similarity based on multiple visual features, the Features Fusion Technique (FFT) is necessary to represent the images as matchless as possible, and to extend the system ability for accuracy result [6].

Fusion data is more enriched in information than those brought from a single source, it is an important aspects of an intelligent system which combine of data from multiple resources to form useful information. Although the Feature Fusion Vector (FFV) obtained by fuse multiple feature vectors gives extra knowledge and can be increased the performance of CBIR system, but at the same time, it may yield a large Feature Vector (FV) contains hundreds of dimensions which lead to be long time-consuming for retrieving similar images from a large database. So, proposing a new FFT is required to avoid these problems.

III. TEXTURE IMAGES

Texture is a structure of surfaces formed by repeating a particular subimage or several elements in a specific spatial positions. Generally, the repetition involves local variations of scale, orientation, or other geometric and optical features of the elements [7], it contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment, such as; clouds, leaves, fabric, tree bark, water, etc [7]. Typically, texture features methods calculate the degree of uniformity, roughness, contrast, frequency, coarseness, directionality, density, regularity, linearity, and phase

[8]. Although texture is quite easy for humans to recognize and describe, it is quite subjective by its nature and is extremely difficult to precisely define and analyze by digital computers. Since the texture features could be very useful.

Texture is one of the most important characteristic which used to classify and recognize objects and used in finding similarities between images in multimedia databases. Fundamentally, texture formation methods can be classified into five approaches (structural, statistical, signal processing, stochastic, and morphology based method), Out of these five groups of methods, structural and statistical methods are the most widely used because they can be directly applied onto any type of texture [9].

Most of the texture analysis methods consist of two successive stages: feature extraction and feature based classification [10]. Therefore, the texture features could obtained using different types of methods; they can be used individually or in combination with each other.

IV. PROPOSED SYSTEM LAYOUT

The architecture of the proposed CBIR system is illustrated in Figure (1), it is consist of three main parts:

- Building feature vectors database: for each image in the image database, preprocessing sub stage should be done. To describe the contents, image should be processed first in order to extract the features. The processing involves footnote removing, filtering, normalization, segmentation, and object identification. The output of this stage is a set of significant regions and objects. Other sub stages of the proposed CBIR system is the Feature Extraction (FE) function, four FE methods are proposed which are dealing with texture images.
- Building a feature vector for query image: by applying same procedure in the first part on the image which required to retrieve the similarities.
- Retrieving process: The proposed CBIR system select the feature values that are effected in texture, and supported two retrieval process, firstly, image retrieval based on individual image features. Secondly, IR based on fusion feature technique, the similarity matching process depend on the distance between query image vector and each vector in obtained feature DB. Then some measures for testing the performance should be applied.

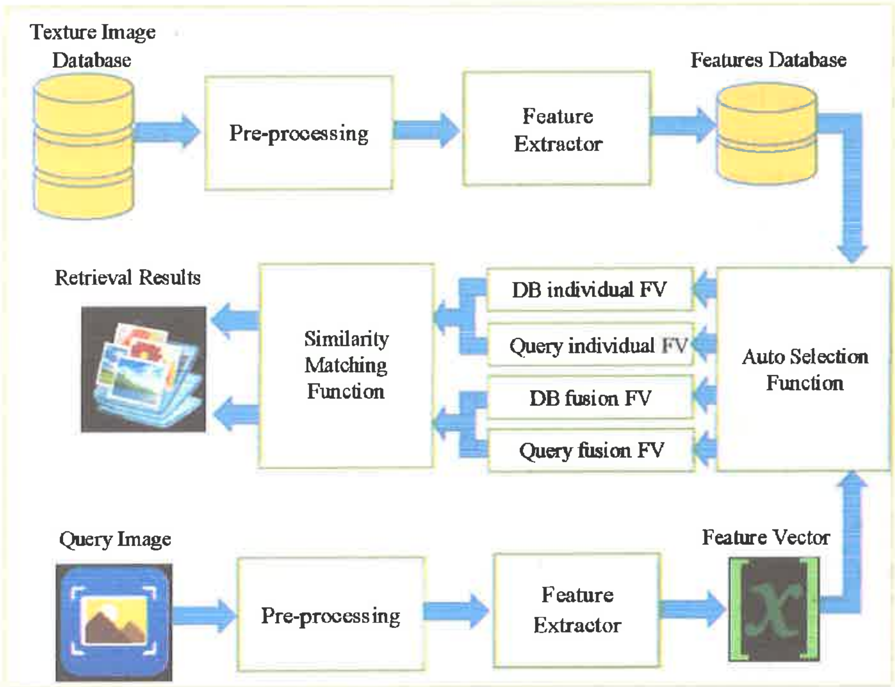

Figure (1) Proposed System Architecture.

V. FEATURE EXTRACTION

One of important texture characteristic is directionality, in this work, the directional attributes have been used to distinguish between the textures regions, where different sets of gradient based features were used as an extracted feature vectors, they are:

1. Co- occurrence Feature Vector (V1)

The co-occurrence matrix $C(i, j)$ calculates the co-occurrence of pixels with gray values i and j at a given distance d. The distance d is defined in polar coordinates (d, θ), with discrete length and orientation. In practice, θ takes the values $0°; 45°; 90°; 135°; 180°; 225°; 270°;$ *and* $315°$ [9]. The co-occurrence matrix $C(i, j)$ can now be defined as in equation (1):

$$C(i,j) = card\begin{cases} ((x_1,y_1),(x_2,y_2)) \in (XY) \times (XY); \; for\; f(x_1,y_1) = i, f(x_2,y_2) = j \\ (x_2,y_2) = (x_1,y_1) + (d\cos\theta, d\sin\theta); for\; 0 < i,j < N \end{cases} \quad \dots (1)$$

Let G be the number of gray-values in the image, then the dimension of the co-occurrence matrix $C(i, j)$ will be $N \times N$.

Features can be extracted from the co-occurrence matrix to reduce feature space dimensionality and the formal definitions of five features from the co-occurrence matrix are done as in equations 2 to 6 [9]:

$$Energy = \sum_i \sum_j C(i,j)^2 \qquad (2)$$

$$Inertia = \sum_i \sum_j (i-j)^2\, C(i,j) \qquad (3)$$

$$Correlation = \frac{\sum_i \sum_j (ij)C(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j} \qquad (4)$$

$$Difference\ Moment = \sum_i \sum_j \frac{1}{1+(i-j)^2}\, C(i,j) \qquad (5)$$

$$Entropy = -\sum_i \sum_j C(i,j) \log C(i,j) \qquad (6)$$

Where

$$\mu_i = \sum_i i \sum_j C(i,j)$$

$$\mu_j = \sum_i C(i,j) \sum_j j$$

$$\sigma_i = \sum_i (i-\mu_i)^2 \sum_j C(i,j)$$

$$\sigma_j = \sum_i (j-\mu_j)^2 \sum_j C(i,j)$$

The feature vector is constructed using $\mu(W, \theta, \sigma_x, \sigma_y)$, $std(W, \theta, \sigma_x, \sigma_y)$ and *Skew* as feature components, where σ_x and $, \sigma_y$ are the scaling parameters, W is the radial frequency of the sinusoid and $\theta \in [0, \pi]$ specifies the orientation [11].

2. First Order Gradient Feature Vector (V2):

A set of first order derivative in image processing are implemented using the magnitude of the gradient. For the function $I(x,y)$, the gradient of I at coordinates (x,y) is defined as the two dimensional vector as in equation (7) [10].

$$\frac{dy}{dx} = \begin{bmatrix} G_x & G_y \end{bmatrix} \qquad (7)$$

Where G_x and G_y are the horizontal and vertical derivatives, respectively. This vector holds geometrical information that points to the direction of the greatest rate of change in I at (x, y). The gradient approximations results can be combined to give the gradient magnitude at each point in the image, using equation (8):

$$G = \sqrt{G_x^2 + G_y^2} \qquad (8)$$

Sobel operator is used here for computing G_x and G_y as in equation (9):

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \ and \ G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \qquad (9)$$

A feature vector (V2) is constructed by computing 12 features representing the means of G_x, G_y, and G for four directions.

3. Min- Max Gradient Feature Vector (V3)

This set of features depends on the minimum and maximum values of the first derivative for each direction, equations (10), and (11):

$$G_{min}(x, y) = Min(G_x(x, y), G_y(x, y)) \qquad (10)$$
$$G_{max}(x, y) = Max(G_x(x, y), G_y(x, y)) \qquad (11)$$

The established feature vector (V3) consists of 8 values, each represents the mean values of G_{min} or G_{max} for each of the four directions. At any point G_{max} takes its value to be the greater between the horizontal and vertical derivative (i.e. it measures the maximum rate of changes at any point despite its direction; that makes it invariant under rotation). The same is true for G_{min}.

4. Second Order Gradient Feature Vector (V4)

The established second order gradient feature vector consist of 2nd gradient values along the horizontal, vertical, diagonal and second diagonal direction as in figure (2):

x-1, y-1	x-1, y	x-1, y+1
x, y-1	x, y	x, y+1
x+1, y-1	x+1, y	x+1, y+1

Figure (2) Sub Image Coordinates

They can be computed using the following equations 12 to 15:

$$G_{x2}(x,y) = 2I(x,y) - I(x-1,y) - I(x+1,y) \quad (12)$$
$$G_{y2}(x,y) = 2I(x,y) - I(x,y-1) - I(x,y+1) \quad (13)$$
$$G_{dm2}(x,y) = 2I(x,y) - I(x-1,y-1) - I(x+1,y+1) \quad (14)$$
$$G_{ds2}(x,y) = 2I(x,y) - I(x+1,y-1) - I(x-1,y+1) \quad (15)$$

The feature vector (V4) consists of 16 values, 4 for each direction.

The above four introduced features vectors (V1, V2, V3, and V4) methods tend to capture different image texture characteristics. The use of different combinations between these features could improve the classifier accuracy.

VI. CLASSIFIERS

Appropriate classifier should be selected when the features are extracted. A number of classifiers are used, each chosen classifier is found suitable to determine a particular kind of feature vectors. The classifier that used commonly is Nearest Neighbor classifier. The Nearest Neighbor classifier is used to calculate the distance of similarity of the feature vector of the query image with image feature vectors stored in the database. It is obtained by finding the shortest distance between the query image and the database vectors. Classifier used two times, first for individual feature vectors, then for fusion feature vectors.

VII. EVALUATION METHOD

Accuracy is a direct measurement of the quality and user satisfaction of the image retrieval process. To evaluate the efficiency of IR system, two well-known metrics are used, reliability which outlined in equation (16) is defined as the ratio of the relevant images that are retrieved to the total number of retrieved images, and recovery which outlined in equation (17) is defined as the ratio of the relevant images that are retrieved to the total number of relevant images in the database. Reliability and recovery metrics give an estimate of retrieval efficiency in the range from worst case (equal to 0) to perfect retrieval (equal to 1) [12]:

$$Reliability = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ retrieved\ images} \quad (16)$$

$$Recovery = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ database} \quad (17)$$

VIII. EXPERIMENTAL RESULTS AND ANALYSIS

The testing results was evaluated using Describable Textures Dataset DTD which is a texture database, it is developed in 2012 at the Johns Hopkins Centre for Language and Speech Processing (CLSP) [13], consisting of 5640 images, organized according to a list of 47 terms (categories) inspired from human perception. There are 120 images for each category. Image sizes range between 300x300 and 640x640, and the images contain at least 90% of the surface representing the category attribute, each image provided by key attribute (main category) and a list of joint attributes, some texture pictures from DTD are shown in Figure (3).

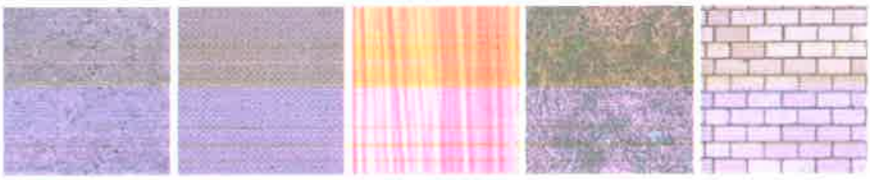

Figure (3) some images from DTD

The dataset is split into two parts, training, and testing, for first experiment a set of 40 randomly chosen images from each class were used for training (building feature database), while the rest 80 samples were used for testing the classifier efficiency, the number of training/testing is changed in other experiments. The conducted tests have been directed toward finding the similar images which is belong to the same class according to distance measure.

Table (1) shows the comparison in term of Success Retrieval Rate between the results that were obtained using individual feature descriptors V1, V2, V3, and V4, then the combination of descriptors {V1, V2}, {V3, V4}, and {V1, V2, V3, V4} with results that were obtained using proposed fusion vector. Three experiments are applied with different training/ testing number of images for each class, these are 40/80, 60/60, and 80/40. It is clear that the fusion had achieved higher retrieving accuracy.

Table (1) retrieving accuracy for three experiments.

Descriptor	Success Retrieving rate for Train/test 40/80	Success Retrieving rate for Train/test 60/60	Success Retrieving rate for Train/test 80/40	Average
V1	0.683	0.725	0.775	0.728
V2	0.692	0.750	0.767	0.736
V3	0.608	0.658	0.692	0.653
V4	0.658	0.708	0.733	0.700
{V1, V2}	0.733	0.783	0.800	0.772
{V3, V4}	0.725	0.758	0.775	0.753
{V1, V2, V3, V4}	0.850	0.900	0.917	0.889
Fusion Vector	0.908	0.958	0.975	0.947

It can clearly be seen that the retrieving results of fusion (0.947) more effective than individuals, and the success rate increasing when the number of training images be increased. The performance of the combination techniques are evaluated and summarized in Table (1). Since we cannot rely on a single feature in the image retrieval because it cannot represent the images as much as possible, so fusing the most appropriate features descriptors to increase the accuracy of system's performance has been used.

REFERENCES

[1] Sia Ka Cheung. (2013). "Issues on Content-Based Image Retrieval", Msc. Thesis, the Chinese University of Hong Kong.

[2] Ismael, M, Maher, (2017). "A survey of content-based image retrieval with high-level semantics", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 5.

[3] Deepak S. Shete, M.S. Chavan, (2012). "Content Based Image Retrieval Review", International Journal of Emerging Technology and Advanced Engineering, (IJETAE), Vol. 2, Issue 9.

[4] Datta R., Joshi, D., Li, J., Wang J.Z., (2008). "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Transactions on Computing Surveys, 40(2).

[5] Er. Kirti Sachdeva, O.P.Vinocha, (2014). "A Comparative Study of Factors Affecting Performance of Local Binary Pattern (LBP) Variant along with Distance Metrics for Face Recognition", International Journal of Scientific & Engineering Research, Vol. 5, Issue 1.

[6] Ashish Oberoi, Manpreet Singh, (2012). "Content Based Image Retrieval System for Medical Databases (CBIR-MD)- Lucratively tested on Endoscopy, Dental and Skull ages", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1.

[7] Zhang, J., Li, G., He, S. (2008). "Texture-based image retrieval by edge detection matching GLCM", The 10th IEEE International Conference on High Performance Computing and Communications.

[8] Smita S Patill, A. A. Junnarkar, D. V. Gore. (2014). "Study of Texture Representation Techniques", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Vol (3), Issue 3.

[9] Prasad P. S.; Varma V. T.; Harish V. S.; and Kumar K. S. (2012). "Classification of Different Textures Using SVM and Fuzzy logic", International Journal of Advanced Computer Research Vol. 2, No.4, Issue. 6.

[10] Proch Zka, A.; Mare. J.; and Yadollahi. M., (2012). "Biomedical Image Enhancement, Segmentation and Classification Using Wavelet Transform", WSEAS International Conference on Systems; July, 2012.

[11] Ryszard S. Chora, (2007). "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", International Journal of Biology and Biomedical Engineering, Poland.

[10] C. Panagiotakis and G. Tziritas. (2015)." A speech/music discriminator based on rms and zero-crossings". IEEE Transactions on Multimedia, Feb. 2015.

[11] T. Wold, D. Blum, and J. Wheaton. (2006). "Content-based classification, search, and retrieval of audio". IEEE Multimedia, 2006.

[12] M. S. Nixon, et al., (2008). "Feature extraction and Image Processing", book, Elsevier.

[13] DTD, Describable Textures Dataset, (2012), "Johns Hopkins Centre for Language and Speech Processing (CLSP), https://www.robots.ox.ac.uk/~vgg/data/dtd/