PalArch's Journal of Archaeology of Egypt / Egyptology

"Image Retrieval Based on Colour, Texture and Shape Feature Similarity Score Fusion Using Genetic Algorithm"

Abhijit T. Somnathe[1], Dr. S. Ram Reddy[2]
[1]Research Scholar, JJTU.
[2]Research Guide, JJTU.
Email: abhijitsomnathe@gmail.com[1], Sramreddy25@gmail.com[2]

Abhijit T. Somnathe, Dr. S. Ram Reddy, Image Retrieval Based on Colour, Texture and Shape Feature Similarity Score Fusion Using Genetic Algorithm, -- Palarch's Journal Of Archaeology Of Egypt/Egyptology 18(7). ISSN 1567-214x

Keywords: image retrieval; color feature, texture feature, shape feature, feature fusion; genetic algorithm.

ABSTRACT

This paper proposes an image retrieval method based on colour, texture and shape feature similarity score fusion. For checking similarity score we'll use genetic Algorithm. Single feature describes image content only from one point of view, which has a certain limitations. Fusing multifeature similarity score will improve the system's retrieval performance. In this paper, the retrieval results from color, texture and shape feature are analyzed, and the method of fusing multi-feature similarity score is described. For the purpose of assigning the fusion weights of multi-feature similarity scores reasonably, the genetic algorithm is applied. For comparison, other three methods are implemented. They are image retrieval based on color feature, texture feature and fusion of color-texture feature similarity score with equal weights. The experimental results show that the proposed method is superior to other methods.

I. INTRODUCTION

With the rapid development of multimedia and network technology, people can access a large number of multimedia information. For people who want to make full use of multimedia information resources, the primary question is how to query the multimedia information of interest. Text query can be applied to multimedia information retrieval, but it has inherent deficiencies. One hand, text annotation of multimedia information will spend a lot of manpower and resources and it is inefficient. On the other hand, annotated text is usually a person's perception of multimedia information. It is subject to impact of individual difference and state of human and environment, and the described results may be more one-sided. In addition, it is clearly incomplete to describe content-rich multimedia information with a small amount of text. Content Based Image Retrieval (CBIR) techniques appeared in 1990s. It solves the above problems well. It uses low-level features like color, texture and shape to describe image content, and breaks through the limitation of traditional text query technique.

Content-Based Image Retrieval (CBIR) emerged as a promising substitute to surpass the challenges met by text-based image retrieval solutions. In fact, digital images, which are mined using CBIR system, are represented using a set of visual features. CBIR system can be implemented based on single feature. Single image feature describes the content of an image from a specific angle. It may be suitable for some images, but it also may be difficult to describe other images. Moreover, describing an image with single feature is also incomplete. Representing an image with multi-features from multi-angles is expected to achieve better results. Information is multisource, and information fusion approach is diverse. The problem how to organize multi-source information in a suitable way to achieve the intended results attracts extensive attention from the researchers in this field. .Information fusion can be carried out in feature level .Information fusion in feature level has advantage in some extent. Because different features reflect the different characteristics of the image, if those features are integrated reasonably, the results will both reserve the discriminate information of multi-feature and eliminate the interference resulted from the difference of multi-feature.

CBIR systems represent an improvement taking advantage of the digital information stored in the image itself when image collections are not semantically annotated with textual labels. Thus, visual features are extracted from images in order to describe its content, and later be compared with the image query. These visual features used in CBIR systems can be classified into low level features (colour, texture and shape) and high level features, which are usually obtained by combining low level features with a predefined model. High level features are not usually suitable for general purpose systems as they have a strong dependency on the application domain, so the extraction of good low level image descriptors in an important research activity in this field. Nevertheless, although the low level features can easily describe the content of simple images, complex images and high level concepts cannot be properly described. This gap between high level concepts closer to human perception and low level features used to describe images is called semantic gap, and different methods have been proposed to deal with it. In many cases, the strategies proposed are based on the integration of the information provided by the user into the decision process.

The main goal of image retrieval is to retrieve a set of semantically similar images in database based on a query image. This similarity matching can be performed by computing the distance score of the feature descriptors between the query and target images in database. Many methods have been developed for the content-based image retrieval task such as. The image retrieval offers a convenient way to browse and search a set of similar images which can reduce the user time for searching a set of images with similarity and user preference constraints.

Most state of the art methods lack the ability to successfully incorporate human intuition into retrieving images. Retrieving the image that the user wants is a challenge due to inefficiency in explicit description. In order to supplement the absence of the user competence, we developed automatically detected facial attribute with relevance feedback based face image retrieval.

II. LITERATURE REVIEW

Many researchers suggest different methods of image retrieval. But they are particular application based methods. Some are discussed below,

(Nishant Shrivastava and Vipin Tyagi, 2014) have discussed the image retrieval technique which retrieves similar images in three stages. A fixed number of images are first retrieved based on their colour feature similarity. The relevance of the retrieved images is further improved by matching their texture and shape features respectively. This eliminates the need of fusion and normalization techniques, which are commonly used to calculate final similarity scores. This reduces the computation time and increases the overall accuracy of the system. Moreover, in this technique, global and region features are combined to obtain better retrieval accuracy. This concept only focused the similarity score assignment based on features not considering the semantics of the image. Accordingly the accuracy of the image retrieval gets low.

(Mourão et al. ,2015)presented a medical information retrieval system with support for multimodal medical case-based retrieval. The system supports medical information discovery by providing multimodal search, through a novel data fusion algorithm, and term suggestions from a medical thesaurus. This algorithm was not be a generic. This method is not provided for the new modality. Ahmad et al. discussed an efficient framework to model image contents as an undirected attributed relational graph, exploiting colour, texture, layout, and saliency information. This method encodes salient features into this rich representative model without requiring any segmentation or clustering procedures, reducing the computational complexity. In addition, an efficient graph-matching procedure implemented on specialized hardware makes it more suitable for real-time retrieval applications. This method exploits the local salient maxima. Consequently the computation time was increased.

(Jiabo Huang et al. ,2019) discussed an unsupervised visual hashing approach called semantic-assisted visual hashing (SAVH). Distinguished from semi-supervised and supervised visual hashing, its core idea is to effectively extract the rich semantics latently embedded in auxiliary texts of images to boost the effectiveness of visual hashing without any explicit semantic labels. Multiple-feature hashing is transformed to a similarity preserving problem with linearly combined kernel functions, which are corresponding to the similarity measures for individual features. Dosovitskiy et al. presented a class of loss functions, which we call deep perceptual similarity metrics (DeePSiM), allowing to generate sharp high resolution images from compressed abstract representations. Instead of computing distances in the image space, compute distances between image features extracted by deep neural networks.

(Liang et al., **2010)** studied the problem of content-based image retrieval. In this problem, the most popular performance measure is the top precision measure, and the most important Component of a retrieval system is the similarity function used to compare a query image against a database image. However, upto now, there is no existing similarity learning method proposed to optimize the top precision measure. To fill this gap, in this paper, present a similarity learning method to maximize the top precision measure. This minimization problem is solved as a quadratic programming problem. This algorithm was not useful for some other similarity function as similarity measure instead of linear function, such as Bayesian network.

(Bala et al. 2016)explained the feature descriptor, local texton XOR patterns (LTxXORP) is proposed for content based image retrieval. This method collects the text on XOR pattern which gives the structure of the query image or database image. First, the RGB (red, green, blue) color image is converted into HSV (hue, saturation and value) color space. Second, the V color space is divided into overlapping sub blocks of size2 × 2 and textons are collected based on the shape of the textons. Then, exclusive OR (XOR) operation is performed on the texton image between the center pixel and its surrounding neighbors. Finally, the feature vector is constructed based on the LTxXORPs and HSV histogram.

(Zhou et al. 2019) explained that a sparse bilinear similarity function is introduced to model the relative characteristics encoded in kin data. The similarity function parameterized by a diagonal matrix enjoys the superiority in computational efficiency, making it more practical for real-world high-dimensional kinship verification applications. Then, ESL learns from kin dataset by generating an ensemble of similarity models with the aim of achieving strong generalization ability. Specifically, ESL works by best satisfying the constraints (typically triplet-based) derived from the class labels on each base similarity model, while maximizing the diversity among the base similarity models. This method was not interested in investigating the effect of face alignment pipelines on kinship verification, and fusion of multiple feature representations in the ESL framework to further improve the verification performance.

Rapid growth for storing and capturing multimedia data with digital devices, in recent years, information of multimedia retrieval has one of the most important researches and key challenging problems with image retrieval. Thus, Content-Based Image Retrieval (CBIR) is used that deals with retrieval of similar images from a large database for a given input query image. A large number of diverse methods have been proposed for CBIR using low level image content like edge, colour and texture. For combination of different types of content, there is a need to train these features with different weights to achieve good results.

III. PROBLEM STATEMENT

In expensive image-capture and storage technologies have allowed massive collections of digital images to be created in lots of application areas such as medicine, remote-sensing, entertainment, education and on-line information services. Storing of such images is relatively above-board, but an accessing and searching image database is intrinsically difficult and key challenging problems with image retrieval. Image retrieval is done by two ways,

1] Text-based approach

2] Content based approach.

In text-based approach, image is search through image databases using manual keyword annotations and keyword indexing. Although this approach can be useful, there are two rigorous problems with it. First, textual tagging of images requires a lot of user effort and is also a very tedious task. If the database is very large, then it is not practically feasible to annotate each image with textual tags. Second is that there is no standard vocabulary that is used for textual tagging. The perception of images may vary from person to person, resulting in different tags for images that are otherwise similar. Also the vocabulary that is available may not be sufficient to completely describe the image. Therefore, new ways of indexing, browsing and retrieval of images are needed, which can automatically generate descriptors for images.

Content-based image retrieval (CBIR) is a technique in which images are extracted directly based on their visual descriptor such as colour, texture, and edge. An image has several types of features; every feature has different effect on image retrieval. How to organize these features and properly assign weights to get satisfying retrieval results is a one of the challenge in Content-based image retrieval (CBIR).

IV. IMAGE FEATURE EXTRACTION

The image content is mainly embodied in color, texture and shape etc. The color feature, texture feature and shape feature describes the image content from different angle. More features will provide more information on the image content. This paper focuses on fusion method of multifeature similarity score. For convenience, this paper only discusses the fusion method of two-feature similarity score. Without loss of generality, the used features are color feature and texture feature. The following part describes the used extraction method of color feature and texture feature.

A. Color feature extraction

HSV color model forms a uniform color space, which uses a linear gauge. The perceived distance between colors is in proportion to Euclidean distance between corresponding pixels in HSV color model, and conforms to eye's feeling about color.

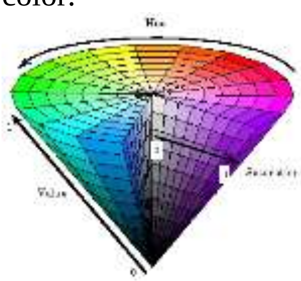

Fig. HSV Color space

So it is very suitable for color based image similarity comparison. In this paper, the color histogram in HSV color space is taken as the color feature describing image content. For calculating color histogram in HSV color space, HSV color space must first be quantified. According to human cognitive about color, three components of HSV space are quantified in non-uniform manner. Hue is quantized into 16 bins and is among [0, 15]. Saturation is quantized into 4 bins and is among [0, 3]. Value is quantized into 4 bins and is among [0, 3]. Among those three components, human cognitive about color is mainly based on hue, and then saturation, finally value. So, quantized results are coded as

$$C = 16H + 4S + V \qquad (1)$$

where C is an integer between 0 and 255. Thus the color feature can be obtained by calculating histogram of an image in HSV space.

B. *Texture feature extraction*

In this paper, the statistical properties of image co-occurrence matrix are taken as texture features of an image. Firstly, color image is converted to grayscale image, and the image co-occurrence matrix is gained. Then, the following four statistical properties are calculated to describing image content. They are contrast, energy, correlation and Homogeneity. All these statistical properties are calculated in 4 directions, so we can get 16 texture features. At last, we calculated the means and variances of these four kinds of statistical properties, and took the results as the ultimate texture features, denoted as

$T = (\mu 1, \mu 2, \mu 3, \mu 4, \mu 5; \sigma 1, \sigma 2, \sigma 3, \sigma 4, \sigma 5)$ (2)

C. *Convolutional Neural Network*

The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation. Convolutional networks are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers.

Fig. CNN

The convolutional neural network architecture is defined by a three-dimensional arrangement of neurons, instead of the standard two-dimensional array. The first layer in such neural networks is called a convolutional layer. Each neuron in the convolutional layer only processes the information from a small part of the visual field. The convolutional layers are followed by rectified layer units or ReLU, which enables the CNN to handle complicated information. CNNs are mainly used in object recognition applications like machine vision and in self-driving vehicles.

V. IMPLEMENTATION

The proposed methodology process flow with the implementation coding is given below,

1. Initially a database is created with the images - The image database downloaded from 'http: //wang.ist.psu.edu/docs/related/' website. These databases have thousands of images having a same resolution.

2. Then utilize the deep CNN (Convolutional Neural Network) to extract the image feature. In traditional CBIR systems, low-level features such as the colour, shape and texture features are usually extracted to construct a feature vector for describing images and then, based on a proper similarity measure, images are retrieved by comparing the feature vector corresponding to the query image and those corresponding to images in the data set. . Many researchers devote most of their attention to the selection of appropriate feature extraction method. . However, they usually fail to extract the internal structure contained in the features which is crucial for distinguishing data points. In our paper, we aim to find this internal structure from the original data space. Moreover, deep learning paradigm is that features need not to be extracted from the raw data beforehand, but the raw data themselves are processed by the network that produces an internal feature representation of the data suited for the task at hand. The CNN can also serve as good descriptors for image retrieval.

Deep learning using CNN (Implicit feature extraction)

Step 1: First, try to segment the image using an automated global intensity threshold technique (Otsu's method). The graythresh function uses Otsu's method, which chooses the threshold to minimize the intra class variance of the black and white pixels. The pixel line gives the intensity threshold defined by this function and im2bw segments the image according to the threshold defined via the graythresh command. Then show the new binary image, J2.

Otsu's method: It is based on entirely on the set of histogram counts. Pixels can take on the set of values i=1, 2, 3…L. The histogram count for pixel value i is n_i, and the associated probability is $p_i=n_i/N$, where N is the number of image pixels.

The threshold task is formulated as the problem of dividing image pixels into two classes.C_0 is the set of pixels with values [1,…K] and C_1 is the set of pixels with values in the range [K+1,…L]

The overall class probabilities, ω0 and ω1, are:

$\omega_0 =$

$\omega_1 =$

The class means, μ_0 and μ_1 are the mean values of the pixels in C_0 and C_1. They are given by:

$\mu_0 = / \omega_0 = \mu(k)/\omega(k)$

$\mu_1 = / \omega_1 =$

where

$\mu(k)=$

and μT, the mean pixel value for the total image, is:

$\mu T=$

The class variances, σ_0^2 and σ_1^2, are:

$\sigma_0^2 = {}^2 p_i/\omega_0$

$\sigma_0^2 = {}^2 p_i/\omega_0$

Otsu mentions three measures of "good" class separability: within-class variance (λ), between-class variance (κ), and total variance (η). These are given by:

$\lambda=\sigma^2_B$

$\kappa=\sigma^2_T/\sigma^2_W$

$\eta=\sigma^2_B/\sigma^2_T$

where

$\sigma^2_W=\omega_0\sigma^2_0+\omega_1\sigma^2_1$

$\sigma^2_B=\omega_0(\mu_0-\mu_T)^2+\omega_1(\mu_1-\mu_T)^2=\omega_0\omega_1(\mu 1-\mu_0)^2$.

He goes on to point out that maximizing any of these criteria is equivalent to maximizing the others. Further, maximizing η is the same as maximizing σ^2_B, which can be rewritten in terms of the selected threshold, k:

$\sigma^2_B(k)=[\mu T\omega(k)-\mu(k)]2\omega(k)[1-\omega(k)]$.

Step 2: Apply a filter to find the darkest pixel in each region. You can use either the imerode command to find the darkest pixel in a region of size 'nsize' around each pixel. Then show this image using imshow.

Step 3: Now we want to subtract this from the initial image I. This image processing levels the background intensity of the image. Remember that our image I is stored as uint8, though, so we need to transform it to a double precision matrix before subtracting. The imshow command will show the new leveled image. Now we transform it back to a uint8 image by rescaling the values to integers between 0 and 255. We are going to store this as our new and improved image, I2. Let's also look at the image histogram. Notice how the two distributions have separated by leveling the image.

Step 4: Let's get rid of everything below a certain size using the bwareaopen command and then show the new image.

3. In order to compute the similarity/distance between the images in the database and the query image, the feature vector of query image and feature vector of database images have been compared. The Similarity score fusion is done by the Hybrid GA-Chaotic Fuzzy optimization algorithm. The weight is to be assigned to individual content features. The task of this algorithm consists of finding the weight that maximizes the retrieval accuracy to the context defined by the query image. Here we can have the hybridization with the genetic algorithm, fuzzy c-mean and Improved Glow Worm search optimization. FCM can be terminated once a pre-set convergence criteria is met. Accordingly it can fall prey to local optima problem. In order to avoid this, the Improved Glow worm optimization is incorporated with the FCM. Due to the Chaotic behaviour of Improved Glow worm algorithm the convergence rate will be increased. Furthermore the cross over and mutation step of the genetic algorithm have to be optimized more because of the random searching behaviour. Due to random search we cannot attain the optimal solution. In order to gain that the Chaotic FCM is incorporated to the GA. In accordance with that the optimal similarity score fusion assignment can be made.

4. Next, the user gives feedback to the relevance of the initially returned results and submits user feedback to the CBIR system. A relevance feedback algorithm refines the initial retrieval results based on the user's relevance feedback, and returns an improved set of results to the user. The Relevance Feedback Paradigm is employed to improve the retrieval capabilities of CNN by modifying the weights of the similarity features according to the feedback of the user. Typically, a number of rounds of users' relevance feedback are needed to achieve satisfactory results. Here with the combination of user's feedback (explicit feedback), implicit feedback is also used. That means, here we are using the feedback log with the regular relevance feedback. Log database is used to collect and store user’s relevance feedback. When feedback log data is unavailable, the log-based relevance feedback algorithm behaves exactly like a regular relevance feedback algorithm we need to systematically organize the log data of users' feedback. Assume a user labels N images in each round of regular relevance feedback, which is called a log session. Thus, each log session contains N evaluated images that are marked as either "relevant" or "irrelevant." System accepts explicit and implicit feedback and sends the explicit feedback to the retrieval system and at the same time sends the implicit feedback to the log session. Then sends the processed data to log database. Depending upon these relevance feedback recomputed similarity between query image and image in data base. System is terminated when user is satisfied with

retrieval result.

VI. PROCESS FLOW DIAGRAM

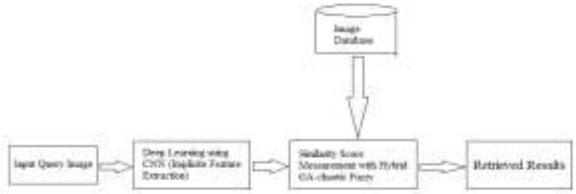

Fig. Process Flow

VII. RESULT

We select an image of monument from the database of nearly 1000 images of different types. We use MATLAB for simulation purpose i.e. for finding similar images with respect to input image.

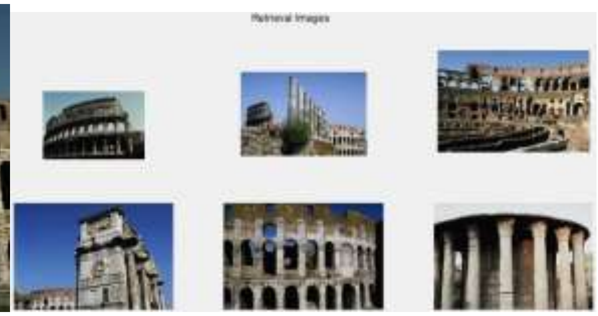

Fig. Input Image Fig. Output Image

We select a particular animal (Elephant) as an input image. Using MATLAB we found the output as follows.

Fig. Input Image Fig. Output Images

Now we choose a horse image as an input and try to find out the relevant images available in database out of these nearly 1000 images. The result obtained is as follows.


Fig. Input Image Fig. Output Images

Now we consider a landscape with mountain in background and a water body in front. The result using MATLAB coding is as follows.


Fig. Input Image Fig. Output Images

Now we consider food image like vegetables, fruits image as an input image and try to find similar images from the database relevant to input image using MATLAB.


Fig. Input Image Fig. Output Images

VIII. CONCLUSION

Image library is used to evaluate the proposed algorithm. It contains different categories, each of which has about 100 images. The total number of images is about 600. They are Alphabets, Flower, Landscapes, Dinosaurs, Automobiles, Animals and so forth. The color feature and texture feature of every image are extracted to build feature database.

The method based on color feature is better than the method based on texture feature. However, relative to the color based image retrieval method, performance of other image retrieval methods based on multi-feature similarity score fusion doesn’t increase much. This is mainly due to that compared with image retrieval method based on color feature; the performance of image retrieval method based on texture feature is poor. There are two possible reasons for it. One is that the color difference of different images in this image library is more obvious, and the performance of image retrieval method based on color feature is better. The other is that for this image library, the extracted texture feature may be insufficient to reflect the differences between different classes, which make performance of image retrieval based on texture feature poor. Better performance of image retrieval method based on texture feature is expected to increase the performance of image retrieval method based multi-feature similarity score fusion.

This project proposed an image retrieval method based multi-feature similarity score fusion. For a query image, multiple similarity score lists based on different features are obtained. Then using genetic algorithm, multi-feature similarity scores are fused, and better image retrieval results are gained. In this project, when we evaluated the fitness of an individual, we considered only the occurrence frequencies of an image in retrieval result, and not the location of an image in retrieval result.

REFERENCE

1) Nishant Shrivastava, Vipin Tyagi,(2014),"Content based image retrieval based on relative locations of multiple regions of interest using selective regions matching" Information Sciences 259:212–224 DOI:10.1016/j.ins.2013.08.0431

2) A Mourão, F Martins, J Magalhaes (2015) " Multimodal medical information retrieval with unsupervised rank fusion" Computerized Medical Imaging and Graphics Volume 39, Pages 35-45

3) Jiabo Huang, Qi Dong, Shaogang Gong, Xiatian Zhu (2019)," Unsupervised Deep Learning by Neighbourhood Discovery "Proceedings of the 36th International Conference on Machine Learning, PMLR 97:2849-2858

4) Z Liang, H Fu, Y Zhang, Z Chi, D Feng - (2010), "Content-based image retrieval using a combination of visual features and eye tracking data" ETRA '10: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications Pages 41–44https://doi.org/10.1145/1743666.1743675.

5) Anu Balaa TajinderKaur (2016) , "Local texton XOR patterns: A new feature descriptor for content-based image retrieval" Engineering Science and Technology, an International Journal Volume 19, Issue 1, , Pages 101-112.

6) Yao Ding, Yanzhao Zhou, Yi Zhu, Qixiang Ye, Jianbin Jiao;(2019)." Selective Sparse Sampling for Fine-Grained Image Recognition" Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 6599-6608

7) Gudivada V. N., Raghavan V. V.,(1995) "Content based image retrieval systems," IEEE Computer, vol. 28, pp. 18-22, .

8) Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang,(2008), "Image retrieval: Ideas, influences, and trends of the new age," ACM Computing Surveys, vol. 40, pp. 1-60.

9) B.G. Prasad, K.K. Biswas,(2004), "Region-based image retrieval using integrated color, shape, and location index," Computer Vision and Image Understanding, vol. 94, pp. 193–233.

10) Young Deok Chun, Nam Chul Kim, Ick Hoon Jang,(2008) "Content-Based Image Retrieval Using Multiresolution Color and Texture Features," IEEE Transaction on Multimedia, vol. 10, pp. 1073-1084.

11) Tai X. Y., Wang L. D.(2008) , "Medical Image Retrieval Based on Color-Texture Algorithm and GTI Model, " Bioinformatics and Biomedical Engineering, , ICBBE 2008, The 2nd International Conference on, pp. 2574-2578.