Engineering Applications of Artificial Intelligence 26 (2013) 2372–2382

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Using visual features to design a content-based image retrieval method optimized by particle swarm optimization algorithm

Bae-Muu Chang [a], Hung-Hsu Tsai [b,*], Wen-Ling Chou [b]

[a] Department of Information Management, Chienkuo Technology University, Chang-Hua, Chang-Hua 500, Taiwan, ROC
[b] Department of Information Management, National Formosa University, Hu-Wei, Yun-Lin 632, Taiwan, ROC

ARTICLE INFO

Article history:
Received 21 September 2012
Received in revised form
4 July 2013
Accepted 24 July 2013
Available online 2 September 2013

Keywords:
Content-based image retrieval
Particle swarm optimization
Visual features
Similarities
Weights

ABSTRACT

This paper presents a content-based image retrieval method using three kinds of visual features and 12 distance measurements, which is optimized by particle swarm optimization (PSO) algorithm. For convenience, it is called the CBIRVP method hereafter. First, the CBIRVP method extracts three kinds of features: color, texture, and shape features of images. Subsequently, it employs appropriate distance measurements for each kind of features to calculate the similarities between a query image and others in the database D. Also, the PSO algorithm is utilized to optimize the CBIRVP method via searching for nearly optimal combinations between the features and their corresponding similarity measurements, as well as finding out the approximately optimal weights for three similarities with respect to three kinds of features. Finally, experimental results demonstrate that the CBIRVP method outperforms other existing methods under consideration here.

1. Introduction

With the constantly increasing number of digital images, an effective digital management and the retrieval method for a large number of files is very important. First, an image is specified by adhering a text annotation (a set of keywords) to the image. The image retrieval procedure is to match the text annotation of a query image with those of images in the image database. Subsequently, the method exhibits candidate images with similar text annotations. However, employing the text annotation is more impractical (Lu and Chang, 2007). The main problems are as follows: (1) When the database is growing more and more large, it is very time-consuming to necessarily mark up the text annotations for huge number of images. (2) Because of users′ different cognitions, they might mark up different text annotations for the same image. (3) It is very difficult to utilize the text annotations to clearly describe the diversity and the ambiguity of visual contents of an image. For these reasons, this paper proposes the techniques for content-based image retrieval (CBIR) to overcome the image retrieval problems of annotation-based methods mentioned above. Thus, the CBIR research has become an emerging issue over the past decade (Manjunath et al., 2001; Sim et al., 2004; Wang et al., 2011).

Earlier CBIR researches only employ a single feature, for instance, the color, the texture, or the shape feature. The method of Swain demonstrates that color histograms of multicolored objects provide an efficient information to index into a huge database of images (Swain and Ballard, 1991). To solve the identification problem, it presents a technique, histogram intersection, which matches model and image histograms, and a better version of histogram intersection, which achieves real-time indexing into a large database of stored models. Also, to solve the location problem, it provides a technique called histogram backprojection which efficiently performs this task in crowded scenes (Swain and Ballard, 1991).

The method of Manjunath whereas utilizes color and texture descriptors to undergo extensive evaluation and development (Manjunath et al., 2001). It employs several color descriptors which roughly include scalable color descriptor (SCD) and color structure descriptor (CSD), dominant color descriptor (DCD), and color layout descriptor (CLD). In (Manjunath et al., 2001), the texture descriptors contain a homogeneous texture descriptor (HTD) and an edge histogram descriptor (EHD). The HTD offers a quantized characterization of homogeneous texture regions for similarity matching. It calculates the local spatial-frequency statistics of image textures. The EHD captures the spatial distribution of edges, somewhat, the same principle as the CLD. The distribution of edges is a good texture feature which is useful for image retrieval when the underlying texture feature is not homogeneous.

However, an image contains various kinds of visual features. It is difficult to achieve satisfying retrieval results by using one or

* Corresponding author. Tel.: +886 937 749299.
E-mail addresses: bmchang@cc.ctu.edu.tw (B.-M. Chang), thh@nfu.edu.tw (H.-H. Tsai), hj206012001@yahoo.com.tw (W.-L. Chou).

0952-1976/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2013.07.018

two features. In recent years, there have been several related researches about the image retrieval methods by using the combination of visual features to enhance the papers mentioned above (ElAlami, 2011; He et al., 2010). Thus, in the CBIR, it is very important to extract the effective visual features and combine these extracted features. There have been some related researches for the combination of various features in recent years. The method of He utilizes the average values of $L^{*}U^{*}V$ color spaces and the 3-level DWT using db4 as the image features. In He′s paper, the Earth Mover′s Distance (EMD) measurement is employed as the similarity degree (He et al., 2010). The method of Wang utilizes the DCD, the steerable filter decomposition, and the pseudo-Zernike moments as the image features (Wang et al., 2011). The combination of color, texture, and shape features provides a feature set for image retrieval. The method of ElAlami utilizes the 3D color histogram and the Gabor filter algorithm to extract the color and the texture features (ElAlami, 2011). It employs the GA to select the features by reducing the feature dimensionality.

However, these papers mentioned above neither select a nearly optimal similarity function for each kind of visual features, nor search for three weights for a combination of three similarities. Thus, a novel color image retrieval method is proposed in this paper. It respectively employs various kinds of color, texture, and shape features to retrieve the images. In addition, it utilizes the PSO algorithm not only to select the nearly optimal combinations between the visual features and the similarities, but also to obtain the approximately optimized weights for the similarity function. More specifically, the paper develops a nearly optimal similarity function for matching two images, which is performed by a linear combination of three distances for three kinds of visual features. Moreover, the function is also optimized by the PSO algorithm, which is exploited to search for a nearly optimal combination including distance methods and weights for three kinds of visual features.

The remainder of this paper is arranged as follows: Section 2 introduces the related researches for feature extraction and reviews the PSO algorithm. Section 3 states the CBIRVP method in this paper. Section 4 exhibits experimental results. The concluding remarks are given in Section 5.

2. Backgrounds

2.1. Color feature extraction

Color is one of low-level visual features, which is mostly used in the applications of image processing. It has the characteristics of easy calculation and invariant in image scaling, rotation and translation. Namely, color features of an image can be represented by a set of k bins for each channel. Also, many proposed methods perform the similarity matching employing this kind of color features (Martinez, 2004; Swain and Ballard, 1991).

The MPEG-7 standard includes various well-defined descriptors for color low-level features (Martinez, 2004). Its applications have good results in image retrieval. The DCD represents main colors, such as red, green, and blue, for several sets of image contents and describes an image by using these colors which have been quantified (Manjunath et al., 2001). The SCD is defined in the HSV color space with fixed color quantization and employs the Haar transform encoding (Manjunath et al., 2001). The HSV color space is developed to obtain an intuitive representation of color and to nearly achieve the way in which humans perceive and process colors. It is uniformly quantized into 256 bins and respectively includes 16 levels in H, four levels in S, and four levels in V. The Haar transform encoding is convenient for a scalable representation of description, as well as the complexity scalability for feature extraction and matching process (Manjunath et al., 2001). The CLD is a compact color descriptor which employs representative colors in an 8×8 grid followed by a discrete cosine transform (DCT) and the YCbCr color space is exploited (Manjunath et al., 2001). The feature extraction can be derived as follows. An image is divided into 64 blocks. Average color of each block is regarded as its corresponding representative color. The average colors of 64 blocks are transformed into a series of coefficients by performing 8×8 DCT. Some low-frequency coefficients are chosen by using the zigzag searching and are quantized to represent a CLD (Manjunath et al., 2001).

2.2. Texture feature extraction

Texture is another kind of important low-level visual features. It can express the relationship between the innate surface properties of an obiect and the surrounding environment (Wang et al., 2011). Several popular texture feature descriptions are briefly computed in the applications of image retrieval. The method of Sim exploits the discrete Fourier transform (DFT) and the modified Zernike moments for invariant texture retrieval (Sim et al., 2004). First, the proposed descriptor is obtained by computing the power spectrum of DFT version of an original texture image for translation invariance. Second, the power spectrum image is normalized for scale invariance. Finally, the modified Zernike moments are claculated using normalized power spectrum images for rotation invariance. The method of Zhong is developed in the DCT domain (Zhong and Defe´e, 2005). Two histograms, quantized AC-patterns and DC direction vectors, are calculated to represent the features of images. A DCT block without DC coefficient is referred as an AC-pattern. The direction vectors for all blocks are employed in calculating DC direction vector histograms. For a 3×3 DCT block, eight differences between the central DC coefficient and its neighbors are calculated. The ninth difference is the difference between the central coefficient DC and the mean of nine coefficients of the block. The nine differences are descendingly ranked by their absolute values and the first γ direction-values are formed a direction vector for the block, where integer γ is in (Barel et al., 2010; Lu and Chang, 2007). The method of He first applies the DWT to extract the texture feature of an image (He et al., 2010). It then utilizes an image segmentation and the C-means algorithms to obtain six segmented regions of the image. For each segmented region, its color feature is extracted by calculating average for each domain in $L^{*}U^{*}V$ space. Moreover, the region′s texture feature is extracted by computing the mean and the standard deviation of 3-level DWT coefficients in each subband for the region. The method of Han proposes two new invariant representations, a rotation-invariant and a scale-invariant Gabor representations, for promoting conventional texture image retrieval (Han and Ma, 2007). A specific scale band covering all different orientations yields a rotation-invariant Gabor transformation. The texture feature for the corresponding specific scale band includes the mean and the standard deviation of the magnitude of coefficients of the rotation-invariant Gabor transformation. A specific orientation band covering all scales forms a scale-invariant Gabor transformation. The texture feature for the corresponding specific orientation band also contains the mean and the standard deviation of the magnitude of coefficients of the scale-invariant Gabor transformation.

An image is performed by a steerable filter decomposition to obtain several subband images at different orientations (Wang et al., 2011). The mean and the standard deviation of each subband image are computed to form the texture features of the image.