Pattern Recognition 47 (2014) 3293–3303

Contents lists available at ScienceDirect

Pattern Recognition

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

The method for image retrieval based on multi-factors correlation utilizing block truncation coding

Wang Xingyuan*, Wang Zongyu

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

ARTICLE INFO

Article history:
Received 16 December 2013
Received in revised form
18 March 2014
Accepted 18 April 2014
Available online 2 May 2014

Keywords:
Content based image retrieval
Block truncation coding
Multi-factors correlation
Structure element correlation
Gradient value correlation
Gradient direction correlation

ABSTRACT

In this paper, we proposed multi-factors correlation (MFC) to describe the image, structure element correlation (SEC), gradient value correlation (GVC) and gradient direction correlation (GDC). At first, the RGB color space image is converted to a bitmap image and a mean color component image utilizing the block truncation coding (BTC). Then, three correlations will be used to extract the image feature. The structure elements can effectively represent the bitmap which is generated by BTC, and SEC can effectively denote the bitmap's structure and the correlation of the block in the bitmap. GVC and GDC can effectively denote the gradient relation, which is computed by a mean color component image. Formed by SEC, GVC and GDC, the image feature vectors can effectively represent the image. In the end, the results demonstrate that the method has better performance than other image retrieval methods in the experiment.

1. Introduction

In recent years, image databases are increasing fast. It is difficult to capture useful images in a huge image database, and it becomes worse that we frequently obtain useful images through the Internet which is widely spreading. Existing methods are usually difficult to change, so it is urgent to find a new way to retrieve images accurately. Thus image retrieval becomes an important topic in image processing and pattern recognition. Generally speaking, images can be retrieved in three ways: text based, content based and semantic based [1–8]. The text based retrieval approach is used widely, such as, Baidu and Google. With this method we can retrieve images using keywords that are annotated on the images. However, we often obtain images that are unrelated to our expected results, and the results of the image retrieval rely on our understanding of the query images. There are two drawbacks to this approach. Firstly, images in the database are annotated manually by annotators; it is time-consuming for a huge image database and requires much human labor for annotation. Secondly, the results of the retrieval are inaccurate, because the results of the retrieval are related to an understanding of the query images. The second approach, i.e. content based image retrieval (CBIR), was proposed in the early 1990s [9–16]. This approach is to retrieve images using low-level features like color, texture and shape that can represent an image. This approach involves querying an example image first, extracting the low-level features on the example image, then computing the similarity between the query image and the images in the image dataset and, finally, sorting images by similarity and displaying the top images. CBIR has been shown to be much more effective and subjective than the text based approach [17–19]. The final approach is the semantic based method. The CBIR also fails to describe the semantic concepts, so researchers proposed some methods for image retrieval using relevance feedback algorithms. The relevance feedback algorithms capture a user's preferences and bridge the semantic gap [20,21]; the results of using this method which is based on relevance feedback is more approach human perception.

In CBIR, color, texture and shape are the most important features. HSV color space is widely used in extracting color features. In this space, hue is used to distinguish color, saturation is the percentage of white light added to a pure color, and value refers to the perceived light intensity [4,22]. The advantage of HSV color space is that it is closer to human conceptual understanding of colors. In order to cut down computing complexity and extract the color features in an efficient way, the HSV color space is quantized to 72 bins, in general [22].

Generally, in CBIR, the descriptors representing the image are based on color, shape or texture and are used to describe the image. Various algorithms have been designed to extract features for image retrieval. The multi-texton histogram (MTH) proposed for image retrieval integrates the advantages of a co-occurrence matrix and a histogram by representing the attribute of co-occurrence matrix using a histogram [1]. A novel image feature detecting and describing method is proposed, called micro-structure descriptor (MSD). The MSD is built based on the underlying colors in micro-

* Corresponding author.
E-mail addresses: wangxy@dlut.edu.cn (X. Wang), chenxizongyu@163.com (Z. Wang).

http://dx.doi.org/10.1016/j.patcog.2014.04.020
0031-3203/

structures with a similar edge orientation [2]. In [3], the author is using the block color co-occurrence matrix and the block pattern histogram to represent an image. In [8], a structure element' descriptor is proposed, and a structure histogram is used to extract an image feature. In [9], three types of image features are proposed to describe the color and spatial distributions of an image; in these features, the K-means algorithm is adopted to classify all the pixels in an image into several clusters according to their colors. The MPEG-7 edge histogram descriptor (EHD) is extracted on spatial distribution of edges, and it is an efficient texture descriptor for images [23]. The edge orientation autocorrelogram (EOAC) is proposed for a shape based descriptor [24]. A very effective method has been proposed to detect and describe local features in images, called scale-invariant feature transform (SIFT) [25]. The texton co-occurrence matrices (TCM) can describe the spatial correlation of textons for image retrieval which is proposed in [26]. In [27], an adaptive color feature extraction scheme is proposed by considering the distribution of an image; the binary quaternion-moment-preserving (BQMP) threshold technique is used.

In this paper, we proposed multi-factors correlation (MFC) to describe an image, structure element correlation (SEC), gradient value correlation (GVC) and gradient direction correlation (GDC). In the first place, an RGB color space image is converted to a bitmap image and a mean color component image utilizing the block truncation coding (BTC). Then, three correlations will be used to extract the image feature. The structure elements can effectively represent the bitmap which is generated by BTC, and SEC can effectively denote the bitmap's structure and the correlation of the block in the bitmap. GVC and GDC can effectively denote the gradient relation, which is computed by a mean color component image. The image feature vectors are formed by SEC, GVC and GDC and they can effectively represent the image. In the end, experiments showed that the MFC method has a higher retrieval precision and recall ratio than the MSD [2], SED [8] and CSD3 [9] methods.

The rest of the paper is organized as follows. Section 2 details the BTC for a color image. Section 3 describes the SEC and the method for extracting the feature vector in a bitmap. In Section 4, the GVC and GDC are described. The similarity measure approach is defined in Section 5. The experimental results and comparisons are presented in Section 6. And Section 7 is the conclusion of the paper.

2. BTC for color image

Block truncation coding (BTC) was first proposed in 1979 [28] and it is mainly used for compressing gray-scale images. BTC divides the image into small nonoverlapping block images. In every block, after block coding, the original image has generated a bitmap and two mean values. In BTC, firstly, the means of every block is computed. Then, compared with the mean value, the pixels in the block will have a value of 1 in the bitmap if the pixel value is greater than the mean value; otherwise it will have a value of 0. In BTC, the two mean values are the mean value of the pixels which is greater than the block mean value and the mean value of the pixels which is less than the block mean value and they are called max mean value and min mean value, respectively. At the decoding stage, in the bitmap, the value is replaced by max mean value if the bitmap has a value of 1; otherwise it is replaced by the min mean value replaced. Therefore, the BTC is the lossy compression algorithm.

The BTC is used in a gray-scale image in the early period and then extended in a Multispectral image, such as a color image [3,29,30], gradually. Most of the color images used the RGB color space which is widely used for representing the color image. After the RGB color space, there are many color spaces, such as YCbCr and HSV.

The steps of the BTC for color image can be described as follows [3,29]:

1. Divide the image into small nonoverlapping block images and the size of every block is $m \times n$; compute the mean value $a(i,j)$, $(i \in [1,m],\ j \in [1,n])$ of every pixel in the block by Eq. (1)
$$a(i,j) = (1/3)(r(i,j)+g(i,j)+b(i,j)) \quad i \in [1,m], j \in [1,n] \tag{1}$$
2. Compute the mean value Th of every block as the threshold.
$$Th = \frac{1}{m \times n} a(i,j). \quad (i \in [1,m],\ j \in [1,n]) \tag{2}$$
3. The bitmap of every block can be computed as follows:
$$bimp(i,j) = \begin{cases} 1, & a(i,j) > Th. \\ 0, & a(i,j) \le Th. \end{cases} \quad (i \in [1,m], j \in [1,n]) \tag{3}$$
4. Compute the max mean value $M_{\max}$ and the min mean value $M_{\min}$ by
$$M_{\max} = \begin{cases} \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j)} \sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j) \times r(i,j). \\ \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j)} \sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j) \times g(i,j). \\ \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j)} \sum_{i=1}^{m}\sum_{j=1}^{n} bimp(i,j) \times b(i,j). \end{cases} \tag{4}$$

Fig. 1. Bitmap of the image. (a) Original image. (b) Bitmap of the image.