Expert Systems With Applications 96 (2018) 347–357

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Taking advantage of multi-regions-based diagonal texture structure descriptor for image retrieval

Wei Song[a,b], Yubing Zhang[a,c], Fei Liu[a], Zhilei Chai[a,b], Feng Ding[a], Xuezhong Qian[a,b], Soon Cheol Park[d,*]

[a] School of Internet of Things (IOT) Engineering, Jiangnan University, Wuxi 214122, PR China
[b] Engineering Research Center of Internet of Things Technology Applications, Ministry of Education, Wuxi 214122, PR China
[c] State Grid Tongling Electric Power Supply Company, Tongling 244000, PR China
[d] Department of Electronics and Information Engineering, Chonbuk National University, Jeonju 561756, Republic of Korea

ARTICLE INFO

Article history:
Received 22 April 2016
Revised 2 November 2017
Accepted 3 December 2017
Available online 6 December 2017

Keywords:
Diagonal texture structure descriptor
Multi-regions
Receptive field
Visual system
Content-based image retrieval

ABSTRACT

With the rapid development of information technology, the capacity of Web images database becomes larger and larger. How to quickly and effectively find the desired images in Web image databases becomes the challenge needed to resolve with high priority. In this paper a novel diagonal texture structure descriptor (DTSD) is proposed, and a new framework considering hue, saturation and value components is utilized for image retrieval. In specific, we firstly use Otsu algorithm to segment image into foreground and background, and the features of multi-regions are respectively considered. That is, we present the contents of these multi-regions distinctively to reduce the influence of each other, which would perform hierarchical feature description and realize more accurate content match for image retrieval. In this study, to simulate the characteristic of human eyes for perceiving colors, hue and saturation components are quantized into various bins which can obtain more detailed description for color difference. Meanwhile, DTSD is extracted based on value component to represent the edge information as the feature of receptive field. Such a method can improve the spatial resolution ability of the descriptor, and identify finer structure of an image. Moreover, histogram with respect to these three components, i.e., hue, saturation and value, is utilized to generate the feature vector of an image. We carry out the experiments on benchmark Corel and UCID image datasets, and the extensive experimental results demonstrate that our method achieves better performance in comparison with state of the art image retrieval algorithms. The proposed method is very promising, which can provide more accurate retrieved results on the basis of color & texture descriptions in multi-regions, and further enhances the performance of the intelligent image retrieval system.

1. Introduction

In recent years, with the development of multimedia and computer technology, the size of digital library is increasing rapidly. The handling of these datasets by human annotation is an extremely difficult work. How to find desired image information efficiently from large image databases has become a challenge needed to resolve urgently. Image retrieval has a wide range of applications in several expert systems. As medical image has become an indispensable tool in modern clinical diagnosis, hospitals produce a large number of medical images every day. Medical image expert retrieval system can quickly return a set of similar medical images corresponding to the input query image, which is conducive to make a systematic diagnosis to the related diseases. In the field of satellite remote sensing, with the increase of the number of transmission satellites and the raise of available bands, it generates massive remote sensing images. Remote sensing image expert retrieval system can quickly locate the disaster areas through automatic comparison of image features, and help experts make rescue strategies in time. In addition, trademark image expert retrieval system can find similar trademarks immediately after entering a new trademark image for registration, which can effectively protect the legitimate rights and interests of registered trademarks. In recent decade, image retrieval has become a research hot spot in the fields of pattern recognition, artificial intelligence, and expert systems.

* Corresponding author.
E-mail addresses: songwei@jiangnan.edu.cn (W. Song), ybzhang@jiangnan.edu.cn (Y. Zhang), fliu@jiangnan.edu.cn (F. Liu), zlchai@jiangnan.edu.cn (Z. Chai), fding@jiangnan.edu.cn (F. Ding), qxzvb@163.com (X. Qian), goldenview@foxmail.com (S.C. Park).

https://doi.org/10.1016/j.eswa.2017.12.006
0957-4174/

The typical image retrieval technologies can be extensively divided into text-based (Ma, Zhu, Lyu, & King, 2010), content-based (Akakin & Gurcan, 2012; Gudivada & Raghavan, 1995; Yildizer, Balci, Hassan, & Alhajj, 2012) and semantic-based (Djordjevic & Izquierdo, 2007; Sun, Li, Li, Ming, & Cai, 2005; Thuy, Huu, Van, & Quoc, 2017) methods. Text-based image retrieval is a traditional searching method which deals with key words and metadata of images, and the frequently used text descriptors include image names, size, generated and revised time, etc. Text-based image retrieval method has to annotate image manually. However, the current web image databases usually become larger and larger, and the annotation accuracy is subjective to human perception. Thus, it spends much time and requires much human labor. Content-based image retrieval (CBIR) (Aptoula & Lefèvre, 2009; Liu, Zhang, Lu, & Ma, 2007) was first appeared in the 1990s. Compared with the text-based method, CBIR utilizes visual features, e.g., color, texture distribution, object shapes, and spatial orientation of objects, to achieve the goal of indexing and describing images. Several famous CBIR systems are QBIC (Flickner et al., 1995), MARS (Ortega, Rui, Chakrabarti, Porkaew, & Mehrotra, 1998), Photobook (Pentland, Picard, & Sclaroff, 1996) and Visual-SEEK (Brandt, Laaksonen, & Oja, 2002). The third kind of approach is semantic-based method, which includes the objective description of image content and the subjective feelings on which people concentrate. Since such a kind of approach is too difficult to find a direct connection between high level semantics and low level features, an issue named semantic gap (Zhang, Islam, & Lu, 2012; Zhao & Grosky, 2002) is emerged in this field. To tackle with the issue of semantic gap, some researchers used machine learning methods, such as support vector machine (SVM) (Rahman, Antani, & Thoma, 2011; Seo, 2007; Wu, & Chung, 2009), and feedback mechanism (Guo, Jain, Ma, & Zhang, 2002), to reveal the relationship between descriptors and corresponding semantics. For the sake of capturing image semantic concepts precisely, Wang, Chen, and Yang (2011) present a new integrated SVM classifiers based on relevance feedback and further utilize it to image retrieval. However, the parameters selection of kernel function affects the retrieval result greatly and results in the unstable performance. Thus, the fast and efficient CBIR methods are still the requirement and research hot point in this field.

It has been known that the underlying features, e.g., color, texture, and shape, are considered to establish index for retrieval. However, the existed general CBIR methods may ignore the fact that the different regions of image would attract the different attentions to human eyes. To deal with this issue, the regions of interest are taken into account. That is to say, such a kind of method divides image into several regions via image segmentation, and extract the features focused on target region while ignoring other unconcerned regions. It is intuitive that, such a kind of methods can achieve good performance for images with outstanding target region. However, it is difficult for them to tackle with well retrieval for images of complex background, because some background regions may also contain rich contents. In this paper, we first use Otsu algorithm to divide an image into foreground and background, and the features of multi-regions are respectively considered. That is, we represent the contents of these multi-regions distinctively to reduce the influence of each other, which would perform hierarchical feature description and realize more accurate content match for image retrieval.

In addition, in the visual system the individual cell at any level or state on the retina has its particular representative area. The optical stimulation in this area can affect the activity of nerve cells, and this area is defined as the visual receptive field of cells (Kuffler, 1953). It has known that, the smaller receptive field can improve the ability of spatial resolution. Thus, from the view of human eye's visual characteristics, the scene can be divided into several small receptive fields. Hubel and Wiesel (1968) presented the classical receptive field feature detection theory of visual system, which found that the content extracted by human retina is the edge information based on luminance contrast. Inspired by such a theory, we proposed a new diagonal texture structure descriptors (DTSD) based on value component to describe the edge information of receptive field in scene. That is, the brightness variation relations in such a diagonal texture structure are considered to represent the edge information as the feature of receptive field. DTSD can improve the spatial resolution ability of the descriptor, and identify finer structure of an image.

The rest of this paper is organized as follows. Section 2 introduces the related work of this study. In this section, commonly used image feature descriptors are extensively introduced, and Otsu threshold segmentation algorithm is further described. Otsu can effectively divide an image into corresponding foreground and background as the multi-regions. Based on the analysis of Section 2, a new DTSD and framework for the proposed retrieval method are presented in Section 3. The experimental results and analysis are given in Section 4. The conclusions and future research directions are given in Section 5.

2. Related works

2.1. Image feature descriptors

It has been known that the key technology of CBIR is how to define effective features to represent the content of image. Swain and Ballard (1991) presented the idea of color histogram which is regarded as the global color descriptor in CBIR system. Color histogram is commonly used because of its simplicity, but the spatial association of color is ignored in this method. To tackle with this issue, Liu and Yang (2013) proposed color difference histogram (CDH), which counts the perceptually uniform color difference with regard to colors and edge orientations in L*a*b* color space. However, it is difficult to convert RGB color space to L*a*b* color space and results in very huge consumption of calculation. As one of typical visual features, texture feature captures the intrinsic surface characteristics of image and can reveal the characteristic relations. Many techniques have been put forward to describe texture features, such as gray level co-occurrence matrices (Gadelmawla, 2004), the Tamura texture feature (Tamura, Mori, & Yamawaki, 1978), Gabor filtering (Manjunath & Ma, 1996), and local binary patterns (Ojala, Pietikäinen, & Mäenpää, 2002). In practice, texture features can be used in conjunction with color features to improve discrimination power. One of the most commonly used methods is to integrate gray-level textures and color features. In Liu, Zhang, Hou, Li, and Yang (2010), multi-texton histogram (MTH) is proposed by integrating the co-occurrence matrix and histogram as one descriptor. Such a method can express both the spatial correlation of texture orientation and texture color. However, the five kinds of structure of this method cannot fully describe the variety of structures of natural images. In Liu, Li, Zhang, and Xu (2011), a micro-structure descriptor (MSD) is proposed, which represents an image in the light of edge orientation similarity. MSD fuses color, texture, & shape information, and extracts features by simulating human visual processing. However, it lacks global properties of the image, and it is unable to utilize relation between locations of different objects in layout. Wang and Wang (2013) proposed a structure elements' descriptor (SED) which takes advantage of five structure elements to denote five directions. SED can represent the spatial correlation of color and texture. However, this descriptor is unable to represent region based properties of images. The hybrid information descriptors (HID) include mutual information descriptors and self-information descriptors. HID can characterize image by imitating