J. Vis. Commun. Image R. 38 (2016) 73–81

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A novel image retrieval method based on multi-trend structure descriptor ☆

Meng Zhao, Huaxiang Zhang *, Jiande Sun

School of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong Province, China

ARTICLE INFO

Article history:
Received 5 November 2015
Revised 24 February 2016
Accepted 29 February 2016
Available online 3 March 2016

Keywords:
Multi-trend structure descriptor
Image retrieval
Visual perception mechanism
Local descriptor

ABSTRACT

This paper proposes an image feature representation method, namely Multi-Trend Structure Descriptor (MTSD), which is built based on the local and multi-trend structures. The local structures can be regarded as the basic units for image analysis, and the multi-trend structures are introduced to explore the correlation among pixels in local structures according to the information change of pixels. The visual information such as color, edge orientation and intensity map are considered and quantized, and with the local structure as a bridge, we use multi-trend to detect color, edge orientation and intensity map respectively for feature extraction. MTSD can characterize not only the low-level features, such as color, shape and texture, but also the local spatial structure information. We evaluate the performance of the proposed algorithm on Corel and Caltech datasets, and experimental results demonstrate that, MTSD significantly outperforms texton co-occurrence matrix, multi-texton histogram, micro-structure descriptor and saliency structure histogram.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Digital images are one of the most important media formats and provide a large amount of information for human communication. With the expansion of images, image retrieval has become an active area in the field of artificial intelligence and pattern recognition [1]. In recent decades, in order to propose effective and efficient image retrieval methods for large image collections [2], more and more researchers devote themselves to this topic. Text-based image retrieval (TBIR) is a traditional searching approach which deals with keywords and image annotations in databases [3]. It is very sensitive to the keywords annotated by users, and relies on the understanding of query images. To overcome the difficulties of TBIR, content-based image retrieval (CBIR), which is more effective and subjective, was proposed in 1990s [4]. CBIR techniques commonly rely on the low-level features for image representation and retrieval [5], and these features can be extracted from images to form feature vectors by various descriptors. CBIR systems retrieve some of the most similar images to users based on the similarity between the feature vectors between query images given by users and the database images. Until now, CBIR feature descriptors can be divided into two categories: global feature descriptors and local feature descriptors. Global feature descriptors aim at the whole image, while local feature descriptors pay more attentions to the key points and the salient patches in images. Many researchers believe the information that key points and salient patches contain is very vital for the process of human visual perception. Generally speaking, local feature descriptors are considered to have higher discriminative power than global feature descriptors [6].

However, because there is a semantic gap between the low-level features and high-level concepts [7,8], the performance of CBIR techniques is limited. It is well known that high-level concepts are the human visual perception based on images. A lot of researchers have adopted machine-learning techniques to solve this problem, and many classical methods are proposed, such as relevance feedback algorithms [9] and dictionary learning [10]. In addition, many researchers believe that it is suitable to utilize local spatial structure information to improve the image retrieval performance, because it can represent image content precisely and bridge the semantic gap as much as possible [11].

In this paper, we propose a novel feature descriptor for CBIR, called Multi-Trend Structure Descriptor (MTSD). It is a unified framework based on the local and multi-trend structures. MTSD utilizes information change of pixels to represent the local spatial structure information for image feature extraction. The rest of this paper is organized as follows. Related works are introduced in Section 2, and the proposed MTSD scheme is described in Section 3.

☆ This paper has been recommended for acceptance by Y. Li.

* Corresponding author.
E-mail address: huaxzhang@hotmail.com (H. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2016.02.016
1047-3203/© 2016 Elsevier Inc. All rights reserved.

Section 4 shows the experimental results and Section 5 concludes the paper.

2. Related work

The critical issue for CBIR is to look for effective and efficient features to represent images. According to different visual information of images, low-level features can be categorized as color features, texture features and shape features, and various image feature descriptors have been designed based on these features. Among these features, color features are considered to be one of the most significant visual features and have a deep influence on human visual perception. Color histogram is invariant to orientation and scale, and is widely used in CBIR systems for its low computational complexity and effectiveness. In order to describe image spatial characteristics, compact color moments, color coherence vector and color correlograms are proposed [12]. In the MPEG-7 standard, there are many color descriptors, such as dominant color descriptor, color layout descriptor (CLD) and scalable color descriptor (SCD) [13]. Texture features depict the information of smoothness and coarseness [14], and can be considered as an inherent property of image regions. Several texture descriptors have been developed for texture analysis, including gray level co-occurrence matrices [15], the Markov random field (MRF) model [16], Gobor filtering [17] and the local binary pattern (LBP) [18]. In MPEG-7 standard, texture features can be derived via approaches such as homogeneous texture descriptor, texture browsing descriptor and the edge histogram descriptor [19]. In addition to color and texture features, shape features can be characterized by edge orientation. Moment invariants [20] and Fourier transforms coefficients [21] are the classical shape descriptors. In MPEG-7 standard, 3-D shape descriptor and curvature scale space (CSS) descriptor are used to extract shape features [19].

In the past few years, scale-invariant feature transform (SIFT) descriptor [22] has been considered to be the most popular local feature detector. It is famous for the reliable matching between different scenes. In order to reduce the computational complexity, Principal Component Analysis-SIFT (PCA-SIFT) [23] and Speeded Up Robust Features (SURF) [24] are designed as efficient alternatives. Recently, bag-of-visual words (BOW) model has become a new research hotspot in the field of image retrieval and object recognition [25]. In the BOW model, many key-point detectors are used for local feature extraction, such as SIFT and SURF. Finally, an image can be represented as a set of visual words, which is called the standard BOW baseline. However, due to the lack of explicit semantic information and the ambiguity of visual words, BOW model may lead to false matches. In order to overcome the major drawbacks, many methods are proposed to improve the visual vocabulary and combine spatial information, such as coupled binary embedding method [26] and the bag-of-colors framework [27]. Besides, BOW model often results in heavy computational burdens because of the learning processes or the clustering implementation. Lots of approaches are designed for computational efficiency, such as optimized product quantization [28], query Pruning and early termination strategies [29].

In addition, many researchers point out that local structures can describe images effectively [30]. According to this view, lots of algorithms considering local structures have been designed. Texton co-occurrence matrix (TCM) can express the spatial correlation of textons and calculate the statistical information of textons for image feature description [31]. Multi-texton histogram (MTH) integrates the advantages of co-occurrence matrix and histogram together to describe the attribute of co-occurrence matrix using histogram [32]. Micro-structure descriptor (MSD) is based on the underlying colors in micro-structures with similar edge orientation. It integrates color, texture, shape and layout information as a whole [33]. Structure elements' descriptor (SED) can extract color and texture features, and is proposed for representing image local features [34]. Color difference histogram (CDH) combines edge orientation and color features together, and calculates the uniform color difference between two points under different backgrounds [35]. Hybrid information descriptors (HIDs) extract features among different image feature spaces with image structure and multi-scale analysis [36]. Local extrema co-occurrence pattern (LECoP) extracts the local directional information from local extrema pattern, and converts it into feature vectors with the use of gray level co-occurrence matrix [37]. Saliency structure histogram (SSH) combines color volume and edge information to detect saliency regions, and embeds oriented Gabor filters into bar-shaped structures for image representation [38].

In this paper, a novel image feature describing method, called multi-trend structure descriptor, is proposed for CBIR. The proposed method extracts image features from multiple perspectives and pays more attention to the local spatial structure information.

3. Multi-Trend Structure Descriptor (MTSD)

Juleszs textons theory indicates that images consist of certain local structures which often show a certain amount of similarity in analogous images. As a result, local structures can be regarded as an essential image property. By means of local structures, low-level features blend well in images. Thus, it is suitable to use local structures to mine the internal image correlations. In this paper, the proposed algorithm MTSD integrates color, edge orientation and intensity information as a whole for image representation.

3.1. Visual feature extraction and quantization

Lots of visual information is hidden in images. In our proposed method, color, edge orientation and intensity information are selected for feature representation.

Color information is extracted in HSV color space, because it is considered to be the most suitable color space to mimic human's visual system. There are three components in HSV color space: Hue (H), Saturation (S) and Value (V). Based on the theory that human eyes cannot distinguish large numbers of colors at the same time, color quantization is used in our method. The purpose of color quantization is to assign a certain set of colors to represent an image with maximum useful information. We divide H, S and V color channels into 12, 3 and 3 bins, respectively, thus resulting in $12 * 3 * 3 = 108$ combinations obtained in total.

For a given full color image $g(x, y)$ of size $m * n$, we transform it from the RGB color space to HSV color space. The color map M_C can be obtained after color uniform quantization. For a pixel located on (x, y), where $0 \leqslant x \leqslant m-1$, $0 \leqslant y \leqslant n-1$, we define $M_C(x, y)$ as the color quantized value, and $M_C(x, y) = \alpha$, $\alpha \in (0, 1, \ldots, N_C - 1)$, where $N_C = 108$.

Intensity information is given by the V color channel, which represents the brightness of color in HSV color space. By means of intensity uniform quantization, the intensity map M_I is obtained. We define $M_I(x, y)$ as the intensity quantized value in the position of (x, y), where $0 \leqslant x \leqslant m-1$, $0 \leqslant y \leqslant n-1$, and $M_I(x, y) = \beta$, $\beta \in (0, 1, \ldots, N_I - 1)$, where $N_I = 20$ in this paper.

Edge orientation is one of the most widely used shape descriptors, and it can represent the object boundaries and describe the image contents. Many operators can be used for edge orientation detection, such as Sobel operator, Prewitt operator, Robert operator and LOG operator. In this paper, we select Sobel operator to obtain the edge orientation map M_O, because it has the lower computational complexity and better stable performance than the other

Download English Version:

https://daneshyari.com/en/article/529653

Download Persian Version:

https://daneshyari.com/article/529653

Daneshyari.com