J. Vis. Commun. Image R. 25 (2014) 1574–1587

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

A novel image retrieval method based on hybrid information descriptors

Ming Zhang [a,b], Ke Zhang [a,b], Qinghe Feng [a,b], Jianzhong Wang [a,b], Jun Kong [a,b,*], Yinghua Lu [a,b,*]

[a] School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China
[b] Key Laboratory of Intelligent Information Processing, Northeast Normal University, China

ARTICLE INFO

Article history:
Received 15 November 2013
Accepted 25 June 2014
Available online 14 July 2014

Keywords:
Hybrid information descriptors (HIDs)
Mutual information descriptors (MIDs)
Self information descriptors (SIDs)
Visual perception mechanism
Image feature spaces
Multi-scale analysis
Content-based image retrieval (CBIR)
Feature fusion

ABSTRACT

In this paper, we propose a novel image retrieval method called hybrid information descriptors (HIDs) consisting of mutual information descriptors (MIDs) and self information descriptors (SIDs). Based on the physiological structure of human eyes and visual perception mechanism, HIDs are designed to explore the internal correlations among different image feature spaces with image structure and multi-scale analysis, not only characterizing the low-level features, such as color, shape and texture, but also imitating the process of visual information transfer and perception in high-level understanding with the help of the proposed visual optimization model for feature fusion. Comparing with other existing methods applied to content-based image retrieval (CBIR) on four datasets, the usefulness and effectiveness of the HIDs are shown. Extensive experimental results can also demonstrate this.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Nowadays, the topic about image database and visual information retrieval has become an active research area [1,2]. To overcome the difficulties of text-based image retrieval system (TBIR), which is very sensitive to the keywords employed by the user, content-based image retrieval (CBIR) was first proposed in the early 1990s [3]. It has attracted wide interests in the field of computer research in the past decade and has become dominating tendency in the information retrieval area [4]. Therefore, more and more researchers pursue on it.

The image retrieval strategy based on image content attributes can solve the problem that men subjectively describe image as fundamental, which is the essential change against TBIR. Generally, image content feature description has included two levels: low-level features and high-level features. The former includes the low-level visual features, such as, color, shape and texture; the latter is the underlying and abstract semantic features based on the image content controlled by human visual perception mechanism, for instance, image content association, emotion, action, etc. [5,6] take full advantage of visual content such as colors, textures, and shapes to analyze the image semantic feature.

It is well known that the performance of CBIR systems is mainly limited to the semantic gap between low-level features and high-level semantics, so finding out the desired images from multimedia databases still remains a challenging issue. In addition, the sensor gap between the real world object and machine learning is also an unsettled issue.

To reduce this 'semantic gap', Cross-Media retrieval methods which integrate various data with their related attributes to represent high-level semantics are introduced as a powerful tool to enhance the search performance [7]. Wang et al. [32–34] takes full account of multi-modal spectral clustering, sparse multimodal machine, and structured sparsity into heterogeneous feature fusion and learning. The method based on harmonizing hierarchical manifolds constructs the MMD semantics to perform Cross-Media retrieval in [31]. However, how to describe image content objectively and effectively is playing an important role for CBIR.

Some researchers point out that image retrieval driven by human visual perception strategy may validly represent image content, and improve retrieval performance and efficiency [5]. Some researchers simulate the mechanisms of the primary visual cortex [8–10] to extract low-level features for CBIR in the last few years. Recently, it has increasingly become one of the research trends, focusing on content-based image retrieval technology domain. Hence, how to make use of human visual perception mechanism for image feature description becomes a crucial topic.

This paper proposes a set of novel features called hybrid information descriptors (HIDs) for content-based image retrieval task. HIDs-based method is a unified framework when using the mutual

* Corresponding authors at: School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China.
E-mail addresses: zhangm545@nenu.edu.cn (M. Zhang), kongjun@nenu.edu.cn (J. Kong).

http://dx.doi.org/10.1016/j.jvcir.2014.06.016
1047-3203/© 2014 Elsevier Inc. All rights reserved.

information descriptors (MIDs) and the self information descriptors (SIDs) for image retrieval. HIDs can characterize image by imitating the human visual perception mechanism, and implement the image retrieval efficiently. As an image feature descriptor, it can be further integrated into Cross-Media retrieval. The remainder of this paper is organized as follows: Section 2 introduces the related theories about human visual perception mechanism, and reviews the classical CBIR descriptors. What's more, HIDs method is proposed in the next section, including feature extraction and fusion strategy. In Section 4, the HIDs-based retrieval experiments and performance evaluation are reported. At last, conclusions are given in Section 5.

2. Related theoretical basis

2.1. Human visual perception mechanism

Retina has the ability to preprocess a series of visual information to our eyes before forming the visual perception. The direct sensitivity to light is the cods and cones in the retina. Cones with superior distinguishing ability are responsible for the perception of color, while rods are sensitive to dim light and provide the impression of the whole vision [6]. Once the sensate light from the combination of cones and rods transforms into neural signals, they will competitively interact to select relevant vision and suppress irrelevant stimulants by other neurons of the retina [11], then the neural signals evolve into action potentials in retinal ganglion cells, furthermore form vision by transmitting observed information along visual pathway, finally achieving percept in the brain in the form of reconstructed image.

The source of the visual attention is the distinctive difference of visual information features. Various color, shape, texture, as well as the change of a mobile velocity can result in visual difference, even visual attention. Visual attention and perception are similar to feature extraction and fusion processes in pattern recognition fields. In view of this analysis, this paper attempts to apply machine learning theories and methods to simulate human visual attention and perception.

Psychological and neurological studies have proved and most researchers also agree with two-stage model hypothesis on human visual perception mechanism proposed by Treisman [12–17]. He pointed out that human visual process includes two stages: a pre-attentive stage for visual attention and an attentive stage for visual perception. At the pre-attentive stage, all low-level visual information features are concurrently extracted and roughly expressed by retina cells, and they serve as the bases of feature fusion process. At the attentive stage, the importance factors of the visual feature are balanced according to different stimulators, and then all of the bases are re-combined to reconstruct objects. Furthermore, Marr had elaborated the visual reconstruction in [18].

In addition, capturing and perceiving information by human are the procedure of the whole description to local observation in details. This paper exploits multi-resolution processing and features extraction to represent it to go further in the machine learning and pattern recognition area. Since 1980s, the existing theories have seemingly demonstrated the correctness of our idea. Daugman [19] has brought forward that the space responding to Gabor filters is very similar to the mechanism of mammal visual perceiving. Han and MA [20] proposes the Gabor features—the classical exemplification to simulate human to use multi-resolution features extraction—can be against rotation and scaling for image retrieval.

2.2. Image feature descriptors

The key technology of content-based image retrieval is to look for certain suitable and efficient features to retrieve. Various feature descriptions for CBIR have been proposed. From the perspective of visual sense, they can be divided into color features, shape features and texture features.

Image color features are one kind of the dominant visual features, and the important source of visual difference. Color features include color set, color coherence vector, color moments, color histogram descriptor (CHD) and color layout descriptor (CLD) [21] in MPEG-7, etc. However, cones' sensitivity to color is not fully effective in the dim environment, but human can identify the object from different shapes with the help of rods.

Generally, edge, orientation and structure can characterize shape. The widely applied edge detectors are Sobel operator, Canny operator, LOG operator and so on, which can extract the structure of image. To resist the variety in scale, rotation, distort, etc., researchers utilize the normalized inertia, Zernike moments, the histogram of edge direction, the edge map [22] and MPEG-7 edge histogram descriptor (EHD) [23] to represent the shape information.

Texture feature is the internal visual feature of the object surface, which includes the main information of the object structure arrangement and its relationship with surroundings. Meanwhile, texture can be considered as the pattern generated from the gray (color) in the form of a certain change with the space and one inherent property of the image region. Obviously, together with color and shape, texture has the property of statistics, structure and spectrum. To characterize image further, texture features can be derived via such approaches as the gray-level co-occurrence matrix, the Tamura feature, wavelet coefficients, and Gabor filter-based features [24].

In order to be suitable for based-content image retrieval system, the visual feature descriptors above usually are combined in the form of parallel, sequence, or integration to describe image content [25]. In the guidance of parallel method, [26] analyzes color and texture respectively, and proposes wavelet based features with the application of CBIR. Micro-structure descriptor [27] adopts sequential approach, builds image units based on statistical texture, then expressed with indexed color layout information. Some methods integrate local descriptors to address the problem of image search, like the aggregating local descriptors [28] and RootSIFT [29], can retrieve in large scale image datasets.

In practice, for CBIR system, integrative method is more effective, moreover, this paper analyzes image features from multiple perspectives, take the internal correlation among various image feature spaces into account, and make use of color and shape feature representation in parallel and interaction, aiming at acting up to the visual perception rule for the image retrieval. It is the original intention of the proposed HIDs.

3. Hybrid information descriptors for content-based image retrieval

Julesz [30] indicates that image texture consists of certain structures, and congener images have similar texture structures. Liu et al. [27] expands this idea of structure to natural image. As a result, the structure is essential property of an image, and can be used for describing the image contents effectively. Color features and shape features act as the important role in content-based image retrieval task. When put into practice, both the description of color feature and shape feature can effectively characterize the image structure. Furthermore, they are interdependent and blending well in nature, and can be used to mine the internal correlations of features. Therefore, under the physiological structure and property of human eyes and human visual perception mechanism, taking the shape information into account when perceiving color features, as well as the color information based on shape features;

Download English Version:

https://daneshyari.com/en/article/529786

Download Persian Version:

https://daneshyari.com/article/529786

Daneshyari.com