J. Vis. Commun. Image R. 25 (2014) 1308–1323

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Content-based image retrieval using local visual attention feature

Hong-Ying Yang [a,*], Yong-Wei Li [a], Wei-Yi Li [a], Xiang-Yang Wang [a,b,*], Fang-Yu Yang [a]

[a] School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, PR China
[b] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, PR China

ARTICLE INFO

Article history:
Received 8 August 2012
Accepted 8 May 2014
Available online 21 May 2014

Keywords:
Image retrieval
Salient point
SURF
Visually significant image point
Weighted color histogram
Spatial distribution entropy
Color complexity measure
Similarity

ABSTRACT

Content-based image retrieval (CBIR) has been an active research topic in the last decade. As one of the promising approaches, salient point based image retrieval has attracted many researchers. However, the related work is usually very time consuming, and some salient points always may not represent the most interesting subset of points for image indexing. Based on fast and performant salient point detector, and the salient point expansion, a novel content-based image retrieval using local visual attention feature is proposed in this paper. Firstly, the salient image points are extracted by using the fast and performant SURF (Speeded-Up Robust Features) detector. Then, the visually significant image points around salient points can be obtained according to the salient point expansion. Finally, the local visual attention feature of visually significant image points, including the weighted color histogram and spatial distribution entropy, are extracted, and the similarity between color images is computed by using the local visual attention feature. Experimental results, including comparisons with the state-of-the-art retrieval systems, demonstrate the effectiveness of our proposal.

1. Introduction

With advances in information technology, there is an explosive growth of image databases, which demands effective and efficient tools that allow users to search through such a large collection. Traditionally, the most straightforward way to implement image database management systems is by means of using the conventional database-management systems such as relational databases or object-oriented databases. The system of these kinds is usually called keyword-based, in which the images are annotated with keywords. However, as the databases grow larger, the traditional keywords based method to retrieve a particular image becomes inefficient and suffers from the following limitations: (1) It is difficult to express visual content like color, texture, shape, and object within the image precisely. (2) For a large dataset, it requires more skilled labor and need very large, sophisticated keyword systems. (3) Further, the keywords increase linguistic barrier to share image data globally. To overcome several of these limitations, many content-based image retrieval (CBIR) systems have been proposed in recent decades, including QBIC, Photobook, MARS, NeTra, PicHunter, Blobworld, VisualSEEK, SIMPLIcity, and others [1]. In a typical CBIR, features related to visual content such as color, and texture are first extracted from a query image, the similarity between the set of features of the query image and that of each target image in a DB is then computed, and target images are next retrieved which are most similar to the query image. CBIR is a very active research topic in recent years, and comprehensive and extensive literature survey on CBIR is presented in [2,3].

Extraction of good visual features which compactly represent a query image is one of the important tasks in CBIR [4]. From early CBIR research, we can see that, the low-level features they applied in representing images are often "*global features*", which are extracted from an entire image [1,3–5]. This way of image retrieval receives more encouragement from many promising CBIR methods [1,2]. However, the performance of these CBIR approaches is far from users' expectation. The problem can be due to the following two reasons. First, it is not unusual that targets, for which users search through an image retrieval system, is not images but the visual objects in images. Global features extracted from the images cannot represent the characteristics of objects in these images. Second, features used in most CBIR work are low-level features (e.g. colors, textures, etc.). The semantic gap between low-level features and highlevel semantic understanding of images is often hard to bridge. This particular problem is also known as the semantic gap problem in image retrieval research. Therefore, in order to solve the problems, which are the bottleneck for further improvement of CBIR performance, "*local features*" based image retrieval approaches begin to receive more attention [2,3,6].

Local features based image retrieval is an image retrieval approach which focuses on contents from salient points (or

* Corresponding authors. Address: School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, PR China (X.-Y. Wang).
E-mail addresses: yhy_65@126.com (H.-Y. Yang), wxy37@126.com (X.-Y. Wang).

http://dx.doi.org/10.1016/j.jvcir.2014.05.003
1047-3203/

regions) of images, not the content from the entire image in early CBIR. For local features based image retrieval, it first segments images into a number of salient points (or regions), and extracts a set of features, which are known as "local features", from salient points (or regions). A similarity (or distance) measure determining the similarity between target salient points (or regions) in the query and a set of salient points (or regions) from other images is utilized later to determine relevant images to the query based on local features. The motivation of local features based image retrieval approaches are based on the fact that high-level semantic understanding of images can be better reflected by local features of images, rather than global features [2,3,6].

Image retrieval tasks based on salient regions are reported in [7–12]. Pratikakis et al. [7] presented a strategy for unsupervised robust CBIR. The basic components of the proposed scheme are (i) a meaningful watershed-driven hierarchical segmentation that partitions the image into visually consistent homogeneous regions and (ii) a feature set that combines color, texture and spatial characteristics that are further weighted by a novel weighting scheme that is inherent to the proposed segmentation method. In scheme [8], the image was segmented into salient regions using the watershed segmentation algorithm. A set of features, comprising the color-size feature and Gabor texture, was then extracted from each watershed region. Finally, by using the extracted visual features of image regions, the similarity measure between the query and each candidate image was computed. Sylvie et al. [9] developed a fuzzy region-based image retrieval system compatible with real-time use and dedicated to object retrieval. Its contributions are a new representation of the image as a set of fuzzy regions with their features, and a matching between sets of regions, compatible with their features, their composition. In scheme [10], rotation invariant curvelet texture features are proposed for region based image retrieval. Several contributions have been made. Firstly, a comprehensive review of state of the art texture methods in literature is conducted. Secondly, curvelet transform and curvelet texture features based on low order statistics are formally introduced for image retrieval and systematically analyzed. Thirdly, a rotation invariant texture feature based on curvelet transform is proposed. Chung et al. [11] proposed a region-based object retrieval using the generalized Hough transform (GHT) and adaptive image segmentation. The proposed approach has two phases. First, a learning phase identifies and stores stable parameters for segmenting each database image. In the retrieval phase, the adaptive image segmentation process is also performed to segment a query image into regions for retrieving visual objects inside database images through the GHT with a modified voting scheme to locate the target visual object under a certain affine transformation. The learned parameters make the segmentation results of query and database images more stable and consistent. In scheme [12], an improved saliency map computing algorithm is employed first. Then, based on the saliency map, an efficient salient edges and regions detection method is introduced. Moreover, the concepts of salient edge histogram descriptors and salient region adjacency graphs are proposed, respectively, for images' similarity comparison. Finally, an integrated strategy is adopted for content-based image retrieval. Obviously, segmentation into multiple regions is a primary step for region based image retrieval, but segmentation is a difficult task without prior knowledge about the classes present in an image, and it will become more difficult when there are a large number of fragmented objects in a scene, or no specific objects. Therefore many researchers seek different solutions. Applying some salient points to solve the problem is a new trend. These points have special properties which can make them stand out in comparison to their neighboring points.

Zheng et al. [13] presented a shape prior algorithm retrieval based on interest point. Firstly, according to the formulaic composition of a medical image, a Harris point detector is improved to extract some interest points in images. Secondly, by combining invariants for each point and an edge type histogram, the feature vector for matching is constructed. Finally, a strategy for matching vectors is implemented to retrieve medical images. Hiremath and Pujari [14] proposed a novel method for image retrieval using color, texture and shape features. Salient points based on color saliency are computed on the images. Texture and color features are extracted from fixed sized windows around these salient points to serve as local descriptors. Gradient vector flow fields are used to extract edge images of objects. Invariant moments are used to describe the shape features. A combination of these local color, texture and global shape features provides a robust set of features for image retrieval. Stöttinger et al. [15] propose a new image retrieval method using salient points. Based on the Harris corner detector, a way to use multi-channel images is explored and different color spaces are evaluated. To determine the characteristic scale of a salient point, a new color scale selection method is presented. Abdullah et al. [16] described a new method to combine visual MPEG-7 descriptors with spatial information, and employed two approaches for dividing up an image for region-based image retrieval and categorization. They compare fixed partitioning and salient points schemes for dividing an image into patches, and low-level MPEG-7 visual descriptors are used to represent the patches with particular patterns. Pedrosa et al. [17] proposed a method to describe shapes based on salience points. The proposed descriptor utilizes a salience detector and an elastic matching algorithm to measure the similarity between two shapes represented by their salience points. The proposed approach gives good results in recognizing shapes of the same class, even if they are represented by a different number of salience points. Chary et al. [18] proposed an image retrieval algorithm using the connecting edge points. Firstly, each color image is converted into gray form, and the image edges are extracted. And then the connecting edge points are calculated, which are adopted by image retrieval. Stottinger [19] proposed color interest points for sparse image representation. To reduce the sensitivity to varying imaging conditions, light-invariant interest points are introduced. Color statistics based on occurrence probability lead to color boosted points, which are obtained through saliency-based feature selection. Furthermore, a principal component analysis-based scale selection method is proposed, which gives a robust scale estimation per interest point. Minakshi et al. [20] presented a new image retrieval scheme using visually significant point features. The clusters of points around significant curvature regions (high, medium, and weak type) are extracted using a fuzzy set theoretic approach. Some invariant color features are computed from these points to evaluate the similarity between images. However, the above point-based image retrieval approaches are usually very time consuming due to the slow salient point detectors. Besides, most salient point detectors were not designed to give a summary as comprehensive as possible of an image, so some salient points always may not represent the most interesting subset of points for image indexing.

Based on the fast and performant scale and rotation invariant salient point detector-SURF (Speeded-Up Robust Features) [21], and the salient point expansion, we propose a novel content-based image retrieval approach using local visual attention feature. In our approach, the weighted color histogram and spatial distribution entropy of visually significant image points are used to describe image content. The weighted color histogram is not only simple and quick to process, but also can reflect the color perception sensitivity of human visual system. The spatial distribution entropy not only gives the dispersive degree of the pixel patches of a color bin in an image, and but also describes the spatial information of a color image.

This paper is organized as follows. In Section 2, the salient image point detection based on SURF algorithm for color image