J. Vis. Commun. Image R. 25 (2014) 1726–1731

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Fusing multi-cues description for partial-duplicate image retrieval

Chenggang Clarence Yan [a], Liang Li [b,*], Zhan Wang [c], Jian Yin [d], Hailong Shi [a], Shuqiang Jiang [a], Qingming Huang [a,b]

[a] Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[b] University of Chinese Academy of Sciences, Beijing, China
[c] State Key Laboratory of Information Security (DCS Center), Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[d] Department of Computer, Shandong University, Weihai, China

ARTICLE INFO

Article history:
Received 17 April 2014
Accepted 8 June 2014
Available online 26 June 2014

Keywords:
Multi-cues
Local Self-Similarity Descriptor
Spatial pyramid
Partial-duplicate image retrieval

ABSTRACT

In traditional image retrieval, images are commonly represented using Bag-of-visual-Words (BoW) built from image local features. However, the lack of spatial and structural information suppresses its performance in applications. In this paper, we introduce a multi-cues description by fusing structural, content and spatial information for partial-duplicate image retrieval. Firstly, we propose a rotation-invariant Local Self-Similarity Descriptor (LSSD), which captures the internal structural layouts in the local textural self-similar regions around interest points. Then, based on the spatial pyramid model, we make use of both LSSD and SIFT to construct an image representation with multi-cues. Finally, we formulate the Semi-Relative Entropy as the distance metric. Comparison experiments with state-of-the-art methods on four popular databases show the efficiency and effectiveness of our approach.

1. Introduction

With the rapid development of internet and multimedia technology, plenty of partial-duplicate images are generated for picture sharing, information delivering, photo ornamenting, etc. Different from the traditional image retrieval, the duplicate regions among partial-duplicate images are only parts of the whole images while there are various kinds of transformations on scale, viewpoint, illumination and resolution. Although this make the retrieval task more complicated and challenging, partial-duplicate image retrieval is demanded by various real applications (e.g. fake image detection, copy protection, and landmark searching), and has attracted increasing research attentions recently.

Most of existing partial-duplicate image retrieval systems are based on Bag-of-visual-Words (BoW) representation of SIFT extracted from images. However traditional BoW representation loses their discriminative power in partial-duplicate image retrieval since partial duplicate images are duplicates on small local regions rather than entire images. To improve the discriminativeness of BoW representation, research efforts have been dedicated at the following aspects: (1) *Local descriptor aggregation* [1–6]. Winder et al. [5] learn the descriptor via taking advantage of the DAISY configuration (a.k.a. the polar Gaussian pooling approach which has origins in geometric blur). Bronstein et al. [1] construct a spatially-sensitive image descriptor in which both the features and their relations are affine-invariant. Perronnin et al. [3] and Philbin et al. [4] propose a simple and yet efficient way of aggregating local descriptors into a vector of limited dimension inspired by the Fisher kernel representation. The purpose of feature integration is to generate a more discriminative description with spatial information. But they are implemented on the region with a fixed size so as to be limited in the partial-duplicate image retrieval. (2) *Geometric verification* [7–11], which is motivated by the observation that there existing strong geometric constrains among the visual words in an object. For example, Jégou et al. [2] propose a weak geometric consistency constraint in terms of angle and scale to filter the matching descriptors. Philbin et al. [8] exploit the shape information in the affine-invariant image regions to improve query performance. Wu et al. [9] bundle local features into groups and introduce a relative coordinates ordering to improve the retrieval precision. However, the scheme is limited because the relative coordinates ordering is not rotation-invariant. Chum et al. [11] extends the query expiation methods on spatial verification and re-ranking. As there exists plenty of background noise in the partial-duplicate images, the geometric verification is normally ineffective in partial-duplicate image retrieval task.

Based on the above analysis, we design a multi-cues description by fusing structural, content and spatial information for partial-duplicate image retrieval. On the structural description, we

* Corresponding author.
E-mail address: liang.li@vipl.ict.ac.cn (L. Li).

http://dx.doi.org/10.1016/j.jvcir.2014.06.005
1047-3203/

propose a novel rotation-invariant Local Self-Similarity Descriptor (LSSD), which captures the internal structural information in the local textural self-similar regions around interest points. The details of LSSD will be presented in Section 2. On the content description, SIFT, as the most popular local descriptor, has the ability of robust and effective content representation for local point. For the better discriminative power, we combine LSSD and SIFT to form a multi-description for the local interest point. On the spatial description, we construct the spatial pyramid description of both LSSD and SIFT. Finally, according to the unique characteristic of partial-duplicate image retrieval, we propose the Semi-Relative Entropy (SRE) as the distance metric between two images based on the spatial pyramid representation of multi-description. Fig. 1 illustrates the framework of our approach.

In summary, there are three contributions of our work:

- A multi-cues description with structural, content and spatial information is designed for the partial-duplicate image retrieval.
- The proposed LSSD presents good descriptive ability for structural layout and is robust for small local distortions.
- The proposed Semi-Relative Entropy is proved to be reasonable and shows promising performance for partial-duplicate image retrieval.

The remainder of the paper is organized as follows. Section 2 introduces the proposed Local Self-Similarity Descriptor in detail. Section 3 details the multi-cues description framework. Section 4 derives the proposed similarity measure formulation. The results of experiments will be discussed in Section 5. Finally, Section 6 concludes the paper.

2. The local textural self-similarity description

In this paper we present a Local Self-Similarity Descriptor (LSSD) which captures internal structural layouts of local textural self-similarities around interest points and is robust to small local distortions. Fig. 2 illustrates the process of extracting LSSD.

2.1. Distinctive interest point

The Local Self-Similarity Descriptor is built upon interest points. Interest points are commonly employed in a number of real-world applications such as object recognition [12,6,13,14] and image retrieval [3,9], because they can be computed efficiently. In addition, they are resistant to partial occlusion and relatively insensitive to changes in viewpoint. We detect image interest points using the DoG (Difference of Gaussian), proposed by Lowe [15], which has been shown to be robust and effective for detecting interest points. From the detected interest points, we can obtain three cues for each interest point *i.e.* (1) scale; (2) location factor; and (3) orientation factor.

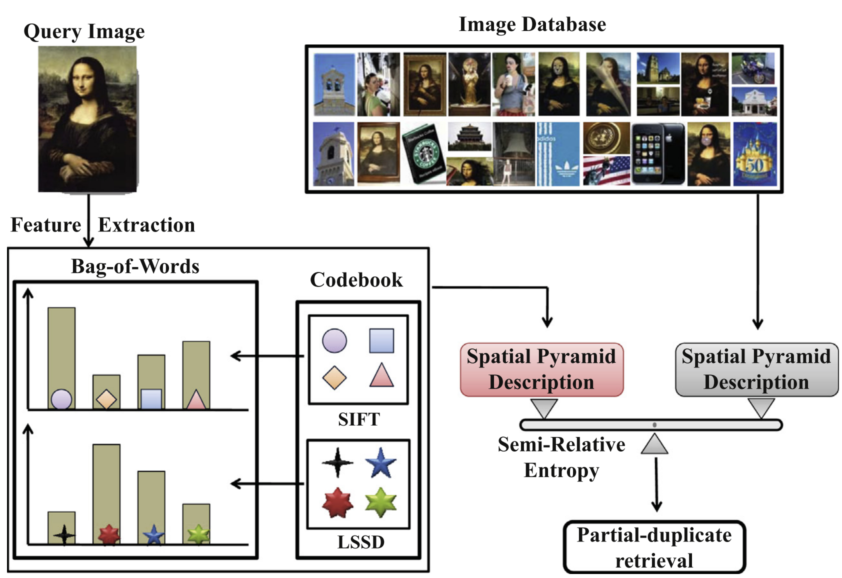

Fig. 1. Flowchart of our proposed approach.

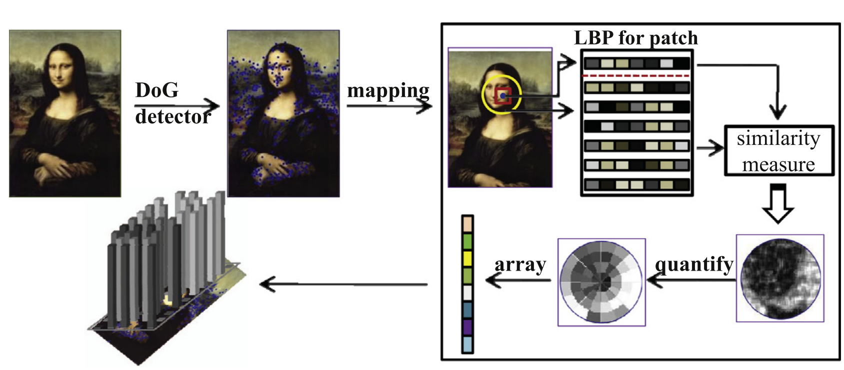

Fig. 2. Extracting Local Self-Similarity Descriptor.