Information Sciences 302 (2015) 94–107

Contents lists available at ScienceDirect

Information Sciences

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

Binary feature from intensity quantization and weakly spatial contextual coding for image search

Dongye Zhuang [a,b,c,*], Dongming Zhang [a,b], Jintao Li [a,b], Qi Tian [d]

[a] Advanced Computing Research Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, China
[b] Beijing Key Laboratory of Mobile Computing and Pervasive Device, China
[c] University of Chinese Academy of Sciences, China
[d] University of Texas at San Antonio, United States

ARTICLE INFO

Article history:
Received 10 May 2013
Received in revised form 7 August 2014
Accepted 29 August 2014
Available online 6 September 2014

Keywords:
Image search
Binary descriptor
Local feature
CBIR

ABSTRACT

During the past few years, a number of local binary feature descriptors for images have been proposed, *e.g.* BRIEF, ORB, BRISK, and FREAK. The binary descriptors have several advantages over the well-established floating vector descriptors such as SIFT and SURF, for their fast computing speed and much low memory consumption. Nevertheless, the binary descriptor still suffer from the poor performance in computer vision applications for its low discriminative power and Hamming distance metric. To improve the capability of binary descriptors, some works focusing on the points selection pattern algorithm in the descriptor extraction are proposed. These works mostly adopt more optimal selection pattern (BRISK and FREAK) to enhance the performance of binary descriptors, rather than random chosen pattern. In our study, however, the points selection algorithm does not have much contributions for the promotion on performance. Therefore, in this paper, we try to solve the problem of low discriminative power and robustness through two novel methods: **Intensity Difference Quantization** and **Weakly Spatial Context Coding**. The experimental results on the public datasets show that our method can significantly boost the performance of binary features and highly enhance the retrieval accuracy of the image search system, even though our proposed method increases slightly memory usage and computing efficiency, compared to the original binary descriptors.

1. Introduction

In this paper, the scheme we focus on is query-by-example image retrieval (*i.e.* given a query image, how to find a subset of the images database and rank the result from the closest to the farthest). During the past ten years, many robust local features have been proposed to make the Content Based Image Retrieval (CBIR) system more efficient and effective. Among these works, SIFT (Scale-Invariant Feature Transform) [21] already has been widely accepted and it is highly discriminant and invariant to a variety of image transformations, but with the cost of expensive computation. In order to improve the computational performance of SIFT, lots of derivative works are proposed, such as SURF (Speeded-up Robust Feature) [2], GLOH (Gradient Location Orientation Histogram) [23], and GA-SIFT [19]. Unfortunately, these approaches that descriptors based on high dimensional floating vector still cannot resolve the conflict between efficiency and effectiveness ultimately.

* Corresponding author. Address: No. 6 Kexueyuan South Road, Haidian District, Beijing 100190, China. Tel.: +86 13810563879.
E-mail address: zhuangdongye@ict.ac.cn (D. Zhuang).

http://dx.doi.org/10.1016/j.ins.2014.08.064
0020-0255/

In addition, these methods pay more attention to handle various photometric or geometric transformations, those are computational expensive and cannot be applied directly to the task of efficient CBIR system. For the sake of these reasons, state-of-the-art methods of image search are mostly based on the Bag-of-Visual-Words model [38].

1.1. Binary features

Recently, a number of binarization schemes have been proposed for their better compact representation and higher computational speed, including Locality Sensitive Hashing (LSH) [7,13], Spectral Hash (SH) [47], Kernel LSH (KLSH) [17], Locality Sensitive Binary Codes (LSBC) [33], Semi-Supervised Hashing (SSH) [43]. These schemes generally map the floating-vector descriptor into Hamming space where the Hamming distance H can be employed. The binary embedding methods address both the computational and memory issues: the Hamming distance between two binary features can be computed extremely efficiently and the memory footprint is drastically reduced.

Recently, many works that using binary strings to directly represent local image patches have been proposed in the top-level international conferences and journals. The notion that binary represented descriptors was introduced by Ojala et al. [26,27], who give birth to the LBP operator that forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Zhou et al. [52] adopted the similar idea of binary descriptor for texture analysis. Afterwards, the binary descriptors have become a important issue in the computer vision community. Gupta et al. [11] proposed the SMD method to construct binary descriptors by the pixel comparison results in image patches. Calonder et al. [3] proposed the method to combine the FAST [34] keypoint detector and the BRIEF (Binary Independent Elementary Feature) algorithm to detect and describe the local feature with very high speed. Rublee et al. [36] promoted the invariance of BRIEF to orientation. Meanwhile, BRISK [18] was proposed to enhance the robustness of BRIEF to orientation and scale. FREAK [1] is another effective binary feature scheme, in which the selection of points is pre-trained by the model that similar to human retina, so as to improve the discriminative power and robustness. Zhou et al. [54] proposed a novel scheme to transform the SIFT [21] descriptor to binary bits through scalar quantization. In their method, the binary features are extracted from the original SIFT [21] descriptor.

In general, the binary descriptors consist of binary bits derived from concatenating the results of simple intensity comparisons. The descriptors can be computed with a very high speed because the descriptor extraction algorithm mainly operates on the integer comparison. Formally, a binary descriptor constructed from a set of binary intensity tests. Considering a smoothed image patch p, a binary test τ is defined by:

$$\tau(p;x,y) := \begin{cases} 0, & p(x) < p(y) \\ 1, & \text{Otherwise} \end{cases} \tag{1}$$

where $p(x)$ is the intensity of the point x. The feature is defined as a vector of n binary tests:

$$f_n(p) := \sum_{1 \leqslant i \leqslant n} 2^{i-1} \tau(p;x,y) \tag{2}$$

Several strategies of pixel pairs sampling are considered in [3], and the experimental results show that selecting the points according to Gaussian distribution around the center of image patch achieves the best comprehensive performance, which means that those points closer to the center of the image patch have more contribution to the binary features. The dimensionality n can be 128, 256, and 512. The Hamming distance is adopted as the metric to measure the similarity between features.

In general, binary features are compactly represented and have very high computational speed. However, the discriminative power and robustness of binary descriptors are lower than the floating-vector based descriptors. In experiments, the accuracy of image retrieval decline sharply because of too much false matches by using Hamming metric in corresponding points matching. In this paper, we search the reason of poor performance of binary descriptors and give a new descriptor extraction algorithm which outperforms the origin algorithm significantly in precision and recall.

1.2. Spatial information and geometric verification

Since the conventional BoVW model discard the spatial geometric information, many researchers exploit the geometric information in BoVW by the approaches from several aspects, *e.g.* introducing a post spatial verification such as RANSAC [9,31,46]; applying the query expansion which reissues the initial retrieval results as queries [8]; combining several point features within an interest region (MSER [22]) to form a bundled features [48]; softly assigning descriptors to multiple words [32]; imposing weak geometry constraints [14]; encoding the spatial relationship among local features in an image [53]; building high-order features [42,51]; and extending to local BoVW models [20]. Most of these methods induce considerable computation, and the spatial verification or query expansion is also costly. Therefore, these works mentioned above are not suitable solutions for binary features. In this paper, we propose the spatial context coding algorithm as described in Section 4.

In addition, another scheme is to learn effective classifiers and ranking functions towards exploring the semantics of visual information [16]. Zha et al. [49,50] proposed a new query suggestion technique named Visual Query Suggestion (VQS), which simultaneously provides both keywords and image suggestions and thus is able to help users specify and deliver their search