Signal Processing: *Image Communication* 32 (2015) 54–68

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Finding more relevance: Propagating similarity on Markov random field for object retrieval

Peng Lu [a], Xujun Peng [b], Xinshan Zhu [c,d,*], Ruifan Li [a]

[a] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China
[b] Raytheon BBN Technologies, Cambridge, MA, USA
[c] School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China
[d] School of Information Engineering, Yangzhou University, Yangzhou 225127, China

ARTICLE INFO

Article history:
Received 29 June 2014
Received in revised form
19 January 2015
Accepted 20 January 2015
Available online 29 January 2015

Keywords:
Image retrieval
Bag-of-visual-words
Markov random field

ABSTRACT

To retrieve objects from large corpus with high accuracy is a challenging task. In this paper, we propose a Markov random field (MRF) based probabilistic retrieval framework. In this framework, the similarities between the query image and dataset images are modeled as the likelihood and the relationships among the images in the dataset are modeled as the prior. Then, the prior and the likelihood are combined to improve retrieval performance. Further, we present an approximate belief propagation algorithm as well as a subgraph extraction algorithm for efficient inference in MRF. Finally, we design a new image retrieval system under our framework. This system can be considered as an extended bag-of-visual-words retrieval system with the probabilistic based re-ranking module. We evaluate our method on three standard datasets: Oxford-5K, Oxford-105K and Paris-6K. The experimental results show that the proposed system significantly improves the retrieval accuracy on these datasets and exceeds the state-of-the-art results.

1. Introduction

In this paper, we are interested in the problem of object retrieval, where the goal is to retrieve all images containing a particular query object specified by a query image, in a large scale and unordered image dataset. This is an important problem with applications in object or location recognition, video search, near duplicate detection and 3D reconstruction. However, it is a challenging task to design an object retrieval system with high accuracy meeting the real-time requirement.

The pioneering work of image retrieval system is based on bag-of-visual-words (BoW) model [1], which is inspired from text-retrieval systems. In this model, an image is represented using bag-of-visual-words, and images are ranked using term frequency inverse document frequency (tf-idf) computed efficiently via an inverted index [2]. Based on this model, many efforts have been made to improve the performance of image retrieval system, which can be grouped into three main categories: (I) optimizing the standard BoW model; (II) enhancing query image information; and (III) exploiting the relationships among images in datasets.

Category I: The basic BoW model has several drawbacks, including lower discriminative feature representation, inappropriate metrics for descriptor comparison and information loss due to descriptor quantization, which cause the retrieval failure.

To address the feature representation problem, Tian et al. [3] proposed an edge oriented difference histogram (EODH) descriptor, which was combined with color-SFIT descriptors in their image retrieval system. In [4], Perd'och and Chum suggested an improved feature detector where discretized

* Corresponding author.

http://dx.doi.org/10.1016/j.image.2015.01.007
0923-5965/

local geometry representation was learnt. By exploring different configurations of normalization, dimension reduction and dynamic range reduction, Winder et al. [5] proposed a better visual descriptor than SIFT for image retrieval with high discriminative capability. In [6], Li et al. proposed an image retrieval framework by concatenating image global features-based similarities with multiple learnt ranking features. Different from the approaches which were relied on global image features, Tao et al. suggested a local point-indexed representation for all images, where the exponential similarity measure was used to calculate the relevance of two local descriptors [7]. This local representation can facilitate the retrieval tasks with very large vocabularies. In [8], Zhou et al. adopted original SIFT features by introducing the spatial relationships among local features in an image. The similar work was also presented in [9], where spatial context information was embedded into the inverted file to speedup retrieval process. Most recently, with the successes of deep learning techniques in computer vision and machine learning areas, they also provided many encouraging results for the content-based image retrieval, especially for effective feature representations [10].

In order to avoid the false rejection problem introduced from the vocabulary quantization in traditional BoW based methods, Philbin et al. described a soft-assignment method in [11], which can include features lost in the quantization stage. And they suggested an approximate nearest neighbor (ANN) based k-means approach in [12] to build large visual vocabulary, which boosted the retrieval accuracy considerably. By closely investigating the relation between nearest neighbor search and clustering in high-dimensional spaces, Avrithis [13] proposed a dimensionality-recursive quantization method that can be used for image retrieval efficiently. To reduce quantization error, some researchers generated multiple vocabularies to quantize feature vectors into visual words. In [14], Zheng et al. proposed a Bayesian based merging scheme, which reduced the correlation between vocabularies by calculating the cardinality ratio of intersection and union sets from indexed features list.

Instead of focusing on improvements on feature representation and vocabulary quantization, other researchers explored methods on the similarity measure between images. In [15], Arandjelović proposed a square root (Hellinger) kernel based approach to measure the distance between SIFT descriptors, which outperformed the standard Euclidean distance measure for image retrieval. Similarly, Jégou et al. [16] presented a Hamming embedding scheme to represent descriptors more precisely where weak geometric consistency constraints were applied. In [17], Mao et al. proposed a general distance metric for the local features, which considers L1, L2 and hamming distance measure as special cases under this framework. Zheng and Wang [18] suggested the visual phraselet as a kind of similarity measure between images which is more discriminative than visual word and phrase. In their other work, Zheng et al. proposed a L_P norm IDF which takes into account the term frequency, document frequency, the complexity of images, as well as the codebook information, instead of conventional IDF [19]. In [20], Qin et al. presented a probabilistic framework to measure the similarity between features where a new cost function was also proposed in this paper to score the similarity between images. Inspired by this work, Yang et al. also proposed an adaptive similarity function to replace Euclidean distance or Gaussian kernel which was used by nearest neighbor (NN) search for image retrieval [21]. Based on this adaptive similarity and the normalized Gaussian kernel, a hashing scheme was constructed to improve the efficiency and accuracy of ANN used for image retrieval.

Category II: Different from previously described basic image retrieval methods that concentrate on querying using a single image, multiple-queries based approaches tackle the problem of finding all images of an object by using multiple images. In [22], Fernando and Tuytelaars focused on an unsupervised approach to discover visual patterns for the specific object, that are robust to noisy query images. By combining multiple queries in a principled manner, Arandjelović and Zisserman [23] described a framework where the query images are obtained by using textual Google image search. In order to capture the mutual information of multiple queries, Chen et al. [24] employed discriminative ranking criterion to rank images in dataset, where a group of correlated instances of a query object were provided.

Unlike multiple-queries based approaches, where many instance images of the query object are explicitly required by users, query expansion based re-rank methods retrieve a number of highly ranked images from the original single query and reissue a new query by adding more relevant terms into the query. In order to improve retrieval performance, Chum et al. [25,26] adopted a query expansion method to enrich the query model by using spatially verified features. In [15], Arandjelović and Zisserman proposed a discriminative query expansion approach(DQE), where an SVM classifier was trained by initial retrieved images and used to re-rank final images in dataset. In [27], Fernando and Tuytelaars designed a set of mid-level patterns based on the images retrieved by the Google image search engine. These patterns were used for improving the retrieval performance.

Category III: Most recent studies used the matching relationships among dataset images to boost the retrieval performance. In order to enhance the images of datasets according to their neighbors, Turcot and Lowe [28] proposed a matching graph [29] based method, which was mainly used for object mining or location recognition [30] in large scale image datasets.

Based on the initially retrieved images, a re-ranking method was proposed by Shen et al. [31] where a voting-based approach was used to evaluate the similarity measure between K-nearest neighbors (K-NN) of the query. Using the idea similar in spirit to [31], in [32], a k-reciprocal nearest neighbor structure was created in the image space to identify an initial set of relevant images in dataset which were then used to re-rank the remaining images. Based on the observation that initial retrieved images which yielded similar ranking lists are relevant to each other, when they were used as queries, Chen et al. [33] applied a ranking consistency verification procedure to refine the existing ranking list for retrieval. In [34], a modified K-NN graphs was constructed to retrieve images with high accuracy and efficiency. Inspired by the method of graph propagation, Niu et al. [35] presented an active learning scheme for selective sampling in relevance feedback by considering neighboring images' relationship in