Information Sciences 417 (2017) 143–153

Contents lists available at ScienceDirect

Information Sciences

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

Local residual similarity for image re-ranking

Shaoyan Sun[a], Ying Li[b], Wengang Zhou[a], Qi Tian[c], Houqiang Li[a,*]

[a] EEIS Department, USTC, Hefei 230027, PR China
[b] ICE School, DUT, Dalian 116023, PR China
[c] One UTSA Circle, San Antonio, TX 78249, U.S.

ARTICLE INFO

Article history:
Received 12 September 2016
Revised 14 June 2017
Accepted 4 July 2017
Available online 5 July 2017

2010 MSC:
00-01
99-00

Keywords:
Image retrieval
Re-ranking
Similarity measure

ABSTRACT

Similarity measurement is an essential component in image retrieval systems. While previous work is focused on generic distance estimation, this paper investigates the problem of similarity estimation within a local neighborhood defined in the original feature space. Specifically, our method is characterized in two aspects, *i.e.*, “local” and “residual”. First of all, we focus on a subset of the top-ranked relevant images to a query, with which anchors are discovered by methods such as averaging or clustering. The anchors are then subtracted from the neighborhood features, resulting in residual representations. The proposed Local Residual Similarity (LRS) homogenizes the feature distances within the local neighborhood. Effective and efficient image re-ranking is achieved by calculating LRS between the query and the top-ranked images. The method constrains that relevant images should appear similar in both original and local residual feature space. We evaluate the proposed method on two image retrieval benchmarks with global CNN representations, demonstrating a consistent improvement on performance with very limited extra computational cost.

1. Introduction

In this paper we concentrate on the task of image retrieval [1–6]. Given a query image, we aim to find its similar counterparts from an image database. Typically, for each image a feature vector is extracted, with which the similarities between the query and database images are measured. Then the database images with high similarity scores to the query can be returned as retrieval results.

The image retrieval community has witnessed a number of effective image representations. For example, the Bag-of-Visual-Words representation [1,2,7] follows the practice in text retrieval to group the frequencies of visual word terms into a histogram. This model has dominated the image retrieval frameworks over a decade. To aggregate local visual features into a compact representation, VLAD [8] is proposed to train small visual codebooks and sum up residual vectors of local features with respect to the codewords. Most recently, the convolutional neural network (CNN) is successfully applied in image retrieval for feature extraction and produces superior accuracy [9–11].

As effective feature representations are crucial to improve image retrieval performance, we thereby study on more effective usage of the existing feature representations by exploiting residual vectors with respect to some local anchor-points in the neighborhood. The residual representation has been widely applied in the field of machine. As mentioned above, Jégou

* Corresponding author.
E-mail address: lihq@ustc.edu.cn (H. Li).

http://dx.doi.org/10.1016/j.ins.2017.07.004
0020-0255/

et al. [8] propose VLAD to aggregate residual vectors of local features. Arandjelović et al. [12], Liu et al. [13], *etc.*, proceed to further improve this representation. Product Quantization [14] is proposed for encoding quantized residual vectors, and Residual Net [15] exploits residual learning to overcome deeper neural network training problem and achieves state-of-the-art classification accuracy. Bai et al. [16] propose a hierarchical residual framework for feature coding and approximate nearest neighbor search. Bai et al. [17] address 3D shape matching by introducing the residual coding algorithm VLAD into a two-layer coding framework.

Motivated by the latent benefit of residual representations, we propose a novel distance measure for image retrieval re-ranking. We first make an initial query with original feature vectors to identify the neighborhood of the query image, and then represent images in the neighborhood as residual vectors with respect to local anchor-points for an extra re-ranking process. While the residual vectors have equivalent Euclidean distance with respect to the original feature vectors, we use cosine distance to measure included angle value of the residual vectors. We demonstrate the feasibility of designing a more effective distance measure with this strategy by exploiting local feature distribution information.

Our work contributes in three aspects:

(1) We propose a novel similarity measure, Local Residual Similarity (LRS), in a neighborhood of the query defined by its top-ranked images. With this, an efficient online re-ranking method is introduced.
(2) We perform extensive evaluations on various strategies for choosing a neighborhood and computing local anchor-points.
(3) We elaborate two strategies to effectively improve the re-ranking performance by imposing reciprocal neighborhood constraints on our method.

2. Related work

CNN based image retrieval has achieved considerable processes in recent years. In [10], Babenko et al. propose to extract features from CNN activations as *Neural Codes*, and demonstrate its practicality in the image retrieval task. Gong et al. [11] extract CNN features from multiple scales and positions in the image, and aggregate the features by pooling. In [18], the authors evaluate multiple aggregating strategies to improve the CNN representation. Apart from these methods that exploit off-the-shelf CNN features, some works train CNN architectures for some specific image retrieval tasks, *e.g.* landmark recognition [19–21].

With well designed feature representations, re-ranking becomes an effective tool to improve the image retrieval results. To refine results in the ranking list, several supervised re-ranking methods are proposed to learn concepts from manually labeled data [22–27]. Wang et al. [23] learn different semantic spaces (*i.e.*, concepts) and generate semantic signatures by projecting correlate visual features into the same semantic space. Some other works employ multiple attributes as constraints for refining [24], and use feedbacks from user interaction in [25,28]. In order to avoid noisy labels introduced by various causes (*e.g.*, undetermined user intentions), Jain et al. [29] alternatively take the statistical information of click data as a reference to facilitate re-ranking.

As a form of automatic relevance feedback without supervision, query expansion [30] is effective to improve recall, which is transferred from the documental domain to the visual field by [31]. Geometrical consistence restriction is applied in [31] to filter out unrelated images in the candidate list, and the rest is used to construct new queries. Chum et al. [32] improve context expansion for each retrieved image, and [5,33] learn weighting factors through SVM. Some other works apply graph diffusion based reranking methods [34–38]. Zhang et al. [35] exploit the reciprocal neighborhood relationship to build a graph for rank list fusion. Pedronette et al. [36] construct a correlation graph with rank information, and apply unsupervised manifold learning to improve image retrieval performance. Further more, some binary signature based methods [39,40] are proposed in concern of efficiency.

Since supervision information is hardly obtained in practical image retrieval systems, the reciprocal neighborhood relationship is widely applied for unsupervised ground truth verification. Jégou et al. [4,41] propose contextual dissimilarity measure (CDM) based on the average distance of features to their neighborhood to maximize the neighborhood symmetry by iteration. Bai et al. [42] take into account the neighborhood of a given visual word vector when the dissimilarity between visual word vectors are measured. Instead of pursuing the symmetry relationships between the nearest neighbors, Qin et al. [43] design separate similarity metrics for highly relevant neighbors and the rest far away points, respectively. Our method harnesses the reciprocal neighborhood relationships for distance measure designing, and also for extensions by integrating some relevant constraints. Different from CDM, which explicitly assigns a score representing the density around each feature, our method implements a query adaptive online distance measure, which requires no extra storage and runs very fast. In addition, we demonstrate that by extending our method with strategies such as CDM we can further improve the performance.

3. Local residual representation for re-ranking

We elaborate our method in this section. First, we introduce the framework of our local residual representation based re-ranking algorithm in Section 3.1. Then we discuss the definition of neighborhood and anchor-points in our method in Section 3.2. In the last, we make some extensions by imposing reciprocal neighbor constraints in Section 3.3.