Pattern Recognition 47 (2014) 1349–1360

Contents lists available at ScienceDirect

Pattern Recognition

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

Ranking consistency for image matching and object retrieval

Yanzhi Chen [1], Xi Li *, Anthony Dick, Rhys Hill

ACVT, School of Computer Science, The University of Adelaide, SA 5005, Australia

ARTICLE INFO

Article history:
Received 24 January 2013
Received in revised form
23 July 2013
Accepted 15 September 2013
Available online 27 September 2013

Keywords:
Ranking consistency
Image matching
Object retrieval
List-wise min-Hash

ABSTRACT

The goal of object retrieval is to rank a set of images by the similarity of their contents to those of a query image. However, it is difficult to measure image content similarity due to visual changes caused by varying viewpoint and environment. In this paper, we propose a simple, efficient method to more effectively measure content similarity from image measurements. Our method is based on the ranking information available from existing retrieval systems. We observe that images within the set which, when used as queries, yield similar ranking lists are likely to be relevant to each other and vice versa. In our method, ranking consistency is used as a verification method to efficiently refine an existing ranking list, in much the same fashion that spatial verification is employed. The efficiency of our method is achieved by a list-wise min-Hash scheme, which allows rapid calculation of an approximate similarity ranking. Experimental results demonstrate the effectiveness of the proposed framework and its applications.

1. Introduction

In recent years, the availability of large, unordered image collections on the Internet has given rise to a need for robust, scalable image retrieval systems. Given a query image, the goal of image retrieval is to efficiently find other images that depict the same object and to rank them in order of similarity. Typically, a large-scale retrieval system focuses on three important issues: (i) image representation, (ii) image similarity measure and (iii) retrieval result refinement as a post-process.

To address issue (i), many retrieval systems use a simple bag of words (BoW) representation, which represent each image as a histogram of *visual words* extracted from the image [1]. Although the BoW representation is simple, it lacks robustness to variations in scale, illumination or viewpoint that occurs in images. There has been extensive study on improving the robustness of the BoW representation, *e.g.* soft visual word representation improves the feature matching between images [2,3]; distance metric learning reduces the quantisation errors [4]. However, these methods either lack efficiency in memory usage and run time [2,3], or require a time-consuming learning stage [4]. To address issue (ii), many retrieval systems apply fast computation of similarity measure between query/response image pairs, *e.g.* the dot-product similarity of the weighted BoW representations [1,5], or approximate similarity measures [6]. Although these methods are efficient, they fail in varied image conditions [1,5] and some are only effective for near-duplicate image retrieval [6]. To address issue (iii), refinement methods usually re-rank the initial retrieval results by exploiting extra information about the images, *e.g.* raw feature correspondences [5]; relevance feedback to refine the original query [7,8]; and the ranking preference provided by the users to train a classifier [9,10]. These methods struggle to balance efficiency and effectiveness: the re-ranking method [5] based on feature correspondence requires RANSAC processing, leading to a relatively high computational cost and a restriction to rigid object matching; both the relevance feedback and ranking preference methods need an additional stage after the query (*e.g.* a re-query process in [7,8] and a learning stage in [9,10]).

In this paper, we focus on issue (iii). We aim to design an effective and efficient re-ranking method for result refinement, which is based on an improved image similarity measure. In contrast to previous result refinement methods, our method utilizes consistency information collected after initial retrieval results, *without prior knowledge of the images*. It is observed that consistency in ranked results indicates that query images are likely to contain similar content. Therefore, we propose an image retrieval system that exploits the ranking consistency information among images. Based on this ranking information, we propose to refine the image retrieval results, while retaining efficiency and not relying on low level information, *e.g.* spatial or geometric feature information.

Firstly, we propose a simple yet effective image similarity criterion, named *ranking consistency*, in which the similarity between two images is measured by the similarity of the ranked lists that result from using them as queries. The usage of ranking consistency in the image domain is motivated by the ranking

* Corresponding author. Tel.: +61 8 8313 1855; fax: +61 8 8313 4366.
E-mail address: xi.li03@adelaide.edu.au (X. Li).

[1] Current address: United Technologies Research Centre (China) Ltd.

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.09.011

Fig. 1. Ranking consistency overview. The examples are top ranked results of *all souls 1*, where the input image A and B are relevant, but both of them are irrelevant to C. Our method generates some top ranked results for each image by list-wise min-Hash. The similarity between images is measured by the similarity between their top ranked results.

result comparison used in information retrieval. Fig. 1 illustrates our key idea with some top retrieved results by similar query images A and B, and dissimilar images A and C. Note that image C is a highly ranked false positive result of query A. The retrieved results using image C as a query are completely different from the results of images A and B, supporting the fact that images with similar contents are consistent in their top search results. The observation motivates us to use ranking consistency as a verification method: *images whose content matches a query can be inferred on the basis of their ranking consistency*. The ranking consistency criterion can work with any retrieval system, as it only requires ranks of images. In addition, our ranking consistency criterion does not need the comparison of any geometric information, unlike [5].

Secondly, we propose an efficient image re-ranking method, *ranking verification*, to re-rank an initial set of retrieved results by the embedded ranking consistency information. The ranking verification requires online computation of a fixed number of queries based on top ranked images (typically $K=200$ in our implementation). Therefore, the re-ranking process is either inefficient using the standard dot-product vector comparison [1,5] or less accurate due to the approximate similarity comparison [6]. Instead, we use a novel variant of min-Hash to obtain rapid approximate ranking lists, whose consistency is then measured to evaluate the final image result ranking. The effectiveness of our method is due to two factors. First, our final ranking is the result of multiple min-Hash queries, so errors in individual queries can be tolerated. Second, we take into account multiple words from each image, which increases the average recall of the approximate method.

Finally, the ranking verification requires little extra cost per image. It only needs to store hash keys for each image instead of geometric information of each feature, as is required by spatial verification. Moreover, the ranking consistency similarity can be intrinsically used in many object retrieval related problems, *e.g.* expansion of the query model, as an alternative to spatial verification. We also illustrate a graph structure of dataset images built on ranking consistency, which is useful for object mining in large image sets.

2. Related work

Recently, there has been a great deal of research to increase the accuracy of object retrieval by measuring the spatial consistency between the query and result images [5,11,12]. For example, spatial verification [5] is a widely used method to filter out false positives from the top-K results. It relies on matching a minimum number of inlier features between images, and then estimates a rigid transformation via RANSAC [13]. Spatial adjacency is used in [12] to accurately localise objects, by searching for consistent min-Hashes within a restricted image region. The concept of visual phrases proposed in [11] is defined by the spatial offset of the features located in the image space. However, as mentioned above, these methods rely on the geometric information between pairwise images, and thus are less effective when the dataset images lack geometric information.

In contrast, information retrieval has used ranking information to enhance and evaluate retrieval systems for many years. These methods consider ranking information in the following ways: (i) learning to rank with relevance feedback; (ii) re-ranking with results consistency.

Relevance feedback: It aims to refine the ranking model by some labeled data indicating relevance or irrelevance to the query. This is known as relevance feedback in information retrieval [14]. Relevance feedback is also used in some image retrieval systems to refine the ranking functions. These methods can be categorised into two groups. The first group of methods focus on formulation of a new query to take into account the relevant features to the original query. For example, query expansion [7] adds more relevant features from an automatic scheme of sampling selection; Bayesian relevance feedback [15] needs users to identify retrieved images as being relevant or not, and then adjusts the query by Bayesian decision theory; Trademark retrieval [16] dynamically improves both the query formation and similarity measure with relevance feedback. The second group of methods usually use pairwise ranking method, *e.g.* Ranking SVM [17], to sort the document (or image) relationship to the given query [9,10]. In order to train a ranking classifier, these methods need to know either some ranking preference in advance [9], or user-interaction information [10].

Results consistency: It uses the relevance of ranked results to improve retrieval performance as a post-processing step. Ranked results are often compared, using a ranking similarity measure, *e.g.* Spearman's ρ [18] or Kendall's τ [19]. The similarity measurement scores documents sharing many common results highly, which indicates the ranking consistency in these documents. The consistency of ranking has been considered in a number of image retrieval methods, such as [20–22]. In these works, the initial retrieved results are processed with some high level information, *i.e.* a relevance model to evaluate the linked text search results for similarity measurement [20]; a distance matrix defined by the similarity of ranking lists to take into account contextual information [21,22]. However, these methods require expensive post-processing.

Similar to previous ranking refinement methods, our ranking verification method benefits from an improved image similarity measure. Typically, similarity measures can be categorized into two groups [23]: individual image similarity measure and semantic concept measure. Image similarity measures are usually based