ARTICLE IN PRESS

Neurocomputing xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Neurocomputing

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

Enhancing instance search with weak geometric correlation consistency

Zhenxing Zhang*, Rami Albatal, Cathal Gurrin, Alan F. Smeaton

Insight Centre for Data Analytics, School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland

ARTICLE INFO

Keywords:
Multimedia indexing
Information retrieval
Instance search
Weak geometric correlation consistency

ABSTRACT

Finding object instances from within in large image collections is a challenging problem with many practical applications. Recent methods inspired by text retrieval have achieved good results, however a re-ranking stage based on spatial verification may still be required to boost performance. To improve the effectiveness of such instance retrieval systems while avoiding the computational complexity of a re-ranking stage, we explore the geometric correlations among local features, and we incorporate these correlations with each individual match to form a transformation consistency in rotation and scale space.

This weak geometric correlation consistency can be used to effectively eliminate inconsistent feature matches in instance retrieval and can be applied to all candidate images at a low computational cost. Experimental results on three standard evaluation benchmarks show that the proposed approach results in a substantial performance improvement when compared with other state-of-the-art methods. In addition, the evaluation results from participating in the Instance Search Task in the TRECVid evaluation campaign also suggest that our proposed approach enhances retrieval performance for large scale video collections.

1. Introduction

Given a query image of an object, the objective of this work is to find images that contain recognisable instances of the object from a large image collection, henceforth referred to as “instance search”. A successful application of instance search requires efficient retrieval of instance images with high accuracy, possibly under varying imaging conditions such as rotation, viewpoint, zoom level, occlusion, and so on.

Instance search is an interesting, yet challenging, problem and has attracted significant research attention in recent years. Most of the state-of-the-art approaches [11,16,19] have been developed based on the Bag-of-Visual-Words (BoVW) representation first introduced by Sivic et al. [3]. This representation framework successfully made use of the discriminative power of local feature descriptors, for example SIFT [1] and SURF [2] which are generally robust to multiple changes in imaging conditions, and are applied to build a statistical representations for each image in the database. At query time, the BoVW representation may take advantage of indexing techniques such as inverted files [4] to provide fast retrieval over large collections.

However this representation leads to a loss of the ability to encode spatial information between local features, so spatial verification [16] was subsequently introduced to improve retrieval accuracy. Based on the observation that the layout of local features from the query object and its instances should share the same or similar geometric structure, there could be only one feature correspondence for any given feature in the query object. So the geometric layout of query objects was adopted to verify the spatial consistency between initial matched local features. Generally, such spatial verification algorithms [11,16] were applied to train models to capture the transformation in spatial space (normally 3-D transformation including position, rotation, and scale changes) between feature correspondences and to fit them to the initial correspondences to eliminate inconsistent matches in order to refine the results. However these techniques such as RANSAC [25] are normally computationally expensive, and so in practice they are only be applied as a post-processing step to a limited set of top-ranked images in the initial results.

In our work, we address the challenges of improving the efficiency and robustness of examining the consistency between local feature matches to enhance the retrieval performance of instance search systems. However, in contrast to previous spatial verification technologies, we followed the work proposed by Jégou et al. [5] to efficiently apply spatial verification. Instead of estimating full spatial transformation models, which is what Jégou et al. [5] propose, we focused on building weak geometric constraints in 2-D space, specifically in the rotation and scale parameters, to examine each individual feature correspondences. This leads to a much reduced computational expense and although having only two spatial parameters are not sufficient to map objects from one image to another, the weak geometric constraints could help us to filter out inconsistent feature correspondences at a

* Corresponding author.
E-mail addresses: zzhang@computing.dcu.ie (Z. Zhang), ramibatal@gmail.com (R. Albatal), cgurrin@computing.dcu.ie (C. Gurrin), alan.smeaton@dcu.ie (A. F. Smeaton).

http://dx.doi.org/10.1016/j.neucom.2016.09.104
Received 30 March 2016; Received in revised form 17 July 2016; Accepted 21 September 2016
Available online xxxx
0925-2312/

very low computational cost. The real consequence of this is that reduced cost makes this suitable for very large data collections and this is an enhancement on the original work by Jégou et al. [5]. In the work from Jégou et al. [5], we observe that their approach considered feature matches independently and ignored the geometric correlation between local features, and thus it performed less effectively when performing search in more challenging datasets like FlickrLogos-27 [18]. In our work, we believe that the geometric correlation between reliable feature matches should also be consistent with the weak geometric constraints, just like each individual feature match. Based on that, we propose a scheme to incorporate the geometric correlations between matched feature correspondences to form a weak geometric correlation consistency to improve the effectiveness of spatial verification. Thus the main contribution of the paper is the construction of weak geometric constraints in 2-D space allowing this to be done with low computational cost while still supporting geometric correlations but in a much more efficient way when scaled to large datasets and we demonstrate this on some large data collection

This paper is organised as follows. In the next section we present some of the most relevant work in the area and following that, in Section 3 we introduce the idea of weak geometric correlation consistency. In Section 4 we present a brief motivation for the need for improvement in instance search matching, and our experiments are introduced and presented in Section 5. Our experimental results on the TRECVid INS task are presented in Section 5. In summary, the paper illustrates that our proposed method is more reliable, and also more tractable for large image collections, and leads to an overall significant improvement of instance search performance compared to state-of-the-art methods.

2. Related work

In this section, we briefly review the development of visual instance retrieval systems and discuss existing approaches to improving retrieval performance by using the geometric information.

The idea of re-ranking a visual search and carrying out some form of dimensionality reduction in the process is now receiving some attention in the research community. One approach [29] is based on exploiting the overall manifold structure of the whole of the dataset and preserving relationships between example images which have been labelled, effectively mapping a data set to a lower dimensionality under constraints of preservation.

Sivic and Zisserman [3] were the first to address instance search using a BoVW representation combined with scalable textual retrieval techniques. Subsequently, a number of techniques have been proposed to improve performance. The work reported in [11] suggested using a very high dimensional vocabulary (1 million visual words) during the quantisation process. This method improved the retrieval precision with more discriminative visual words, and also increased retrieval efficiency with more sparse image representations, especially for large scale database. Chum et al. [13] brought query expansion techniques to the visual search domain and improved instance recall by expanding the query information. For further improvement on the retrieval performance, both approaches added the spatial verification stage to re-rank their results in order to remove noisy or ambiguous visual words.

While geometric-based constraints are of ultra-importance for some image-matching applications like fingerprint matching, as described in [28], recent work reported in [7–10, 14, 17] extended the BoVW approach by encoding the geometric information around the local features into the representation and refine the matching based on visual words. Those methods were very sensitive to the change in imaging condition and made them only suitable for partial-duplicate image search.

For many high-dimensional data processing applications which involve image data like search, clustering, duplicate detection, etc., we often have too many features in the image representation, many of which are often redundant and noisy. Feature selection is one technique for dimensionality reduction that involves identifying a subset of the most useful features. In [30], a novel unsupervised feature selection algorithm, named clustering-guided sparse structural learning (CGSSL), is proposed that, like our work, helps to discriminate among features, in our case using geometric constraints.

One of the other ways to reduce the feature space is to learn a subspace, at retrieval time, and to perform the matching within that subspace. In [6], the authors proposed exactly that, and to guarantee their subspace to be compact and discriminative, the intrinsic geometric structure of data, and the local and global structural consistencies over labels were exploited simultaneously in a similar way to the way we exploit geometric constraints.

Even more recently, alternative approaches have been developed to implicitly verify the feature matches with respect to the consistency of their geometric relations, i.e., scaling, orientation, and location, in the Hough transformation space. Avrithis and Tolias [12] developed a linear algorithm to effectively compute pairwise affinities of correspondences in a 4-dimensional transformation space by applying a pyramid matching model constructed from each single feature correspondence. Jégou et al. [5] increased the reliability of feature matches against imagining condition changes by applying weak constraints to verify the scaling and orientation relations consistency according to the dominant transformation found in the transformation space. Similarly, [15] proposed to represent the feature points' geometric information using topology-base graphs and verified the spatial consistency by performing a graph matching.

Finally, if we broaden out the function from just instance search to other content-based operations we find that there are other examples of work which uses some form of spatial coding, to encode the spatial relationships among local features in an image. Work described in [10] shows significant improvement in performance, but accuracy and computational cost, for the task of detecting near-duplicate images in web search.

Our proposed method follows the direction of implicitly verifying feature matches do exist in an instance search in order to reduce the computational cost. However compared to existing work, which focused on individual correspondences, our proposed method also considers the spatial consistency for the geometric correlations between matched feature correspondences, while maintaining the efficiency and increasing the effectiveness of the instance search systems.

3. Weak geometric correlation consistency

In the BoVW representation for images, local features are first extracted from each image to encode invariant visual information into feature vectors. Generally, a feature vector is defined as $\vec{v}(x, y, \theta, \sigma, q)$, where variables $\{x, y, \theta, \sigma\}$ stand for the local salient point's 2-D spatial location, dominant orientation, and most stable scale, respectively while q represents a 128-D feature vector to describe the local region. For a query image I_q and candidate image I_c, a set of initial matching features $C_{initial}$ could be established by examining the feature vector q. The task of spatial verification is to eliminate the unreliable feature matches and only retain the matches set C_{stable} that link the patches of the same object. The following equation formulates this process:

$$C_{stable} = \{m_i \in C_{initial} \text{ and } f_{sp}(m_i) = 1\} \tag{1}$$

where m_i stands for the ith feature match in the initial match set and f_{sp} stands for the spatial verification function for assessing a geometric consistency. Take the weak geometric consistency of [5] for example, the verification function in their work could be expressed as follows:

$$f_{sp} = \begin{cases} 1 & \text{if} \Delta\theta \in D_\theta \text{ and } \Delta\sigma \in D_\sigma \\ 0 & \text{if otherwise} \end{cases} \tag{2}$$