Pairwise Geometric Matching for Large-scale Object Retrieval

Xinchao Li, Martha Larson, Alan Hanjalic
Multimedia Computing Group, Delft University of Technology
Delft, The Netherlands
{x.li-3,m.a.larson,a.hanjalic}@tudelft.nl

Abstract

Spatial verification is a key step in boosting the performance of object-based image retrieval. It serves to eliminate unreliable correspondences between salient points in the given two images and is typically performed by analyzing the consistency of spatial transformations between the image regions involved in individual correspondences. In this paper, we consider the pairwise geometric relations between correspondences and propose a strategy to incorporate these relations with significantly reduced complexity, which makes it suitable for large-scale object retrieval. In addition, we also combine the information on geometric relations from both the individual correspondences and pairs of correspondences to further improve the verification accuracy. Experimental results on three reference datasets show that the proposed approach results in a substantial performance improvement compared to the existing methods, without making concessions regarding computational complexity.

1. Introduction

In this paper we address the challenge of improving the efficiency and reliability of image matching in an object-based image retrieval scenario. Under object-based image retrieval, further referred simply to as "object retrieval", we understand the problem of finding images that contain the same object(s) or scene elements as in the query image, however, possibly captured under different conditions in terms of rotation, viewpoint, zoom level, occlusion or blur. Many object retrieval approaches and methods [1, 2, 4, 7] have been proposed in recent literature, largely inspired by the pioneering work of Sivic and Zisserman [8] and built on the bag-of-features (BOF) principle for image representation. An analysis of the state-of-the-art reveals that these approaches and methods are typically centered around the idea of detecting and verifying correspondences between salient points in the given two images. The initial set of correspondences are detected based on matches between visual feature statistics measured in different images around found salient points. The correspondence verification step then serves to filter out unreliable correspondences. This verification is typically a spatial (geometric) one and involves geometric constraints to secure consistency of transformation of different image points. Spatial verification is the key to achieve high precision for object retrieval, especially when searching in large, heterogeneous image collections [8, 7].

Spatial verification can be done either explicitly, by iteratively building an optimized transformation model and fitting it to the initial correspondences (e.g., RANSAC-based model fitting approaches [7]), or implicitly, e.g., by verifying the consistency of the image points involved in the correspondences in the Hough transform space [2]. Our proposed method belongs to the category of implicitly verifying the correspondences. In contrast to most of the existing work in this direction, which focuses on individual correspondences, we are considering the pairwise relations between correspondences as well.

Pairwise geometric relations between the correspondences have not been frequently exploited for spatial verification. This may be due to the fact that the typical number N of initially detected correspondences is usually large, resulting in high computational complexity of pairwise comparisons, which can be modeled as $\mathcal{O}(N^2)$. This complexity makes exploitation of pairwise relations less attractive when operating on large image collections. Early work in this direction [3, 5] focuses on exploiting the pairwise relations directly from the initial correspondences and did not incorporate the global geometric relations derived from individual correspondences. The complexity of spatial verification in this case becomes too high to be applicable in a large image collection. Compared to these methods, our contribution is twofold. First, we significantly reduce the number of correspondences and in this way make the proposed spatial verification more tractable. Second, our pairwise geometric matching method combines both the global geometric relations derived from individual corre-

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage of CVPR 2015.

spondences and the local pairwise relations of pairs of correspondences for improved retrieval performance.

As illustrated in Figure 1, the geometric relations in terms of rotation and scaling between vectors formed by a pair of correspondences are closely related to the global geometric relations between images that are encoded in the transformation of the image regions surrounding the salient points. We start out from a standard representation of an image using local features. This representation typically involves detection of salient points in the image and representation of these points by suitable feature vectors describing local image regions around these points. Given two images and their salient points, the initial correspondences are found by the proximity between these points in the local feature space. We first employ a one-versus-one ('1vs1') matching strategy for the initial correspondence set to address redundancy among one-to-many correspondences, which typically arises when detecting correspondences between two images [4, 2]. By removing this redundancy, a significantly reduced correspondence set is generated. Then, similarly to [6], we reduce this set even further, by deploying Hough voting in the scaling and rotation transformation space. After these two steps, a large fraction of original correspondences are filtered out, which enables us to exploit pairwise geometric relations for spatial verification at a significantly reduced computational cost. Finally, a simple pairwise weighting method is devised to incorporate both the global geometric relations derived from individual correspondences and the local pairwise relations of pairs of correspondences.

The paper provides details of the implementation of the three-step method described above and contains extensive experimental results obtained on three reference datasets and in an object retrieval context. In particular, we assessed the quality of the pairwise image matching by deploying the proposed spatial verification method, the impact of improved image matching on the object retrieval performance and the related run time efficiency.

The results indicate the suitability of the proposed pairwise geometric matching method as a solution for large-scale object retrieval at an acceptable computational cost. The superiority of this method compared to the state-of-the-art solutions becomes evident in a context in which a high number of outliers in the initial correspondences generated by BOF and errors in detected features' scale, rotation and position hinder the fit of a specific model (e.g., RANSAC). Pairwise geometric matching encodes not only scale and rotation information derived from the local points, but also their locations. This is achieved by using global scale and rotation relation to enforce the local consistency of geometric relation derived from the locations of pairwise correspondences. Mapping the locations of points to pairwise rotation and scale makes the method more tolerant to

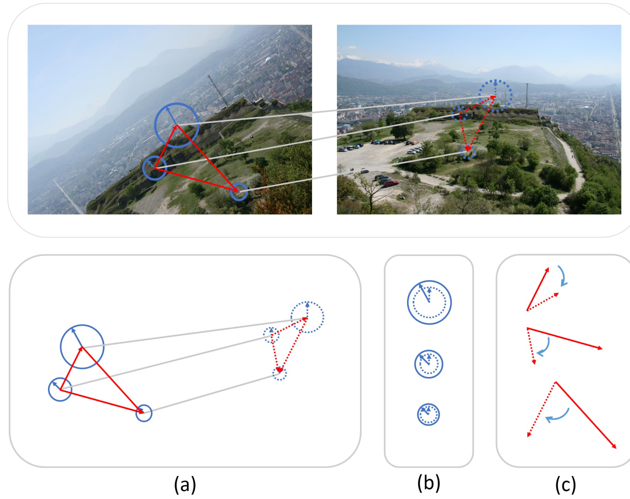

Figure 1. (a) Three correspondences found for two images, (b) global rotation and scale relations between images encoded in the transformation of the matched salient points from individual correspondences, (c) rotation and scale relations between vectors formed by pairwise salient points involved in the correspondences. Transformations in cases (b) and (c) are closely related to each other and can be used to emphasize each other for spatial verification.

the point detection noise. At the same time, using a number of filtering steps, the method significantly reduces the number of correspondences to work with, which makes it possible to maintain high image matching reliability at a significantly reduced computational cost.

References

[1] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *Proc. CVPR '12*, 2012. 1

[2] Y. Avrithis and G. Tolias. Hough pyramid matching: Speeded-up geometry re-ranking for large scale image retrieval. *IJCV*, 107(1):1–19, 2014. 1, 2

[3] G. Carneiro and A. D. Jepson. Flexible spatial models for grouping local image features. In *Proc. CVPR '04*, 2004. 1

[4] H. Jégou, M. Douze, and C. Schmid. On the burstiness of visual elements. In *Proc. CVPR '09*, 2009. 1, 2

[5] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *Proc. ICCV '05*, volume 2, pages 1482–1489 Vol. 2, 2005. 1

[6] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2

[7] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR '07*, 2007. 1

[8] J. Sivic and A. Zisserman. Video Google: a text retrieval approach to object matching in videos. In *Proc. ICCV '03*, 2003. 1