J. Vis. Commun. Image R. 25 (2014) 1416–1424

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Bipartite graph-based mismatch removal for wide-baseline image matching

Chen Wang, Kai-Kuang Ma *

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 8 October 2012
Accepted 10 December 2013
Available online 25 December 2013

Keywords:
Wide-baseline image matching
Region correspondence
Point correspondence
Coherent region pair
SIFT feature descriptor
Bipartite graph matching
Region similarity measurement metric
Hungarian method

ABSTRACT

The conventional wide-baseline image matching aims to establish point-to-point correspondence pairs across the two images under matching. This is normally accomplished by identifying those feature points with most similar local features represented by feature descriptors and measuring the feature-vector distance based on the nearest neighbor matching criterion. However, a large number of mismatches would be incurred especially when the two images under comparison have a large viewpoint variation with respect to each other or involve very different backgrounds. In this paper, a new mismatch removal method is proposed by utilizing the bipartite graph to first establish one-to-one *coherent region pairs* (CRPs), which are then used to verify whether each point-to-point matching pair is a correct match or not. The generation of CRPs is achieved by applying the Hungarian method to the bipartite graph, together with the use of the proposed region-to-region similarity measurement metric. Extensive experimental results have demonstrated that our proposed mismatch removal method is able to effectively remove a significant number of mismatched point-to-point correspondences.

1. Introduction

Wide-baseline image matching plays a vital role in many image processing and computer vision related tasks, such as object recognition [1,2], 3D reconstruction [3], motion estimation [4], to name a few. The matching operation is conducted over two separately acquired images of the same scene but from two widely disparate viewpoints. To perform this task, the conventional approach first identifies variable sizes of circular- or elliptical-shaped regions of interest on each image. Such identified regions, or the so-called *support regions*, usually contain some distinct features, such as corners and blobs. The center point of each support region is considered as a *feature point*, and a local feature vector or *feature descriptor* is formed based on this support region. For that, the well-known *scale invariant feature transform* (SIFT) method [1] is exploited in this paper to establish the feature descriptors, one for each detected feature point. This is achieved by computing the weighted gradient orientation histograms over a circular support region with the radius being set to the estimated scale; consequently, they are *invariant* to scaling, rotation, translation, and even up to a certain degree of illumination change. These SIFT descriptors are used to establish point-to-point correspondences between the two images under matching based on the *nearest neighbor matching criterion* [1].

It is well-recognized that the conventional wide-baseline image matching method based on the SIFT descriptors alone usually yields a large number of mismatches, especially when the two images under matching undergo a large viewpoint variation or involve very different backgrounds [1]. Such mismatches will impair the system performance of the target application. Therefore, it is imperative to detect and remove those mismatched point-to-point correspondence pairs. In this paper, a bipartite graph-based mismatch removal method is proposed to benefit those applications that involve wide-baseline image matching based on SIFT feature descriptors. The turnkey novelty lies in our proposed strategy on utilizing the developed *region-to-region* matching pairs (called *coherent region pairs* in this paper) as 'reference information' to re-evaluate each initially-established SIFT-based *point-to-point* matching pair and determine whether it is a correct match to retain or a mismatch to remove.

To achieve this goal, five main contributions are made in this paper and summarized as follows: (1) the generated *region-to-region* matching pairs are used to detect and remove those mismatched *point-to-point* pairs, (2) the principle of *spatial connectivity* is used to establish candidate region-to-region pairs, (3) the bipartite graph is proposed to model the entire set of candidate region-to-region pairs, (4) a *region similarity metric* is developed for measuring the degree of region's similarity or coherence, and

* Corresponding author.
E-mail addresses: wa0002en@e.ntu.edu.sg (C. Wang), ekkma@ntu.edu.sg (K.-K. Ma).

http://dx.doi.org/10.1016/j.jvcir.2013.12.013
1047-3203/

(5) the most *coherent region pairs* (CRPs) are identified by the Hungarian search method, together with the use of the proposed region-to-region similarity measurement metric.

The remaining of this paper is organized as follows. In Section 2, the relevant background and comparable state-of-the-art methods are succinctly reviewed. To realize our turnkey novelty as mentioned earlier, it is necessary to find out how to generate as many *region-to-region* matching pairs as possible, and this process is detailed in Section 3. In Section 4, a novel *region similarity metric* is proposed for the measurement of the *degree of similarity* or *coherence* between the two regions of each candidate region matching pair. In Section 5, a bipartite graph is exploited to represent the established candidate region matching pairs. The Hungarian method and the proposed *region similarity measurement metric* are then applied to the established bipartite graph for searching the optimal one-to-one CRPs. Section 6 presents our simulation results to show the effectiveness of using the developed one-to-one CRPs to identify mismatched point-to-point matching pairs for removal. Conclusions are drawn in Section 7.

2. Background

One intuitive and straightforward way for conducting mismatch removal is to first assume that a linear spatial transformation (e.g., affine or perspective transformation) is incurred between the two images under matching. A robust estimation of the transformation matrix can be obtained and used for computing the ground truth by using, say, the *random sample consensus* (RANSC) method [5], for example. If the spatial coordinates of the two feature points in an established point-to-point matching pair do not conform with the computed spatial-transformation coordinates, this matching pair will be regarded as a 'mismatch' [6,7]. However, this approach is not reliable, since a large number of mismatches still remain undetected. This is mainly due to the following two reasons. First, when the two images under matching are captured from two widely-disparate views, a simple linear spatial transformation is insufficient to depict the complex spatial transformation that is actually incurred between these two images. Second, when the precision (i.e., the percentage of correct matches) of the SIFT-based point-to-point matches falls below 50%, which is a scenario commonly encountered in the wide-baseline image matching, the spatial transformation matrix estimated by the RANSC-like methods becomes highly inaccurate [1].

To tackle this problem, Tuytelaars and Van Gool [8] first proposed an iterative method to remove the incorrect point-to-point matching pairs based on the projective transformation matrices (or the so-called *homographies*) that are estimated based on the feature points considered in various local regions, respectively. Obviously, the performance of this method is highly constraint by the accuracy of the estimated homographies. Ferrari et al. [2] proposed the *topological filter* that involves the *sidedness constraint* as a test condition; that is, for every three correct point-to-point matching pairs (thus, there are three feature points presented on each image), if one feature point lies on the right-hand side (or the left-hand side for the same argument) of the other two feature points on the same image, this topological relationship should remain the same among the three corresponding feature points on the other image as well. By identifying and discarding those point-to-point matching pairs that violate this sidedness constraint, their method had demonstrated some promising results on mismatch removal, but at the expense of high computational complexity. Zhou et al. further proposed two geometric coding techniques [9,10] by encoding the relative spatial relationship between each two feature points in an image into some binary values (or the so-called *signatures*). It is expected that for correct point-to-point correspondences, the signatures of the feature points in one image should be equal to the signatures of the corresponding feature points in the other image. Based on this test condition, the incorrect point-to-point matching pairs can be identified and removed.

Recently, the spatial layout of the feature points has been popularly incorporated into the design of image matching algorithms [11–14]. When a visual object undergoes a similarity transformation from one image to another, the Euclidian distance between any two feature points located on the same visual object in one image should be proportional to that of the corresponding two feature points of the same object appeared in the other image only by a factor of a real-valued constant, which is equal to the ratio of the visual object's scale values measured on the two images, respectively. This distance-preserving property is denoted as the *pairwise spatial consistency principle* [12], which is used to evaluate whether each established point-to-point correspondence is a correct match or a mismatch. However, when a large viewpoint variation is encountered, this principle is oftentimes hard to satisfy and could cause a large number of mismatches being undetected.

The goal of our work is to identify and eliminate mismatches from the established SIFT-based point-to-point matching pairs. The turnkey novelty lies in the strategy of constructing *one-to-one region-to-region* matching pairs, which are coined as the *coherent region pairs* (CRPs) in our work, followed by utilizing them as the reliable reference information to verify whether each existing *point-to-point* correspondence is a mismatch or not. If a point-to-point matching pair is considered as a correct match, then the following conditions must be met: if one end point of a given correspondence pair (or a link) falls in a region of a CRP in one image, then the other end point of the same link must fall in the corresponding region of the same CRP presented in the other image. Otherwise, this point-to-point matching pair will be regarded as a mismatch and should be discarded.

Now, the first question is: how to construct CRPs. For that, we utilize the established SIFT-based point-to-point matching pairs to guide the merging of the segmented image regions that are spatially connected to form the *candidate* region matching pairs. Note that, each candidate region matching pair could be one-to-one, one-to-many, or many-to-one correspondence types. Therefore, such pairs can be best represented by a bipartite graph model which is widely used in the field of object and image retrieval [15–17]. To identify the *optimal* and *one-to-one* CRPs over the graph, the Hungarian method [18] is applied by maximizing the total degree of region similarity under the constraint that each region-to-region correspondence must be one-to-one. Note that the region similarity between the two regions of each candidate region matching pair is measured by using the proposed region similarity measurement metric. Extensive simulation results have shown that our proposed bipartite graph-based mismatch removal method could effectively identify a large number of mismatches for removal.

3. Proposed region-to-region matching method

3.1. Establishment of SIFT-based point-to-point matching pairs

Given a wide-baseline acquired image pair the SIFT feature descriptors [1] are formed on each image—one feature descriptor for each feature point detected by the *affine covariant region detector* [22,23]. For each SIFT feature descriptor f given in one image, two most similar SIFT feature descriptors g and h can be identified from the other image according to the l_2-norm distance measurement (with respect to the feature descriptor f). Further assume that the descriptor g is ranked higher than the descriptor h, both with respect to the descriptor f. If the *nearest neighbor matching criterion*