ARTICLE IN PRESS

ISPRS Journal of Photogrammetry and Remote Sensing xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Recent developments in large-scale tie-point matching

Wilfried Hartmann *, Michal Havlena, Konrad Schindler

Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, Switzerland

A R T I C L E I N F O

Article history:
Received 28 May 2015
Received in revised form 30 August 2015
Accepted 16 September 2015
Available online xxxx

Keywords:
Image matching
Unordered image sets
Feature selection
Approximate match graph
Visual vocabulary

A B S T R A C T

Feature matching – i.e. finding corresponding point features in different images to serve as tie-points for camera orientation – is a fundamental step in photogrammetric 3D reconstruction. If the input image set is large and unordered, which is becoming increasingly common with the spread of photogrammetric recording to untrained user groups and even crowd-sourced geodata collection, the bottleneck of the reconstruction pipeline is the matching step, for two reasons. (i) Image acquisition without detailed viewpoint planning requires a denser set of viewpoints with larger overlaps, to ensure appropriate coverage of the object of interest and to guarantee sufficient redundancy for reliable reconstruction in spite of the unoptimised network geometry. As a consequence, there is a large number of images with overlapping viewfields, resulting in a more expensive matching step than, say, a regular block geometry. (ii) In the absence of a carefully pre-planned recording sequence it is not even known which images overlap. One thus faces the even bigger challenge to determine which pairs of images even *can have* tie-points and should therefore be fed into the matching procedure. In this paper we attempt a systematic survey of the state-of-the-art for tie-point generation in unordered image collections, including recent developments for very large image sets.

Contents

* Corresponding author.
E-mail address: hartmann@geod.baug.ethz.ch (W. Hartmann).

http://dx.doi.org/10.1016/j.isprsjprs.2015.09.005
0924-2716/

Please cite this article in press as: Hartmann, W., et al. Recent developments in large-scale tie-point matching. ISPRS J. Photogram. Remote Sensing (2015), http://dx.doi.org/10.1016/j.isprsjprs.2015.09.005

1. Introduction

Reconstructing 3D objects from 2D photographic images is one of *the* fundamental tasks of photogrammetry and computer vision. Over the past decades, the switch to digital photography and the continuing growth of computing power, storage, and transmission bandwidth have brought about important changes. Both the amount of available image data for 3D reconstruction and the number of applications that use photogrammetric techniques have greatly increased. At the same time, processing methods have further developed and nowadays allow one to generate, in a highly automated fashion, fairly accurate reconstructions even from compact cameras or even mobile phone cameras, paving the way for recent trends like photogrammetric recording from small drones, or crowd-sourced images for 3D reconstruction (Agarwal et al., 2009). The reconstruction of 3D objects consists of two steps: first the orientation of the images in a common 3D coordinate frame (a.k.a. camera pose estimation); and second the subsequent generation of a dense point cloud or surface model. The present paper deals with the first step, with a particular emphasis on large, unordered image sets.[1] Dense matching and surface reconstruction have also witnessed exciting developments (e.g. Hirschmuller, 2008; Hiep et al., 2009; Jancosek and Pajdla, 2011; Bulatov et al., 2011; Ahmadabadian et al., 2013; Stentoumis et al., 2014), but are not discussed here. In large projects dense reconstruction is always carried out as a separate step which operates locally, using for each portion of the scene only the small subset of the camera network that covers that portion.

Fig. 1 illustrates the standard 3D modeling pipeline. It starts with the acquisition of images that cover the scene of interest (or with the collection of existing images, e.g. when working with crowd-sourced or historic data). Next, repeatable and stable *interest points* are extracted from the input images. The neighborhood of each interest point is encoded with a *descriptor* – this can in principle be the raw intensity pattern, but more often one chooses descriptors that are more invariant to geometric distortions and illumination changes. By comparing descriptor values of the points in different images, one finds point *matches* between image pairs. The matches are filtered through (usually pairwise or triplet-based) geometric verification to discard mismatches and keep only correct *tie-points*, i.e. projections of the same 3D scene points. The correspondences serve as input for the *pose estimation*, normally by a combination of epipolar geometry estimation, triangulation, and spatial resection. Finally, the camera orientations and the sparse tie-point cloud are refined with bundle adjustment. The point cloud can optionally be densified with dense matching of the oriented images.

This paper is dedicated to image matching. One consequence of the new photogrammetric workflows is that in some applications one is faced with large *unordered* input image sets. In this context "unordered" means that it is not known in advance which images share a common viewfield and should be matched. Examples include crowd-sourced imagery from the Internet; recordings from small micro aerial vehicles (MAVs) with low-quality navigation systems; but also general close-range projects – experience shows that untrained users have great difficulties to record images in a predefined pattern, and even experts are slowed down significantly when they have to follow strict recording protocols in complex environments such as industrial installations.

Classically, photogrammetry has preferred *ordered* image sets (Luhmann et al., 2014), in which the overlapping images (or even the approximate relative orientations) are known before the start of the pose estimation. These can be obtained in two ways: either through carefully planned recording, or by observing the camera poses during image acquisition with external sensors. Metrology has focussed on settings where the overlaps are unknown, but there is a small number of marked tie-points (e.g. high-contrast stickers) that can be disambiguated by their relative positions. We do not suggest that such procedures are obsolete – in situations like aerial mapping or industrial quality control ordered image sets, respectively artificial tie-points, have obvious advantages. Rather, the problems and solutions surveyed in this paper arise in new applications that broaden the scope of image-based 3D measurement. Still, the capability to orient arbitrary images of the same scene can also benefit the classical approaches, e.g. not following a strict planning can save field time, allows one to react flexibly to unexpected problems with lighting, occlusions etc., and simplifies larger recording campaigns with multiple people. The extreme case, which has received a lot of attention in computer vision, is 3D reconstruction from crowd-sourced images taken from photo-sharing sites like Flickr (Yahoo!, 2005) or Panoramio. Note, in that setting matching also serves to weed out irrelevant pictures with misleading tags, such as portraits with blurred backgrounds or images from different locations with the same name.

Modern projects can consist of thousands or even millions of unordered images (Heinly et al., 2015). This means that image matching becomes the bottleneck – even with today's computing power naive brute-force matching is infeasible. To address situations where one can no longer compare every pair of interest point descriptors for every pair of potentially overlapping images, different techniques have been developed. There are three main lines of attack to bring down the matching time without compromising the reconstruction, see Fig. 2: reduce the number of feature points per image, by finding those which are most suitable for matching; reduce the number of images, by finding those which are most important to cover the scene; or reduce the number of potential image pairs, by finding those which are most likely to overlap and match. In the following we will present a systematic overview of these developments. In Section 2 we review and compare different methods to speed up the matching of a given image pair, by reducing the number of feature points per image. Note that this part is applicable also with ordered image sets. Section 3 then presents ways to limit the number of image pairs to be tested, by either selecting promising image pairs, or by pruning the input image set without losing important views, or by avoiding the pairwise matching of images altogether.

[1] For the purposes of this paper relative orientation into a photogrammetric model is sufficient. A metric scale or full absolute orientation can of course be introduced via known distances or ground control points (GCPs).

Please cite this article in press as: Hartmann, W., et al. Recent developments in large-scale tie-point matching. ISPRS J. Photogram. Remote Sensing (2015), http://dx.doi.org/10.1016/j.isprsjprs.2015.09.005