ISPRS Journal of Photogrammetry and Remote Sensing 121 (2016) 113–127

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Skeletal camera network embedded structure-from-motion for 3D scene reconstruction from UAV images

Zhihua Xu [a,b,c], Lixin Wu [d,*], Markus Gerke [c], Ran Wang [e], Huachao Yang [d]

[a] College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China
[b] Key Laboratory of Environment Change & Natural Disaster, Ministry of Education, Beijing Normal University, Beijing 100875, China
[c] Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands
[d] School of Environment Science & Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China
[e] School of Geography, Beijing Normal University, Beijing 100875, China

ARTICLE INFO

Article history:
Received 19 November 2015
Received in revised form 23 August 2016
Accepted 24 August 2016

Keywords:
3D scene reconstruction
Hierarchical degree bounded maximum spanning tree
Skeletal camera network
Structure-from-motion
Unmanned aerial vehicles

ABSTRACT

Structure-from-Motion (SfM) techniques have been widely used for 3D scene reconstruction from multi-view images. However, due to the large computational costs of SfM methods there is a major challenge in processing highly overlapping images, e.g. images from unmanned aerial vehicles (UAV). This paper embeds a novel skeletal camera network (SCN) into SfM to enable efficient 3D scene reconstruction from a large set of UAV images. First, the flight control data are used within a weighted graph to construct a topologically connected camera network (TCN) to determine the spatial connections between UAV images. Second, the TCN is refined using a novel hierarchical degree bounded maximum spanning tree to generate a SCN, which contains a subset of edges from the TCN and ensures that each image is involved in at least a 3-view configuration. Third, the SCN is embedded into the SfM to produce a novel SCN-SfM method, which allows performing tie-point matching only for the actually connected image pairs. The proposed method was applied in three experiments with images from two fixed-wing UAVs and an octo-copter UAV, respectively. In addition, the SCN-SfM method was compared to three other methods for image connectivity determination. The comparison shows a significant reduction in the number of matched images if our method is used, which leads to less computational costs. At the same time the achieved scene completeness and geometric accuracy are comparable.

1. Introduction

3D scene reconstruction plays an important role in heritage documentation (El-Hakim et al., 2004), urban planning (Verdie et al., 2015), virtual reality (Feng et al., 2014), and disaster management (Bulatov et al., 2014; Ferworn et al., 2011; Gerke and Kerle, 2011). Many studies have investigated the benefits of Structure-from-Motion (SfM) techniques for 3D scene reconstruction from multi-view images (Pollefeys et al., 1999; Scaramuzza et al., 2006; Sturm and Triggs, 1996). Nowadays, many commercial and open source software packages integrating a SfM method are available (Koutsoudis et al., 2014). Unmanned aerial vehicles (UAV), showing great advantages in operational cost and flexibility, have been increasingly used to date to capture multi-view images (Colomina and Molina, 2014; Lucieer et al., 2014; Zhang and Kovacs, 2012). In particular, UAV images which are sequentially acquired with high geometrical resolution and large overlap, are suitable for 3D scene reconstruction. Many studies have investigated the use of UAV images and SfM methods for 3D reconstruction in landslides (Niethammer et al., 2012), agriculture (Dandois and Ellis, 2013), topography (Woodget et al., 2015), and disaster scenarios (Vetrivel et al., 2015).

In the literature, many efforts have been undertaken to reduce the computational cost associated especially to image matching and bundle adjustment within SfM. In this work, the attention is mainly devoted to image matching, one of the most time-consuming phases. An effective approach for computational savings is to embed a topologically connected camera network (TCN) within SfM as a constraint (Rupnik et al., 2013; Xu et al., 2014a). A TCN is also referred to as image connectivity graph, which identifies the connections between the images (Snavely et al., 2010). The TCN-embedded SfM allows only the connected image pairs to be involved in matching. This method has been successfully used for processing UAV images in both nadir and oblique

* Corresponding author.
E-mail addresses: hzxurs@mail.bnu.edu.cn (Z. Xu), awulixin@263.net (L. Wu), m.gerke@utwente.nl (M. Gerke), wangr0225@163.com (R. Wang), 932400012@qq.com (H. Yang).

http://dx.doi.org/10.1016/j.isprsjprs.2016.08.013
0924-2716/

views. For instance, Douterloigne et al. (2010) utilized Global Navigation Satellite System (GNSS) data to generate a TCN for matching nadir UAV images. Rupnik et al. (2015) used the on-board sensors associated with GNSS and an Inertial Measurement Unit (IMU) to generate a TCN for dealing with oblique images. Moreover, a rough point cloud-derived TCN was employed for matching unordered UAV images or video frames (Alsadik et al., 2015; PhotoScan, 2013).

Many methods for generating a TCN are decision-making based, either on image footprint conjunction (Douterloigne et al., 2010; Rupnik et al., 2013) or on tie-points distribution (Alsadik et al., 2015; Castillo et al., 2015; PhotoScan, 2013). The major shortcoming of the mentioned approaches is that they operate on a pure direct adjacency principle, e.g. a minimum overlap between images. This, however, does not consider the entire image topology and may lead to adding matching pairs which are not relevant for the network, but cause additional matching cost, or it may lead to a loss of relevant image pairs, respectively.

In the literature, a spanning tree has been commonly used for optimally removing edges which are considered non-essential from a weighted graph (Held and Karp, 1970; Helmi and Rahmani, 2014). In particular, the extraction of the skeletal camera network (SCN) follows a maximum spanning tree (MST) approach (Gavril, 1987), which find a subset of edges from a weighted TCN to achieve a maximal total weight (i.e. overlap) with a minimally required number of edges. For a wide range of applications, several variants of MST have been carried out with maximum connectivity and degree bounded considerations (Gouveia et al., 2014; Helmi and Rahmani, 2014; Katagiri et al., 2012).

This paper aims to construct a SCN to improve the computational efficiency of SfM for 3D scene reconstruction from a large set of highly overlapping UAV images. Specifically, we first construct a weighted TCN to identify the connections between UAV images, then extract a SCN by deleting the non-essential edges from the TCN, and finally embed the SCN into a SfM method (SCN-SfM) in order to restrict pairwise matching to images which are connected in the SCN. This paper is organized as follows: Section 2 starts by describing the framework of the proposed method, followed by illustrations of each component of the proposed method and baseline methods used for comparison; Section 3 describes the experimental sites and data used; the experimental results are present in Section 4, followed by discussion in Section 5 and conclusions of this study in Section 6.

2. Methodology

The framework of the proposed SCN-SfM method for 3D scene reconstruction from UAV images is sketched in Fig. 1. It contains three main components, namely (a) TCN construction, (b) SCN extraction, and (c) SCN embedded SfM. We first construct a TCN to represent the dataset by a weighted graph, considering the connections between UAV images, using the flight control data. Subsequently, we obtain an optimal SCN by iteratively deleting the non-essential edges in the TCN using a novel spanning tree, i.e. the hierarchical degree bounded maximum spanning tree (HDB-MST). Finally, the optimal SCN is embedded into the SfM method for 3D scene reconstruction, where only the connected image pairs represented by the SCN are considered for tie-point matching. The three components of the proposed method are described in detail in the following sections.

2.1. TCN construction

We use the flight control data to construct a weighted TCN representing the connections between the images in a given UAV collection (Rupnik et al., 2013; Xu et al., 2014a). For n images, we define the corresponding TCN by a weighted directed graph $G = (V, E)$ with its node set $V = \{v_1, \ldots, v_n\}$ and the edge set $E = \{e_{i,j} : i, j = 1, \ldots n\}$. Each node represents one image, and the edge represents a connected image pair. We represent the graph by a triangular adjacency matrix, in which we write $c_{i,j} = 1$ to denote $e_{i,j} \in E$, and we use the value of $e_{i,j}$ to denote the edge weight between a pair of nodes (v_i, v_j). Here, the edge weight is characterized by the overlap between an image pair. For the TCN construction, the footprints of each image are first computed from flight control data and by projecting the image outline onto an elevation model (Rupnik et al., 2013; Xu et al., 2014a). Next, the connections between the image pairs are identified through an image topology analysis (Xu et al., 2014a; Xu et al., 2015), followed by the calculation of overlap between the connected images. Fig. 2 illustrates five main categories of an overlapping image pair (v_1, v_2). The overlap of a connected image pair can be estimated by the following procedure: given a connected image pair (v_1, v_2), we first find the convex hull of the overlapping area $e_{1,2}$ by means of Graham scan algorithm (Kong et al., 1990), then divide the convex hull into multiple triangles and finally estimate the overlap by Eqs. (1) and (2).

$$e_{1,2} = \sum_{i=2}^{m-1} S_{\Delta t_1 t_i t_{i+1}} \tag{1}$$

where $t_i = (x_i, y_i)$ is the ith of m nodes of the convex hull, and $S_{\Delta t_1 t_i t_{i+1}}$ is the area of one triangle composed by node set $\{t_1, t_i, t_{i+1}\}$.

$$S_{\Delta t_1 t_i t_{i+1}} = \frac{1}{2}\left|(x_i - x_1)(y_{i+1} - y_1) - (x_{i+1} - x_1)(y_i - y_1)\right| \tag{2}$$

The use of the K-nearest-neighbors algorithm (Cover and Hart, 1967) can help to find corresponding overlapping images since the number of overlapping candidates is quadratic in the number of images.

Fig. 3 shows the weighted TCN of an example UAV collection with 11 images. Note that the information on the diagonal and the last column are auxiliary information described in the next subsections.

2.2. SCN extraction

The SCN minimizes the number of edges in TCN. We consider a representation in a graph $G = (V, E)$. The minimization is conducted using the proposed maximum spanning tree, i.e. HDB-MST. The application of HDB-MST is achieved by using two main procedures associated with *hierarchical TCN representation* and *SCN extraction from the hierarchically structured TCN* in an iterative configuration. The execution of the proposed HDB-MST is summarized by the following pseudo-codes:

```
while do
  let G = (V, E) be a weighted graph that represent TCN
  hierarchically restructure TCN and divide G = (V, E) into a
   set of n subgraphs as discussed in Section 2.2.1
  extract SCN from the hierarchically structured TCN as
   discussed in Section 2.2.2
  if SCN equals to TCN then
    exit
  else
    update TCN by SCN
  end if
end while
return SCN
```