ISPRS Journal of Photogrammetry and Remote Sensing 104 (2015) 1–17

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Line matching based on planar homography for stereo aerial images

Yanbiao Sun [a,c], Liang Zhao [b,c], Shoudong Huang [c], Lei Yan [a,*], Gamini Dissanayake [c]

[a] Institute of Remote Sensing and GIS, School of Earth and Space Science, Peking University, Beijing 100871, China
[b] Hamlyn Centre for Robotic Surgery, Department of Computing, Faculty of Engineering, Imperial College London, London, UK
[c] Centre for Autonomous Systems, Faculty of Engineering and Information Technology, University of Technology, Sydney, NSW 2007, Australia

ARTICLE INFO

Article history:
Received 25 June 2014
Received in revised form 1 December 2014
Accepted 1 December 2014
Available online 13 March 2015

Keywords:
Line matching
Planar homography
Point feature correspondences
Coplanarity
Bundle adjustment

ABSTRACT

We propose an efficient line matching algorithm for a pair of calibrated aerial photogrammetric images, which makes use of sparse 3D points triangulated from 2D point feature correspondences to guide line matching based on planar homography. Two different strategies are applied in the proposed line matching algorithm for two different cases. When three or more points can be found coplanar with the line segment to be matched, the points are used to fit a plane and obtain an accurate planar homography. When one or two points can be found, the approximate terrain plane parallel to the line segment is utilized to compute an approximate planar homography. Six pairs of rural or urban aerial images are used to demonstrate the efficiency and validity of the proposed algorithm. Compared with line matching based on 2D point feature correspondences, the proposed method can increase the number of correctly matched line segments. In addition, compared with most line matching methods that do not use 2D point feature correspondences, the proposed method has better efficiency, although it obtains fewer matches. The C/C++ source code for the proposed algorithm is available at http://services.eng.uts.edu.au/~sdhuang/research.htm.

1. Introduction

Line extraction and matching is crucial for scene reconstruction, motion segmentation, structure from motion, robot navigation and so forth (Tang et al., 2006; Schindler, 2006). For urban aerial images, line correspondences can help reconstruct profiles of the many man-made objects contained within them, which is an important element of city reconstruction.

The extraction and matching of features, which mainly includes point features and line features, provides important data for many applications in photogrammetry, computer vision, etc. Compared with line feature extraction and matching, point feature extraction and matching is more mature, with various approaches proposed (Ke and Sukthankar, 2004; Novak et al., 2011; Mikolajczyk and Schmid, 2005). Among them, Lowe proposed Scale Invariant Feature Transformation (SIFT), a robust and reliable approach invariant to transformation, scale, and illumination changes (Lowe, 2004). This method has been widely applied in many fields. In addition, to rapidly extract and match point correspondence from a pair of large-size aerial images, L^2-SIFT has been proposed (Sun et al., 2014), in which a large aerial image is split into blocks and matching is conducted for every corresponding block using SiftGPU (Wu, 2007). In this paper, L^2-SIFT will be used to extract and match 2D point correspondences, which provide additional information for line matching.

Extracted line segments are unreliable for matching because strong disambiguating geometric constraints are unavailable; hence, line segment matching is significantly more difficult than point feature matching. Consequently, many approaches have been proposed to conduct line matching from stereo images in the last two decades. In 1997, Schmid and Zisserman proposed automatic line matching across views, achieving matching by combining gray-level information and multiple view geometric relations (Schmid and Zisserman, 1997). Each segment was treated as a list of points, and point correspondences around the line segment were found using epipolar geometry. The average cross-correlation scores of all corresponding points acted as a similarity measure of the lines. In 2001, Heuel et al. used homogeneous vector and matrix to represent geometric entities and their related uncertainties and then applied statistical tests to match and reconstruct 3D lines without any given threshold (Heuel and Forstner, 2001). However, this method can only handle more than three images

* Corresponding author.
E-mail addresses: syb51@pku.edu.cn (Y. Sun), liang.zhao@imperial.ac.uk (L. Zhao), Shoudong.Huang@uts.edu.au (S. Huang), lyan@pku.edu.cn (L. Yan), Gamini.Dissanayake@uts.edu.au (G. Dissanayake).

http://dx.doi.org/10.1016/j.isprsjprs.2014.12.003

that are not stereo images. In 2005, to reconstruct DSM with good quality, Zhang used the cross-correlation method to match edges from linear array images (Zhang, 2005) with the help of the local geometric and photometric attributes. However, this method cannot be directly applied to process aerial images captured by a frame camera. In 2005, Bay et al. proposed automatic line matching for color images, which obtained initial line correspondences followed by further matches iteratively based on a topological filter (Bay et al., 2005). However, this method depends strongly on color information and often fails when processing remote sensing images. In 2009, Wang et al. proposed the mean standard deviation line descriptor method (MSLD) to achieve line matching. In this case, a line segment was regarded as a list of points, and a histogram of the image gradient in a pixel support region was then computed (Wang et al., 2009); the mean and standard deviation of the histogram were obtained as line segment descriptors, similar to SIFT descriptors. In 2012, to improve matches from low-texture scenes, a method with appearance similarities and geometric constraints was proposed (Zhang and Koch, 2012). Direction histograms were used to estimate the perspective transformation and rotation angle, thereby reducing tentative matches. In 2012, Ok et al. used probability density functions and seven relational constraints to obtain more robust matches from urban aerial images (Ok et al., 2012). To improve efficiency and match accuracy, additional information, including DSM, LiDAR, GIS layer, and surface details, was used (Chehata et al., 2002; Habib et al., 2010). Later in 2014, a Gaussian mixture model and expectation maximization were used for line segment matching to achieve line segment registration using satellite images and a GIS layer in Long et al. (2014). While the simple affine transformation used in Long's approach may be good enough for satellite images, it may not be feasible for aerial images. In summary, the above methods depend heavily on gray-level and gradient information so will often fail to process images without strong disambiguating neighboring line segments. They also have high computational complexity.

To improve efficiency and provide enhanced matches from low-texture images, some methods have been proposed using other information, including geometric properties and point correspondence. In 1990, Stelmazyk attempted to use a nearest-line strategy to guide line matching. Unfortunately, his approach was only suitable for image tracking, in which images or line segments have relatively small changes (Stelmazyk, 1990). In 2009, Wang et al. proposed the Line Signatures method with matching to a wide-baseline image, in which the angle and length ratio between line segments are computed to describe a pair of line segments (Wang et al., 2009). However, this requires accurate location of the endpoints. In 2011, Elaksher proposed automatic line matching based on geometric properties, including the length of the perpendicular from the origin to the line and the angle from the positive x-axis to the perpendicular (Elaksher, 2011). However, this method also requires gray-level properties and is therefore slow.

Considering that accurate point correspondences are easily obtained by some mature point feature extraction and matching algorithms, some researchers have proposed line matching methods built on point information. There are two representative papers for this approach. First, the projective invariant method for line matching using two lines and two points was proposed by Lourakis et al. (2000). Unfortunately, this approach has a high computational cost. Second, two point-line invariant methods to process close-range images were proposed by Fan et al. (2012). Similar to other approaches based on geometrical information, these methods assumed that the surrounding points were coplanar with the line in 3D space. One method based matching on the projective-invariant, which can be computed using four 2D image points and a line. The projective-invariant is regarded as a similarity measure between two tentative line segments and determines whether two line segments should be matched or not. In practical experiments, for some line segments, it may be difficult to find four points, and thus, the total number of line matches will be reduced. The proposed alternative point-line invariant method relaxed the requirement of four points to only two points for computing the affine-invariant for a similarity measure. To increase line matches, we propose an approach based on planar homography. The software package based on the affine-invariant method (Fan et al., 2012) is publicly available, and we compared those data to our proposed method.

Similar to the affine-invariant and projective-invariant methods, we assume that surrounding points are coplanar with the line segment in 3D space. Without assuming an affine camera, two different strategies are applied in the proposed line matching algorithm for two different cases. In the first case, when three or more points coplanar with the line segment can be found, we fit a plane to the 3D points and calculate the projective transformation between two lines. In the second case, in which only one or two points can be found, we use a parallel plane to calculate an approximate transformation. Because many line segments in aerial images are approximately parallel to the terrain plane, we consider the terrain plane as the parallel plane. The 3D points, which are used to fit the plane containing the 3D line or the terrain plane, can be obtained from 2D point correspondences. Rather than using the point-line invariant in 2D image space, 2D point correspondences are first triangulated into 3D points by Bundle Adjustment (BA).

Let us first define our notation. We express the camera projection matrix as $P = K[R|t]$ or $P = [A|a]$, where K is the camera calibration matrix; R and t are camera rotation and translation, respectively; $A = KR$; and $a = Kt$. A pair of images is considered to comprise both source and target images, I_S and I_T, respectively. We use a point feature extraction and matching algorithm to obtain point feature correspondences, which are represented as $P^F_{2D} = \{x^i_{PS}, x^i_{PT}\},\ i = 1, 2, \dots, N_P$, where x^i_{PS} and x^i_{PT} are a pair of correspondences from the source image I_S and the target image I_T, respectively, and N_P is the number of matched 2D features. The corresponding triangulated 3D points are written as $P^F_{3D} = \left\{X^i\right\},\ i = 1, 2, \dots, N_P$, in the local coordinate system. $L_S = \left\{l^i_S\right\},\ i = 1, 2, \dots, N_{Ls}$, represents a set of line segments extracted from the source image, where N_{Ls} is the number of line segments, and each line segment is presented as $l^i_S = \{x^1_{LS}, x^2_{LS}\}$, including two end-points x^1_{LS} and x^2_{LS}. Similarly, $L_T = \left\{l^i_T\right\},\ i = 1, 2, \dots, N_{LT}$, expresses line segments from the target image. In addition, lines in three-dimensional space are expressed by L, where a matched line segment $L^k = \left\{l^i_S, l^j_T\right\},\ i = 1, 2, \dots, N_{Ls},\ j = 1, 2, \dots, N_{LT}$, corresponds to two line segments of the two images. Finally, we use $P^L = [n^T_L, d]^T$ to express a 3D plane, where n_L, a 1×3 vector, is the normal vector of the plane, and d is the depth information of the plane. Thus, the main work of this paper was to find the corresponding line segment $l^j_T \in L_T$ in the target images I_T when matching a putative line segment $l^i_S \in L_S$ from the source image I_S.

In Section 2, we briefly introduce planar homography, which provides an efficient tool for mapping two corresponding line segments. Section 3 discusses the implementation of line matching based on an accurate projective transformation accomplished by fitting a plane using three or more triangulated 3D points when sufficient points are found. Section 4 discusses line matching based on an approximate transformation using the approximate terrain plane to match line segments when few points are found.