ISPRS Journal of Photogrammetry and Remote Sensing 139 (2018) 241–254

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

An enhanced multi-view vertical line locus matching algorithm of object space ground primitives based on positioning consistency for aerial and space images

Ka Zhang [a,b,c,*], Yehua Sheng [a,b,c,1], Meizhen Wang [a,b,c,1], Suxia Fu [a,b]

[a] Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, 1# WenYuan Road, Nanjing 210023, PR China
[b] State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), 1# WenYuan Road, Nanjing 210023, PR China
[c] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, 1# WenYuan Road, Nanjing 210023, PR China

ARTICLE INFO

Article history:
Received 25 July 2017
Received in revised form 14 March 2018
Accepted 15 March 2018

Keywords:
Multi-view image matching
Object space ground primitive
Positioning consistency
Vertical line locus

ABSTRACT

The traditional multi-view vertical line locus (TMVLL) matching method is an object-space-based method that is commonly used to directly acquire spatial 3D coordinates of ground objects in photogrammetry. However, the TMVLL method can only obtain one elevation and lacks an accurate means of validating the matching results. In this paper, we propose an enhanced multi-view vertical line locus (EMVLL) matching algorithm based on positioning consistency for aerial or space images. The algorithm involves three components: confirming candidate pixels of the ground primitive in the base image, multi-view image matching based on the object space constraints for all candidate pixels, and validating the consistency of the object space coordinates with the multi-view matching result. The proposed algorithm was tested using actual aerial images and space images. Experimental results show that the EMVLL method successfully solves the problems associated with the TMVLL method, and has greater reliability, accuracy and computing efficiency.

1. Introduction

Image matching is an important research topic in the field of digital photogrammetry and computer vision that aims to automatically find homologous pixels of the same object in two or more overlapping images (Gruen, 2012; Leberl et al., 2010). As a key technical procedure in the automatic extraction of geoinformation from images, image matching plays an important role in applications such as automatic aerotriangulation (Krupnik and Schenk, 1997; Li et al., 2016), 3D information acquisition (Zhang et al., 2008; Tao, 2000; Gerke and Kerle, 2011; Zhang et al., 2015a), 3D scene reconstruction (Pollefeys et al., 2008; Cheng et al., 2011), multi-source data fusion (Gerke and Xiao, 2014; Yang and Chen, 2015), and robot vision navigation (Nalpantidis and Gasteratos, 2010). Image matching methods can be divided into two categories based on the characteristics of the matching primitive used: the image-space-based method and the object-space-based method.

The image-space-based matching method searches for homologous pixels in image space. It involves four steps: matching cost computation, cost aggregation, disparity computation and disparity refinement (Scharstein and Szeliski, 2002). In this method, a pixel or feature (point, line or region) in the image is used as the matching primitive, and a grayscale or feature vector in the image space is utilized to define and compute the matching cost.

There are two general types of image-space-based matching methods: local image matching methods and global image matching methods. In local image matching methods, a support window centered on the matching pixel with a prescribed size (such as 5×5 pixels) is designated, and all pixels in the support window are used in the process of cost aggregation. Next, the disparity value of the matching pixel that satisfies local optimization is produced using geometric constraints such as epipolar lines and triangular meshes (Kim, 2000; Bulatov et al., 2011; Wu et al., 2012). In global image matching methods, a global energy function with a data term and smoothing term is first determined for all potential disparity values of the matching pixel; next, the global optimum disparity value that minimizes the global energy is

* Corresponding author at: Key Laboratory of Virtual Geographic Environment, School of Geography Science, Nanjing Normal University, No. 1, Wenyuan Road, Nanjing, China. Academic visitor in the School of Engineering, Newcastle University, King's Gate, Newcastle upon Tyne, United Kingdom.

E-mail addresses: zhangka81@126.com (K. Zhang), shengyehua@njnu.edu.cn (Y. Sheng), wangmeizhen@njnu.edu.cn (M. Wang), 2448952097@qq.com (S. Fu).

[1] Co-first author.

https://doi.org/10.1016/j.isprsjprs.2018.03.017
0924-2716/

produced by iterative computation of the energy function (Birchfield et al., 2007; Szeliski et al., 2008; Yang et al., 2015). However, global matching methods involve correctly setting many parameters that do not have physical meanings, and they have high computational cost. The matching efficiency will be low when global matching methods are applied in multi-view dense matching (Cassisa, 2010).

Regarding matching cost computation, two types of costs are frequently used: one is based on the grayscale values of pixels and the other on the post-processed features of the grayscale values. The gray-based matching cost is computed by directly comparing the similarity measure between the grayscale values of two pixels. Some frequently used gray-based measures are the absolute difference (AD) (Birchfield and Tomasi, 1998), sum of squared differences (SSD) (Banks and Corke, 2001), sum of absolute differences (SAD) (Manap and Soraghan, 2012), arithmetic distance (Yang et al., 2008), normalized cross correlation (NCC) (Hirschmüller and Scharstein, 2009), and zero mean normalized cross correlation (ZNCC) (Mattoccia et al., 2008). The first step in feature-based matching cost computation is to perform *a* feature transform on images to obtain feature vectors of matching pixels. Several kinds of feature transform may be used, including scale invariant feature transform (SIFT) (Lowe, 2004; Zhang et al., 2011; Yang et al., 2012; Sun et al., 2014), Harris transform (Gueguen and Pesaresi, 2009), wavelet transform (Zhou et al., 2007), census transform (Samadi and Othman, 2013; Miron et al., 2014), contourlet transform (Zhang et al., 2015b), and affine-invariant feature or local self-similarity descriptor extraction (Cheng et al., 2008; Engin et al., 2008). The second step involves computing the similarity measure using the Euclidean distance or NCC between feature vectors.

Regarding cost aggregation, the following methods are often used in local matching methods: fixed matching window with fixed weight (Kang et al., 2001), fixed matching window with adaptive weight (Yoon and Kweon, 2006; Maeztu et al., 2011; Hosni et al., 2013; Stentoumis et al., 2014; Xu et al., 2015) and adaptive matching window (Pap et al., 2012). In global matching methods, commonly used aggregation methods include the max-flow/min-cut algorithm (Boykov and Kolmogorov, 2004), graph cuts algorithm (Wang and Lim, 2011) and various algorithms based on belief propagation (Sun et al., 2003; Yang et al., 2010; Hu et al., 2012; Barzigar et al., 2013). In addition, the disparity values obtained using local or global matching methods should be further processed by disparity refinement to filter errors and fill gaps (Huq et al., 2013; Yang, 2014; Li et al., 2014).

In photogrammetry-based applications such as digital surface model generation and digital line graph surveying, forward intersection should be used to compute the 3D spatial coordinates (X, Y, Z) of ground points after their homologous pixels have been obtained using some image-space-based matching method; after this step, geographic information products can be generated (Haala and Rothermel, 2012; Zhang et al., 2014; Bertin et al., 2015). The process of manufacturing geographic information products often involves the use of interpolation methods, which *may* decrease their accuracy. Therefore, object-space-based matching methods, which can directly acquire the spatial 3D coordinates of ground objects, have received more attention than image-space-based methods. In object-space-based methods, a ground point with known object space plane coordinates (X, Y) is used as the matching primitive, and the goal is to confirm the elevation Z of the primitive. These object-space-based methods are also referred to as image matching of the ground primitive.

Existing image matching methods of *the* ground primitive are mainly developed based on the traditional vertical line locus (VLL) method; for example, single stereo VLL (Helava, 1988) and multi-view VLL (Zhang and Gruen, 2006; Fan et al., 2007). These methods *use* the following general matching strategy: first, based on the approximate elevation range of the ground primitive, the object space elevation Z is used as the searching benchmark, and the object space 3D coordinates (X, Y, Z_i) are confirmed for the ith searching point in the object space ($i \in \{0, 1, 2, \ldots, n\}$, where n is the search quantity). Each searching Z_i is computed by adding i times of step ΔZ to the minimum elevation. Second, each object space searching point is re-projected to all matching images to obtain the image space searching pixels and compute the similarity measure among searching pixels through an image space matching cost (such as NCC). Finally, the Z_i corresponding to the maximum similarity is selected as the elevation of the ground primitive. However, this matching strategy has the following three problems: (1) It is difficult to precisely set the ΔZ value of the searching step to ensure that the searching point passes the ground primitive. If the value of ΔZ is too large, the correct matching pixel of the ground primitive will be missed. If the value of ΔZ is too small, the computing time will be long. (2) Only one elevation value can be obtained for the ground primitive. When the ground primitive has multiple elevations (for example, if the ground primitive locates at vertical ground objects such as the facade of a building or a telephone pole), existing methods cannot obtain a result that reflects the actual elevation distribution of the ground object. (3) A correct validating procedure for the matching result is lacking. The result corresponding to the maximum matching measure returned by existing methods may not be the actual elevation of the ground primitive.

Therefore, in this study, we introduce an enhanced multi-view vertical line locus (EMVLL) matching algorithm based on positioning consistency for aerial and space images. The algorithm considers multi-view image matching of the ground primitive with known object space plane coordinates and follows the traditional vertical line locus method. The proposed method uses synthetic object space information of the ground primitive and image space information of multi-view images, and fuses the processes of multi-view image matching and correctness validation of the matching result. These improvements aim to address the three defects of traditional VLL matching methods. The proposed method is tested with real aerial and space images, and the experimental results show that the proposed method yields more reliable and effective matching results for ground primitives than traditional methods.

The remainder of the paper is organized as follows. Section 2 briefly introduces the traditional multi-view vertical line locus (TMVLL) matching method. Section 3 details the computing principles of the proposed EMVLL matching algorithm. Section 4 presents detailed experimental data and discusses the results. Finally, conclusions and potential future work are presented in Section 5.

2. Brief introduction to the TMVLL matching method

A sketch of the TMVLL method is shown in Fig. 1. In the figure, G is the ground primitive with known object space plane coordinates (X_g, Y_g). $I_0, I_1, \ldots, I_{m\text{-}1}$ represent multi-view images with known interior and exterior orientation parameters ($m \geq 2$ is the number of images), and their photography centers are $C_0, C_1, \ldots, C_{m\text{-}1}$. The process of computing the elevation Z_g of the primitive G in the TMVLL method is as follows:

(1) According to the rough elevation range of the image area and accuracy requirements of the elevation computation, the approximate minimum and maximum elevations, $Z_{\min}$ and $Z_{\max}$, and the searching step ΔZ of elevation are respectively designated for the primitive G.