The P2L method of mismatch detection for push broom high-resolution satellite images

Article *in* ISPRS Journal of Photogrammetry and Remote Sensing · August 2017

DOI: 10.1016/j.isprsjprs.2017.06.009

CITATION
1

READS
74

2 authors:

Yi Wan
Wuhan University
3 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Yongjun Zhang
Wuhan University
112 PUBLICATIONS **665** CITATIONS
SEE PROFILE

ISPRS Journal of Photogrammetry and Remote Sensing 130 (2017) 317–328

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

The P2L method of mismatch detection for push broom high-resolution satellite images

Yi Wan, Yongjun Zhang *

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 27 January 2017
Received in revised form 22 June 2017
Accepted 22 June 2017

Keywords:
Remote sensing
Satellite image
Image matching
Mismatch detection
RANSAC
Rational polynomial coefficients

ABSTRACT

RANSAC-based mismatch detection methods are widely used in the geometric registration of images. Despite their prevalence, setting the detection thresholds for different situations continues to be difficult without an appropriate geometric model. In high-resolution satellite images, simple image-space transformations are commonly influenced by the terrain or elevation errors. This paper introduces a new method, called the P2L method, which uses the distance between the transformed right image point and the segment of the corresponding epipolar line to distinguish the correct matches and mismatches. The affine model of the P2L method is solved to transform the right image points towards the segment of the epipolar line. The images for demonstration were acquired by GeoEye-1, Ikonos-2, and Ziyuan-3; and each type of image pairs had different intersection angles to explore the influence of the elevation error. The correct matches were manually collected and the mismatches were simulated. The experiments in this paper, which used only correct matches, demonstrated that this method was very robust with one specific threshold (five pixels) and was suitable for all the image pairs. The experiments using simulated mismatches and real matching points demonstrated that this method was able to distinguish most of the mismatches; and even for the image pair that had a 54-degree intersection angle, the ratio of mismatches was reduced from 81% to 11%.

1. Introduction

Satellite images are being used widely in mapping, land-cover observation, and change detection. Satellite image products include not only images but also initial geolocation models (e.g., a rigorous sensor model or a rational polynomial coefficients (RPC) model) (Fraser et al., 2006; Sadeghian et al., 2001; Tao and Hu, 2001). The initial geolocation model parameters are observed by onboard GPS, inertial measuring units (IMUs), and star-trackers. However, even the most modern satellite geo-position equipment produces varying degrees of geo-location errors on the ground (Long et al., 2016; Ozcanli et al., 2014). The satellite images, which are not acquired at the same time and by the same sensors, may contain mosaicking errors in their overlapping areas. The mosaicking errors need to be corrected before the images can be simultaneously used in various advanced tasks such as 3D reconstruction, DEM extraction, image fusion, image mosaic, change detection, and map updating. The geometric correction process consists of three steps: matching the corresponding points, estimating the geometric correction model, and applying the geometric correction model (Zitová and Flusser, 2003).

The first step, matching the corresponding points, has been extensively studied in the past few decades, and many empirical algorithms have been created (Hartmann et al., 2016; Ling et al., 2016). The existing image matching methods can be classified as either area-based matching (ABM) or feature-based matching (FBM) (Long et al., 2016, 2015). ABM methods mainly include normalized cross correlation (NCC) (Helava, 1978; Lhuillier and Quan, 2002; Zhang, 2005), phase correlation (Chen et al., 1994; Zakharov and Toutin, 2012), and least squares matching (Gruen, 1985). Most of the ABM methods require good initial geometric relationships. As far as NCC and its improvements, without a good initial geometric relationship, means that all the pixels of the image will be corresponding point candidates. For phase correlation and least squares matching, having good initial geometric relationships can guarantee the convergence. Compared to the ABM methods, the FBM methods (scale-invariant feature transform (SIFT) (Lowe and Lowe, 2004; Sun et al., 2014) and its improvements (Bay et al., 2008; Ke and Sukthankar, 2004; Mikolajczyk and Schmid, 2013;

* Corresponding author.
E-mail addresses: yi.wan@whu.edu.cn (Y. Wan), zhangyj@whu.edu.cn (Y. Zhang).

http://dx.doi.org/10.1016/j.isprsjprs.2017.06.009
0924-2716/

Morel and Yu, 2009; Ye and Shan, 2014) are more robust to the geometric changes of the local feature, such as scale change, rotation, and even affine transformation, and do not rely on good initial geometric relationships. Thus, in ground or aerial photogrammetry, the FBM methods are widely used for tie-point matching for the purpose of geometric correction; and the ABM methods are widely used in dense matching for the purpose of 3D reconstruction. However, for satellite images, the initial geo-reference is relatively accurate. With the guidance of the initial geo-location, the corresponding point searching area can be very small (hundreds or thousands of square pixels) in order to reduce the computation time for the ABM methods (Xiong et al., 2013). Furthermore, with the aid of a digital elevation model (DEM), the geometric changes of the local feature can be compensated before the image matching. Thus, the ABM methods also perform well for tie-point matching of satellite images (Ling et al., 2016).

The geometric correction of satellite images mainly includes on-orbit geometric calibration (Breton et al., 2002; Mulawa, 2008; Takaku and Tadono, 2009; Zhang et al., 2014), rigorous model (Dowman and Michalis, 2003; Fritsch and Stallmann, 2000; Michalis and Dowman, 2008; Poli, 2007; Toutin, 2004) refinement, and RPC model (Fraser et al., 2006; Grodechi and Dial, 2003; Zhang et al., 2012) refinement. On-orbit camera calibration, which can improve the initial geo-reference accuracy, includes camera-alignment-angle error correction; camera focal length error correction; lens distortion correction; and CCD array rotation, translation, scaling, and bending error corrections. The refinement of a rigorous sensor model or RPC model can improve the accuracy of the exterior orientation. For a rigorous sensor model, the solutions are directly related to the satellite orbits and attitudes. However, for a RPC model, the solutions are the parameters of the empirical models. The image-space affine transformation (Grodechi and Dial, 2003), which has been demonstrated as being capable of absorbing the main part of the exterior orientation error, is widely used in the refinement of RPC models.

When the images are refined separately by orbits (with a rigorous model) or by scenes (with a RPC model), the ground control points (GCPs) can be collected on each orbit or scene, which is not feasible for most users. The geometric correction can also be processed using the strategy of block adjustment. When there are enough tie-points, block adjustment can simultaneously solve the geometric correction using sparsely distributed GCPs (Teo et al., 2010; Toutin, 2006; Zhang et al., 2015). This strategy can significantly improve the relative precision of overlapped images. Least-squares-based block adjustment is subject to poorly located tie-points or mismatches so it is important to ensure that the automatically matched tie-points are reliable; otherwise, mismatches may lead the block adjustment to a locally optimal solution.

However, very little of the past literature provides in-depth discussions about the mismatch detection model and methods which are specifically designed for satellite image matching. Mismatch detection is more important and difficult to achieve than the block adjustment itself. In some situations, the satellite image matching task may be challenging. The images used may have been acquired in different seasons, years, and weather conditions or by various sensors with different optical properties, ground resolutions, and initial geo-reference accuracies. Some images are acquired over water, forest, or desert locations, where most of the textures are poor and are not suitable for collecting key-points or matching them. Some images are acquired over residential areas that have a lot of repeating textures. When doing such matching tasks, it is common to experience a mismatch of 50% or more, even though the initial geo-reference errors are only tens of meters and the size of the searching-windows are only tens of pixels.

To eliminate the mismatches, robust methods should be utilized to estimate the geometric transformation between matching points. Among the most widely used robust methods are M-estimators, LMedS, and RANSAC (Salvi et al., 2001; Torr and Murray, 1997). The M-estimators try to reduce the influence of mismatches by modifying the weights according to the residuals of each point. The LMedS and RANSAC methods are similar, and both are based on a random selection of observations that are then used for estimating the geometric models. The LMedS method calculates the median of the residuals for each model and then chooses the one that minimizes the median. The RANSAC method calculates the number of inliers for each model and then chooses the one that maximizes the number of inliers; and the M-estimators cannot ensure its robustness when mismatches make up over 50% of the observations. LMedS is more restrictive than RANSAC and eliminates more observations than RANSAC. Based on this information, RANSAC was chosen as the method of mismatch detection in this paper. The objective of the research presented in this paper was to design an appropriate function model that can express the geometric relationship between the corresponding points on two satellite images and then calculate the residuals for mismatch judgement. The following two problems are considered:

1. The function model must be sufficiently accurate. A coarse model is not suitable for high-resolution satellite images since the error scale of the mismatches of such images tend to be small because of the initial geo-reference. This function model can avoid the influence of factors such as the terrain.
2. The function model must be able to achieve a solution with a small number of observations.

In Section 2, the RANSAC-based methods using the point-to-point strategy and the point-to-line strategy are introduced. In Section 3, the images acquired by GeoEye-1, Ikonos-2, and ZY-3 are utilized for experiments in which the manually collected points and simulated mismatches were used. Section 4 presents the conclusions of this work.

2. The proposed approach

2.1. Two-view geometry of push broom image

Most of the currently used optical high-resolution satellite images are acquired by push broom cameras. The push broom camera uses a line-array charge-coupled device (CCD) and captures one or several image lines at each exposure. The push broom imaging system also obeys the rule of collinearity, and the corresponding points on two images also obey the rule of coplanarity (Fig. 1(a)). In other words, the two views of satellite images follow the epipolar geometry. When the relative location and attitude are accurately reconstructed, the points of one image will lie on the epipolar lines (or curve) of their corresponding points on the other image (Fig. 1(b) and (c)). As a result, on one image coordinate, the distance between a point and the corresponding epipolar line of its matched point reflects the matching error in the vertical direction of the epipolar line.

Owing to the dynamic nature of push broom image collection, photogrammetric processing of satellite imagery is more complicated than aerial frame camera processing. Aerial cameras acquire the entire image at an instant in time with a unique exposure station and orientation. When using epipolar geometry for aerial image matching, only calculating the fundamental matrix **F** is necessary in order that corresponding points $p^{(l)}$ (the point vectors in this article are all column vectors) and $p^{(r)}$ satisfy $p^{(r)\mathrm{T}}\mathbf{F}p^{(l)} = 0$ (Hartley and Zisserman, 2003). High-resolution push broom satellite cameras acquire a single image line at an instant of time.

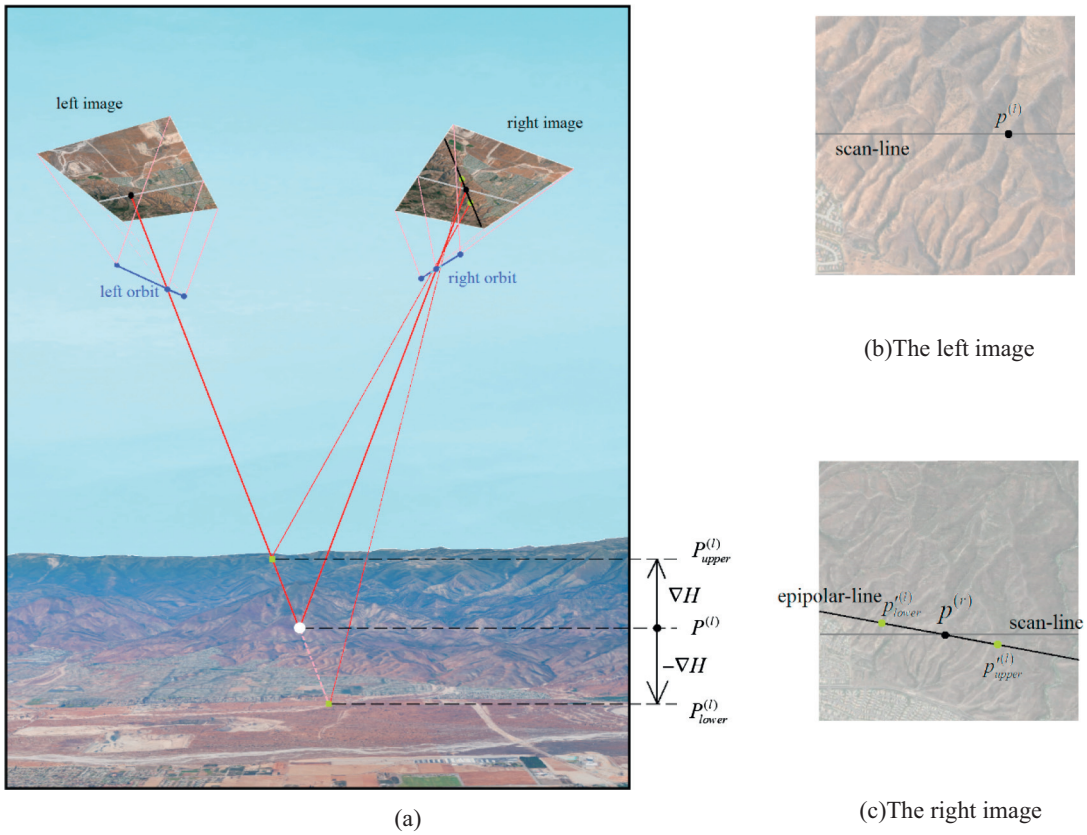

Fig. 1. The epipolar geometry between two overlapped push broom images. The background images are from Google Maps (Google, 2016). (a) The left image point $p^{(l)}$ is back-projected along its ray (the left red line) to a ground point $P^{(l)}$ with an exact elevation or to a ground line-segment $P^{(l)}_{lower}P^{(l)}_{upper}$ with elevation range $[H_0 - \nabla H, H_0 + \nabla H]$. Then, the re-projection of $P^{(l)}$ on the right image is $p'^{(l)}$, and the re-projection of $P^{(l)}_{lower}P^{(l)}_{upper}$ is $p'^{(l)}_{lower}p'^{(l)}_{upper}$. (b) The left image point $p^{(l)}$. (c) The segment of the epipolar-line $p'^{(l)}_{lower}p'^{(l)}_{upper}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Consequently, each line of a push broom image has a different set of exterior orientation parameters. Implementing either the rigorous projection model or the rigorous epipolar geometry model is expensive, time-consuming, and error-prone. Such complicated models require the calculation of a large number of observations and are not suitable for the RANSAC-based methods. Thus, a simplified model to express the satellite image geometry is needed.

2.2. P2P method

In Grodechi and Dial (2003), Grodecki used a polynomial model defined in the domain of image coordinates to represent the adjustable functions for satellite orientation model refinement and demonstrates that a truncated polynomial model, such as an affine model, is sufficient to absorb the orientation error for narrow-field-of-view satellite images.

Grodecki's model is suitable for detecting gross errors in GCPs. On one image, the GCPs ($\{P\}$) onto the image coordinate system is projected as $\{p'\}$ using the initial imaging model (a rigorous model or RPC model); then, by robustly estimating the affine transformation model $\mathbf{A}$ between projections $\{p'\}$ and the observed image points $\{p\}$, the gross errors are detected according to $\text{dist}(p, \mathbf{A}p')$.

This method cannot be directly used for two view images since the GCPs are not available. For the convenience of description, the two view images are the left image and the right image. The point matches are set as:

$$\mathcal{S} = \{(p^{(l)}, p^{(r)})_i \quad | i = 1, 2, \ldots N_{\mathcal{S}}\} \tag{1}$$

where $p^{(l)}$ is the left image point,$p^{(r)}$ is the right image point, $N_{\mathcal{S}}$ is the cardinality of set $\mathcal{S}$. The correct matches are set as $\mathcal{C}_0$, and the mismatches as $\mathcal{M}_0$. Therefore, $\mathcal{S} = \mathcal{C}_0 \cup \mathcal{M}_0$ and $\mathcal{C}_0 \cap \mathcal{M}_0 = \varnothing$.

With the aid of topographic maps or DEMs, the left image points $\{p^{(l)}\}$ are back-projected onto the ground coordinate system using the left initial imaging model. The details of this process can be found in (Sheng, 2005; Zhang et al., 2015). Then, the ground points $\{P^{(l)}\}$ are projected onto the right image coordinate system using its initial imaging model; and the projections are denoted as $\{p'^{(l)}\}$. Finally, the affine transformation model $\mathbf{A}$ is estimated between the projections $\{p'^{(l)}\}$ and the right image points $\{p^{(r)}\}$, and the point-to-point distance $\text{dist}(p'^{(l)}, \mathbf{A}p^{(r)})$ is used for mismatch judgement. This method is called the point-to-point (P2P) method in this paper, and the correspondence set of the P2P method is as follows:

$$\mathcal{S}_{P2P} = \{(p'^{(l)}, p^{(r)})_i \quad | i = 1, 2, \ldots N_{\mathcal{S}}\} \tag{2}$$

In the RANSAC-based estimation and mismatch detection, the j-th set of the random seeds is denoted as $\mathcal{D}_j$, which consists of three corresponding points (i.e., $N_{\mathcal{D}} = 3$). $\mathcal{A}_j = \{\mathbf{A}_{j,k} | k = 1, \ldots N_{\mathcal{A}_j}\}$ is denoted as the set of affine models solved with $\mathcal{D}_j$, and $N_{\mathcal{A}_j}$ is denoted as the number of solutions. In the P2P method, each seeds set outputs one solution, thus $N_{\mathcal{A}} = 1$. The consensus set corresponding to the affine model $\mathbf{A}_{j,k}$ is:

$$\mathcal{C}_{j,k} = \{(p'^{(l)}, p^{(r)})_i \in \mathcal{S}_{P2P} | \text{dist}(p'^{(l)}, \mathbf{A}_{j,k}p^{(r)}) < T_r\} \tag{3}$$

where T_r is the threshold of the point-to-point distance. Then, the coordinates of the consensus sets are used for finding the best estimation of the j -th iteration:

$$\mathbf{A}_j = \arg\max_{\mathbf{A}_{j,k}} N_{\mathcal{C}_{j,k}} \tag{4}$$

The consensus set generated by the best estimation $\mathbf{A}_j$ is denoted as $\mathcal{C}_j$, which is the largest consensus set among those in the j-th iteration, and the corresponding outlier set is denoted as $\mathcal{M}_j = \mathcal{S}_{P2P} \setminus \mathcal{C}_j$. Through enough iterations, the largest $\mathcal{C}_j$ and its corresponding $\mathbf{A}_j$ are the results of the RANSAC-based mismatch detection and robust estimation, as:

$$\begin{aligned} &\mathbf{A} = \arg\max_{\mathbf{A}_j} N_{\mathcal{C}_j} \\ &\mathcal{C} = \{(p'^{(l)}, p^{(r)})_i \in \mathcal{S}_{P2P} | \mathrm{dist}(p'^{(l)}, \mathbf{A}p^{(r)}) < T_r\} \end{aligned} \tag{5}$$

Since the geo-reference error of both the images and the DEM error will affect the back-projection and re-projection process in transforming $p^{(l)}$ s from the left image to the right image as $p'^{(l)}$ s, it remains to be evaluated how the propagation of these errors will affect the robust estimation and the judgement of blunders in Eqs. (4) and (5).

2.3. P2L method

In order to avoid the propagation of the elevation error from affecting the judgment of mismatches, the point-to-line distance based method (P2L method) is proposed. The influence of the elevation error is determined by the intersection angle of the two views of images. Ignoring the elevation error very likely will affect the mismatch detection in images with strong convergence geometry (Teo et al., 2010), (e.g., the stereo image pairs). The elevation error in the back-projection results is not only caused by the DEM error but also by the horizontal error of the left initial imaging model (Zhang et al., 2015). The tolerance of the elevation error in back-projection is denoted as ∇H. When considering the tolerance of the elevation error in the back-projection of the left image point $p^{(l)}$, the result is a space line segment $\overline{P^{(l)}_{lower}P^{(l)}_{upper}}$ (see Fig. 1(a)), which is a segmentation of the left ray. The estimation of the elevation apriori error is discussed in (Zhang et al., 2015), and the tolerance value ∇H is set to three times that of the apriori error. The apriori elevation error is determined by the DEM error, the geolocation error of the left imaging model, and the average slope of the covered area.

Then, the projection of $P^{(l)}_{lower}$ and $P^{(l)}_{upper}$ on the right image is denoted as $p'^{(l)}_{lower}$ and $p'^{(l)}_{upper}$. The projection of line segment $\overline{P^{(l)}_{lower}P^{(l)}_{upper}}$ on the right image is a part of the epipolar line. For push broom images, the epipolar line is not a rigorously straight line, but rather a curve. However, owing to the stable satellite platform, the projection of $\overline{P^{(l)}_{lower}P^{(l)}_{upper}}$ can be approximated as a straight line segment, which is denoted as $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$. So the correspondence set becomes:

$$\mathcal{S}_{P2L} = \left\{ (\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, p^{(r)})_i \quad | i = 1, 2, \ldots N_{\mathcal{S}} \right\} \tag{6}$$

The distance between the point $\mathbf{A}p^{(r)}$ and the line segment $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ is used for judging the outliers. This method, which uses $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$ to judge outliers, is called the point-to-line (P2L) method. The calculation of the distance between a point and a line segment is shown in Fig. 2. There are three reasons for using such a distance for judgement:

1. When the elevation error is limited, the length of the epipolar curve will be limited so that approximating it as a line-segment will not cause large errors.
2. The elevation errors of most of the back-projection will be within the tolerance ∇H. In such condition s (see Fig. 2(a)), the elevation error will not remain in $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$ and will not affect the mismatch detection.
3. Significant elevation errors will be reflected in $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$ and such points will be eliminated (see Fig. 2(b)). This is reasonable because mistakes in elevation will also cause problems in DEM-aided block adjustment.

2.4. Affine model in P2L method

Two-dimensional affine models need at least three pairs of points to solve. However, $\mathcal{S}_{P2L}$ only has corresponding points and line-segments. The correct projection location $p'^{(l)}$ (not affected by the elevation error) is somewhere on the line-segment $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$, but the accurate location is unknown. In order to estimate a more accurate affine model, the equally divided points on $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ are used (see Fig. 3) as candidates. The set of equally divided points on $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ are denoted as:

$$\mathcal{T} = \left\{ p'^{(l)}_k = p'^{(l)}_{lower} + \frac{k}{K+1}\left(p'^{(l)}_{upper} - p'^{(l)}_{lower}\right) | k = 1 \ldots K \right\} \tag{7}$$

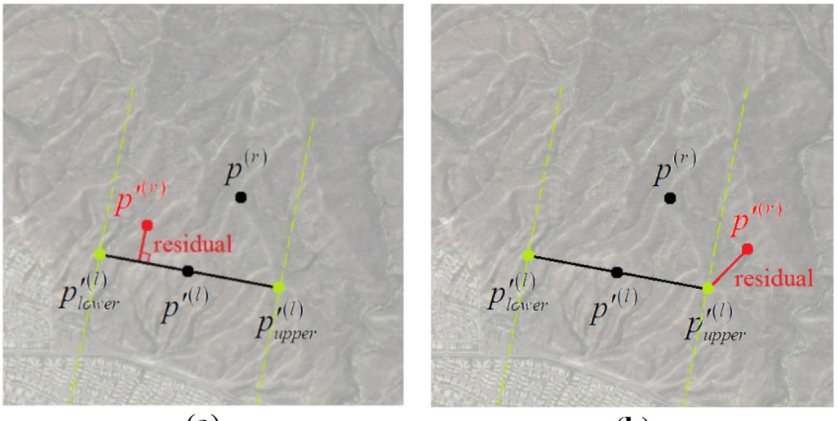

Fig. 2. The computation of the point to line-segment distance. $p'^{(r)}$ is $\mathbf{A}p^{(r)}$. The length of the red line-segment is $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$. (a) The $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$ is the distance between the point to the straight line if the elevation error is within the tolerance. (b) The $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$ is the distance between the point to the nearest endpoints of $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ if the elevation error is greater than the tolerance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

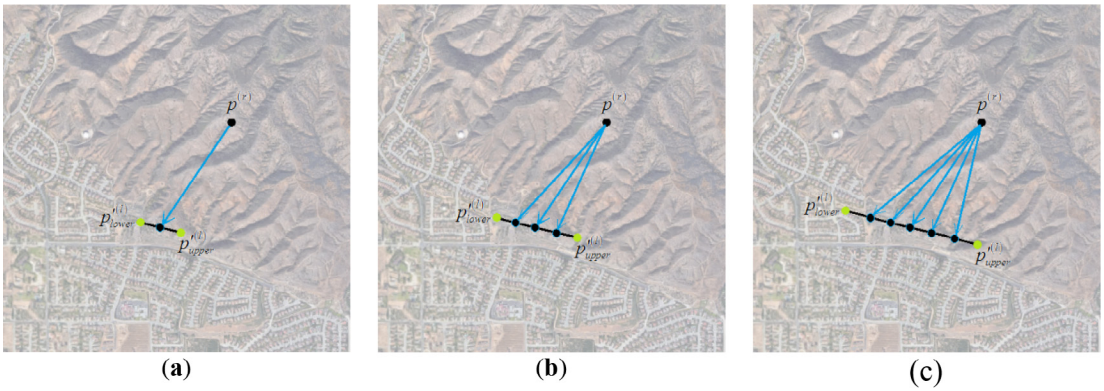

Fig. 3. Solving the affine model that transforms $p^{(r)}$ towards the equally divided points on $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$. With longer epipolar line segments, using more equally divided points as candidates can provide a more accurate solution but will cost more time in finding the best solution: (a) using only the middle point; (b) using three equally divided points; and (c) using five equally divided points.

Each of the seed set $\mathcal{D}_j$ includes three correspondences of the point and line-segment. The sets of equally divided points of the line-segment are denoted as $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$, so the three correspondences will have $N_{\mathcal{T}_1}$, $N_{\mathcal{T}_2}$ and $N_{\mathcal{T}_3}$ point pairs. Brute force computation produces $N_{\mathcal{T}_1} \cdot N_{\mathcal{T}_2} \cdot N_{\mathcal{T}_3}$ solutions. The workflow of each RANSAC iteration of the P2L method is given in Algorithm 1.

Algorithm 1 (*The j-th RANSAC iteration of the P2L method*).

Input: correspondence set $\mathcal{S}_{P2L}$, threshold T_r
Output: the j-th consensus set $\mathcal{C}_j$, the j-th affine model $\mathbf{A}_j$
1. A sample set $\mathcal{D}_j \subset \mathcal{S}_{P2L}$ is randomly selected.
2. Equally divided points on $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ s are generated so that $\mathcal{D}_j = \{(\mathcal{T}_1, p_1^{(r)}), (\mathcal{T}_2, p_2^{(r)}), (\mathcal{T}_3, p_3^{(r)})\}$
3. $\mathcal{C}_j \leftarrow \varnothing$
4. For each $p_1'^{(l)} \in \mathcal{T}_1$:
5. For each $p_2'^{(l)} \in \mathcal{T}_2$:
6. For each $p_3'^{(l)} \in \mathcal{T}_3$:
7. Affine model $\mathbf{A}$ is computed using $\left\{\left(p_1'^{(l)}, p_1^{(r)}\right), \left(p_2'^{(l)}, p_2^{(r)}\right), \left(p_3'^{(l)}, p_3^{(r)}\right)\}$
8. $\mathcal{C} = \{(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, p^{(r)}) \in \mathcal{S}_{P2L} | \mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)}) < T_r\}$ is generated
9. if $N_{\mathcal{C}} > N_{\mathcal{C}_j}$: $\mathcal{C}_j \leftarrow \mathcal{C}$, $\mathbf{A}_j \leftarrow \mathbf{A}$
10. end

The length of $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ is determined by ∇H and the intersection angle. As shown in Fig. 3, the more equally divided points that are used, the less the distance will be between the true position $p'^{(l)}$ and its nearest equally divided point in order that the affine model selected by Algorithm 1 will be more accurate. In the most extreme condition, the candidates can be collected on the epipolar line segments pixel by pixel; thus, a very accurate affine model can be achieved if the seeds are all correct matches, which will cause at most one-pixel-error on the points within the area of the triangle formed by the three seeds. However, using such much equally divided points may lead to too large time-costing when the epipolar line segments are long.

In our implementation, considering both the accuracy and computational efficiency, the number of equally divided points is limited up to seven (see Table 1). The reason is that for most high-resolution satellite images, the field-of-view is very small, which leads to a phenomenon that the epipolar lines on the overlapping area are almost parallel (Morgan et al., 2006; Wang et al., 2011). As a result, the directions of the transformation errors of the affine model solved by the P2L method are almost the same with the epipolar-line direction. Such errors, if limited in scale, will barely affect the P2L distances of the most of correct matches.

Table 1
The number of equally divided points according to the length of $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$.

Length of $\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$ in pixels	0–5	5–20	20–60	60+
Number of candidates, K_i	1	3	5	7

Although, theoretically, the propagation of the elevation error will not influence the mismatch detection according to $\mathrm{dist}(\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$, it remains to evaluate how accurate the relative orientation error (propagation of the geo-reference error of both the images) will be modeled by the affine model estimated by Algorithm 1.

3. Simulated experiments

Seventeen panchromatic high-resolution satellite image pairs (see Table 2) were used to demonstrate and compare the performance and the effect of the P2P and the P2L method. The image pairs were placed in three groups according to their sensor types. Group I included three 0.5 m resolution panchromatic images, which were acquired by the GeoEye-1 satellite over Las Vegas, Nevada USA. Las Vegas is situated on the desert and is surrounded by mountain ranges on all sides. The elevation in the image covered area varies from 500 m to 800 m. Group II included nine panchromatic 1.0 m resolution images, which were acquired by the Ikonos-2 satellite over Shanxi Province, China. Group III included 2.1 m panchromatic resolution images, which were acquired by the Ziyuan-3 satellite over Shanxi Province, China. The topography in Shanxi Province, China is mainly mountains and covers different environments: residential, semirural, forested areas, and many coal mines. The elevation varies from 500 to 3000 m. The product level of the GeoEye-1 images and the Ikonos-2 images is GEO. The product level of the Ziyuan-3 images is level-1B. Both the level-GEO of the digital globe products and the level-1B of the Ziyuan-3 products mean that only the distortion is rectified and the images are not ortho-rectified.

The image pairs in each group had different intersection angles in order to compare the influence of the elevation error in the P2P

Table 2
The detailed information of the image-pairs involved in the experiments.

Group	Pair index	IA (°)[a]	Image center (longitude/latitude, degree)				Width × height (thousand pixel)		Acquisition date/time[b] (mm/dd/yy hh:mm)	
			Image-1		Image-2		Image-1	Image-2	Image-1	Image-2
I GeoEye-1 (0.5 m pan)	1	15	−115.1	36.0	−115.1	36.0	18 × 34	16 × 39	08/15/09 18:40	06/21/09 18:36
	2	26	−115.2	36.0	−115.2	36.0	18 × 34	18 × 34	07/10/09 18:27	08/15/09 18:40
	3	47	−115.0	36.3	−115.0	36.3	18 × 35	13 × 37	09.06/09 18:40	04/27/11 18:24
II Ikonos-2 (1.0 m pan)	1	4	111.4	40.4	111.4	39.8	12 × 91	12 × 58	02/28/02 03:37	11/13/01 03:37
	2	10	112.1	39.4	112.1	39.2	12 × 48	12 × 14	05/13/02 03:35	09/30/02 03:38
	3	16	111.7	38.4	111.6	38.4	12 × 18	12 × 17	09/12/10 03:21	09/12/10 03:21
	4	22	112.9	36.9	112.7	36.9	12 × 50	14 × 54	09/01/08 03:29	09/01/09 03:28
	5	27	112.6	39.2	112.5	39.1	13 × 16	14 × 14	03/11/02 03:39	01/23/08 03:42
	6	32	111.4	34.2	111.4	34.2	12 × 13	12 × 13	09/24/07 03:35	09/24/07 03:36
	7	38	112.7	38.0	112.7	38.2	12 × 28	14 × 34	02/23/02 03:27	10/05/05 03:44
	8	46	110.6	33.9	110.7	33.9	13 × 23	13 × 23	04/06/08 03:39	04/06/08 03:38
	9	54	111.2	34.0	111.1	34.0	13 × 13	13 × 12	04/03/10 03:22	02/21/10 03:27
III Ziyuan-3 (2.1 m pan)	1	3	113.7	39.0	114.0	39.0	25 × 24	25 × 24	07/13/13 03:25	05/20/13 03:26
	2	6	112.8	37.4	113.2	37.4	25 × 24	25 × 24	07/03/13 03:29	05/10/13 03:26
	3	7	111.8	37.8	112.2	37.8	25 × 24	25 × 24	08/16/13 03:33	04/30/13 03:29
	4	11	111.9	38.2	111.8	37.8	25 × 24	25 × 24	08/31/13 03:22	08/16/13 03:33
	5	23	110.8	38.3	111.1	38.2	25 × 24	25 × 24	07/07/13 03:46	08/06/13 03:36

[a] IA means intersection angle, which is estimated by averaging the intersection angles of all corresponding rays.
[b] The acquisition time is listed in the UTC time zone.

method and the P2L method. The influence of the elevation error is determined by the intersection angle. Consequently, if a method is saliently affected by the elevation error, its effect becomes worse when the intersection angle is increased.

About 30–50 corresponding points were manually collected and carefully checked on each image pair to ensure their correctness (see Fig. 4). The accuracy of the matching point was believed to be better than one pixel in the images that had less than a

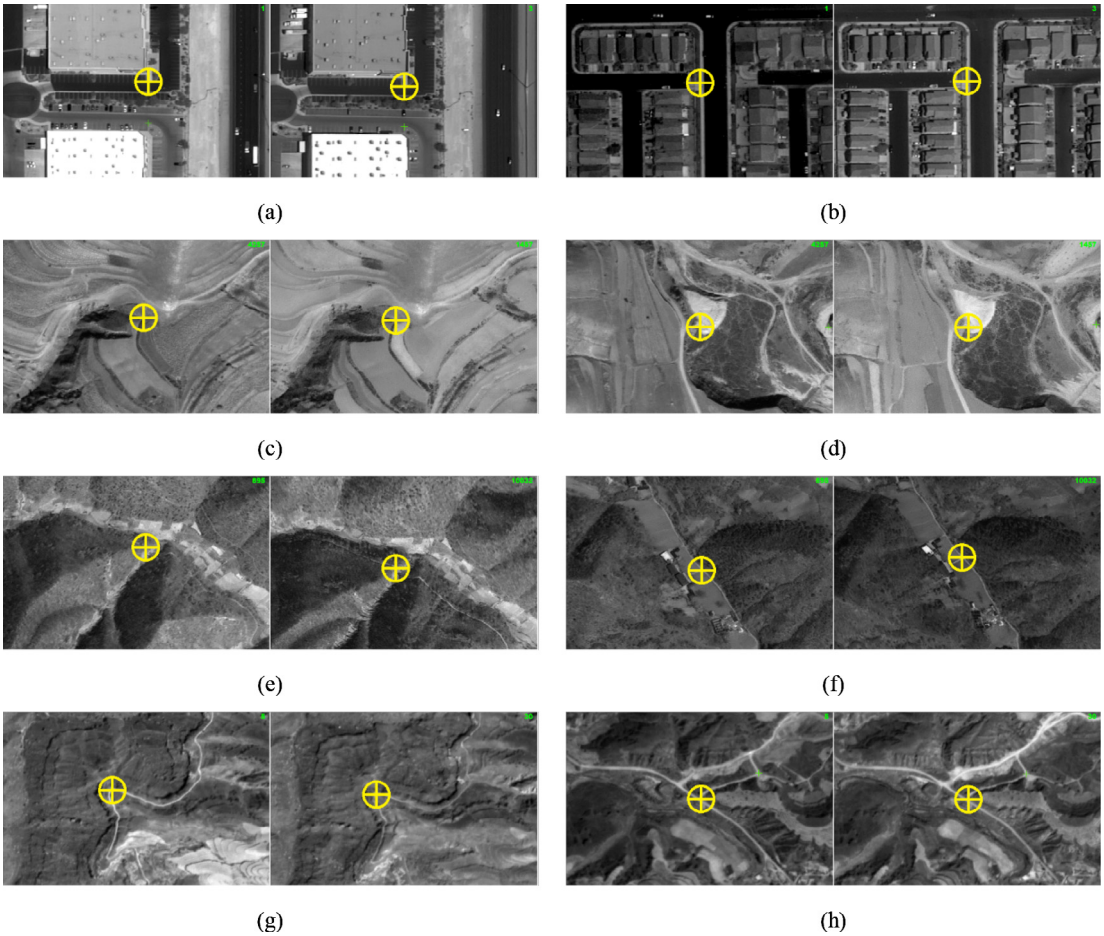

Fig. 4. The manually collected corresponding points on the image pairs. (a) and (b) Corresponding points on image pair I-1 (GeoEye images). (c) and (d) Corresponding points on image pair II-2 (Ikonos images). (e) and (f) Corresponding points on image pair II-8 (Ikonos images). (g) and (h) Corresponding points on image pair III-5 (ZY-3 images).

30-degree intersection angle. Whereas, in image pairs having larger-than-30-degree intersection angles, where the landscape differed significantly, the accuracy was believed to be better than two pixels. The mismatches were randomly simulated. Each mismatch was simulated in three steps:

1. One location, $p^{(l)}$, was randomly selected on the left image.
2. $p^{(l)}$ was back-projected onto the DEM to determine the ground location $P^{(l)}$, and then $P^{(l)}$ was projected to the right image to determine $p'^{(l)}$.
3. Within the right image window, which is 200 × 200 square pixels, and the center is $p'^{(l)}$, one location was randomly selected as $p^{(r)}$.

The SRTM-DEM with 30 m ground resolution was used for elevation interpolation in the back-projection. When using the P2L method, the elevation tolerance was set as $\nabla H = 30m$ so that the elevation range was $[H_0 - \nabla H, H_0 + \nabla H]$ where H_0 was the interpolated elevation on the SRTM. Since the SRTM-DEM is not a standard DEM product, abnormally large error can occur. As a result, not all of the correct manually collected matching points can be used for solving affine models. In this paper, about 10% of the correct matches had significant elevation errors in the experiments.

3.1. The geometric consistency

It has been previously demonstrated that the image-space affine model can absorb most of the exterior orientation error of push broom satellite images (Grodechi and Dial, 2003). The purpose of this experiment was to evaluate the accuracy of using the image-space affine transformation to model the relative orientation error in both the P2P and the P2L methods. If it was accurate, the response value ($\mathrm{dist}(p, \mathbf{A}p')$ or $\mathrm{dist}(\overline{p'^{(l)}_{lower} p'^{(l)}_{upper}}, \mathbf{A}p^{(r)})$) of the correct match would be small if there was no other error propagation (e.g., elevation error). The influence of elevation error would also be evaluated for the P2P and the P2L methods.

$T_j^{(p)}$ was defined as the minimum threshold that can put no less than $100p\%$ of the correct matches ($\mathcal{C}_0$) into the j-th consensus set $\mathcal{C}_j$:

$$T_j^{(p)} = \min\{T_r | N_{\mathcal{C}_j \cap \mathcal{C}_0} \geqslant p \cdot N_{\mathcal{C}_0},\ 0 < p \leqslant 1\} \tag{8}$$

When the j-th sample set is made up of correct matches (i.e., $\mathcal{D}_j \subset \mathcal{C}_0$), $T_j^{(p)}$ can be used to evaluate the accuracy of the geometric model. A large $T_j^{(p)}|_{\mathcal{D}_j \subset \mathcal{C}_0}$ means that the geometric consistency is poor, and distinguishing between the correct matches and the mismatches will be difficult. Thus, by examining the $T_j^{(p)}|_{\mathcal{D}_j \subset \mathcal{C}_0}$ for each image pair, it was possible to determine whether the affine model in the P2P method or the P2L method could compute accurate and stable response values for each matching points pair.

For the P2P method, only one affine model was solved, and one $T_j^{(p)}$ was computed with a particular seed set. However, for the P2L method, more than one solution was possible. With each solution, one $T_j^{(p)}$ was computed, and the minimum solution was selected. To avoid the influence of the abnormally large elevation error, $T_j^{(0.9)}$, rather than $T_j^{(1)}$, was used for evaluation.

$\min_{j=1}^{K}(k)\{\cdot\}$ was denoted as the k-th minimum item among the K items. Among the K number of $T_j^{(0.9)}$ s, the k-th was the smallest:

$$T_{(k,K)}^{(0.9)} = \min_{j=1}^{K}(k)\left\{T_j^{(0.9)}\right\} \tag{9}$$

When K was large enough to consider all the possible seed sets sampled from the inliers $\mathcal{C}_0$, $T_{(1,K)}^{(0.9)}$ was the lower bound of $T_j^{(0.9)}$. When $T_r < T_{(1,K)}^{(0.9)}$, it was impossible to maintain 90% correct matches no matter how accurate the affine model was. When $T_r \geqslant T_{(0.5 \cdot K,K)}^{(0.9)}$, there was a larger than 0.5 probability to maintain at least 90% of the correct matches. In this experiment, the seeds were randomly sampled 1000 times on each image pair (i.e., $K = 1000$). Not all the solutions solved by correct matches were accurate, because some of the correct matches suffered from poor elevations. Consequently, $T_{(1,1000)}^{(0.9)}$, $T_{(750,1000)}^{(0.9)}$ and $T_{(500,1000)}^{(0.9)}$ were illustrated in Fig. 5 and were used for evaluation.

Fig. 5 shows the scale of the threshold $T^{(0.9)}$ s and their relationship with the intersection angles. As shown in Fig. 5(a), when using the P2L method with Group I, the $T_{(1,1000)}^{(0.9)}$ s were within one pixel, the $T_{(500,1000)}^{(0.9)}$ s were within two pixels, and the $T_{(750,1000)}^{(0.9)}$ s were within three pixels. When using the P2P method with Group I, the $T^{(0.9)}$ s were much larger. Especially for the image pair I-3, which had a 46-degree intersection angle, the $T_{(1,1000)}^{(0.9)}$ was about six pixels, and the $T_{(500,1000)}^{(0.9)}$ was about 12 pixels. As shown in Fig. 5(b), when using the P2L method with Group II, almost all the $T^{(0.9)}$ s were within five pixels. Only the $T_{(750,1000)}^{(0.9)}$ of image pair II-9 was about eight pixels. When using the P2P method with Group II, the $T^{(0.9)}$ s were much larger and showed an increasing trend with larger intersection angles. As shown in Fig. 5(c), when

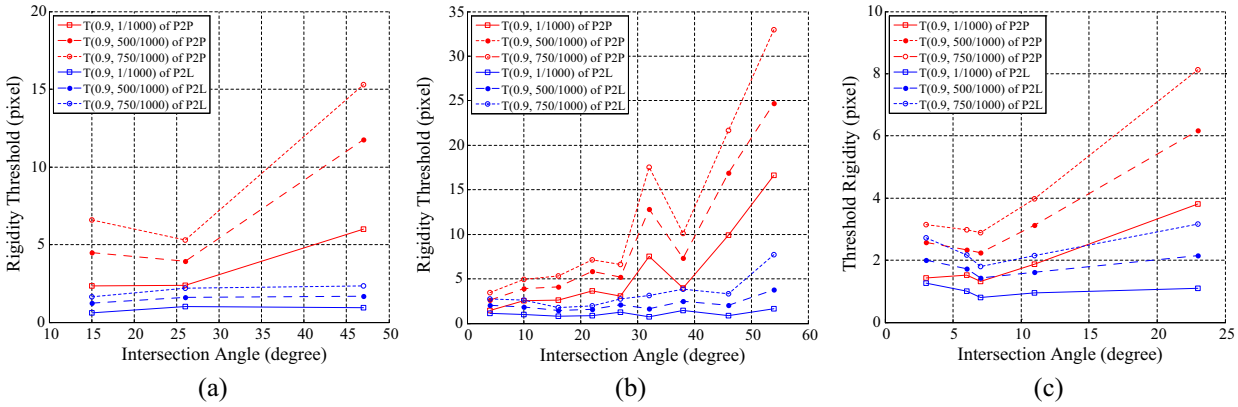

Fig. 5. The minimum threshold that can maintain 90% correct matches. T(0.9, 1/1000) means the minimum threshold among 1000 tests, T(0.9, 500/1000) means the 500th smallest one, and T(0.9, 750/1000) means the 750th smallest one. (a) Group I; (b) Group II; (c) Group III.

using the P2L method with Group III, all the $T^{(0.9)}$ s were within four pixels. When using the P2P method with Group III, the $T^{(0.9)}$ s also showed an increasing trend with larger intersection angles. When the image-pairs had large intersection angles, the elevation error seriously influenced the results of the P2P method because the correct matches also had larger-than-ten-pixel residuals, which can confuse the correct matches and the mismatches. However, the elevation error barely influenced the results of the P2L method. Furthermore, one threshold (i.e., $T_r = 5$) was used for all the image pairs in the P2L method, which means that selecting several thresholds is unnecessary. From this aspect, the P2L method is convenient for automatic systems, while the P2P method is not.

The $T^{(0.9)}$ s were less than five pixels, when computed with the P2P method on the image pair II-1 through II-3 and III-1 through III-4 which have less-than-15-degree intersection angles or when computed with the P2L method on all the image pairs. This result demonstrates that, if there is no influence by the elevation error, the affine transformation can accurately model the relative orientation error between the two images.

3.2. Distinguish the mismatches

The experiment results in Section 3.1 demonstrate that the P2L method needs much smaller thresholds than the P2P method. However, this does not prove that the P2L method can better distinguish the mismatches. When using the P2P method, a point correspondence was put into the consensus set $\mathcal{C}_j$ if the transformed right image points $\mathbf{A}_j p^{(r)}$ fell within the T_r-dilation region of $p'^{(l)}$ (see Fig. 6(a)) which is called the inlier-region. The inlier-region of the P2L method is T_r-dilation of each line segment $\overline{p'^{(l)}_{lower} p'^{(l)}_{upper}}$ (see Fig. 6(b)). The area of the inlier-region reflects its ability to distinguish mismatches. A larger area indicates that more mismatches will be mistakenly put into the consensus set. As shown in Fig. 6, when the same T_r was used, the inlier-region of the P2L method was larger than the P2P method. Thus, more mismatches were remained by the P2L method.

As previously mentioned and shown in Fig. 5, the P2L method needs much smaller thresholds to maintain 90% correct matches. When evaluating the ability to distinguish correct matches and mismatches, $T^{(0.9)}$ was set as the threshold, and then the areas of their inlier-regions were compared. $A(T_r)$ was denoted as the area of inlier region computed with the threshold T_r. Fig. 7 illustrates the inlier region area $A(T^{(0.9)}_{(750,1000)})$. To examine the influence of elevation error, the horizontal coordinates were also the intersection angles. For the image pairs that had small intersection angles (e.g., less than 20 degrees), the areas of the inlier region computed in the P2P method and in the P2L method did not show significant differences. With the increasing intersection angle, the areas of inlier region of both the P2P method and the P2L method increased. For the P2P method, this was caused by increasing the minimum threshold (see Fig. 5). For the P2L method, it was caused by increasing the intersection angle and the length of the line-segment $\overline{p'^{(l)}_{lower} p'^{(l)}_{upper}}$ (see Fig. 1). For the image pairs that had large intersection angles (e.g., greater than 20 degrees), the area of the inlier region of the P2L method tended to be less than that of the P2P method.

The area of the inlier region reflects the theoretical ability to recognize mismatches. The shape of the inlier area reflects the distribution of the unrecognized mismatches. The mismatches maintained by the P2P method were evenly distributed within a circle. However, the mismatches maintained by the P2L method were evenly distributed within the T_r -dilation of the line segment

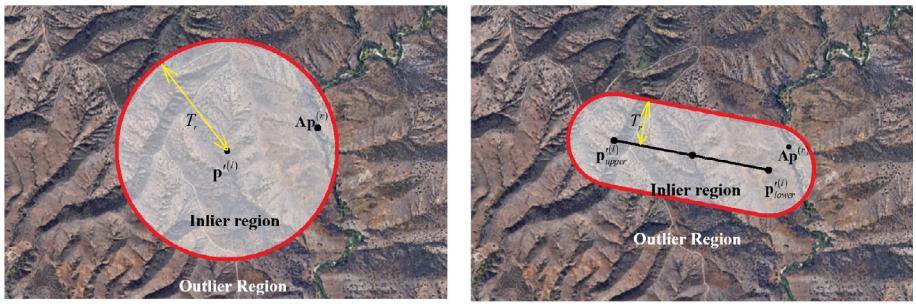

Fig. 6. The inlier region in RANSAC-based method. (a) When using the P2P method, if the transformed position $\mathbf{A}_j p^{(r)}$ falls within the T_r -dilation of $p'^{(l)}$, this point-pair will be judged as inliers. (b) When using the P2L method, if the transformed position $\mathbf{A}_j p^{(r)}$ falls within the T_r -dilation of $\overline{p'^{(l)}_{lower} p'^{(l)}_{upper}}$, this point-pair will be judged as inliers.

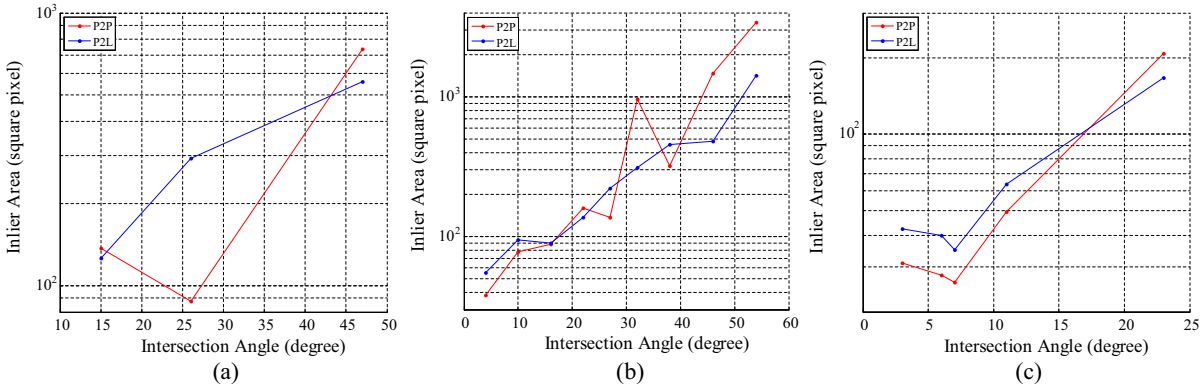

Fig. 7. The area of the inlier region when using $T_r = T^{(0.9)}_{(750,1000)}$. (a) The result of Group I. (b) The result of Group II. (c) The result of Group III.

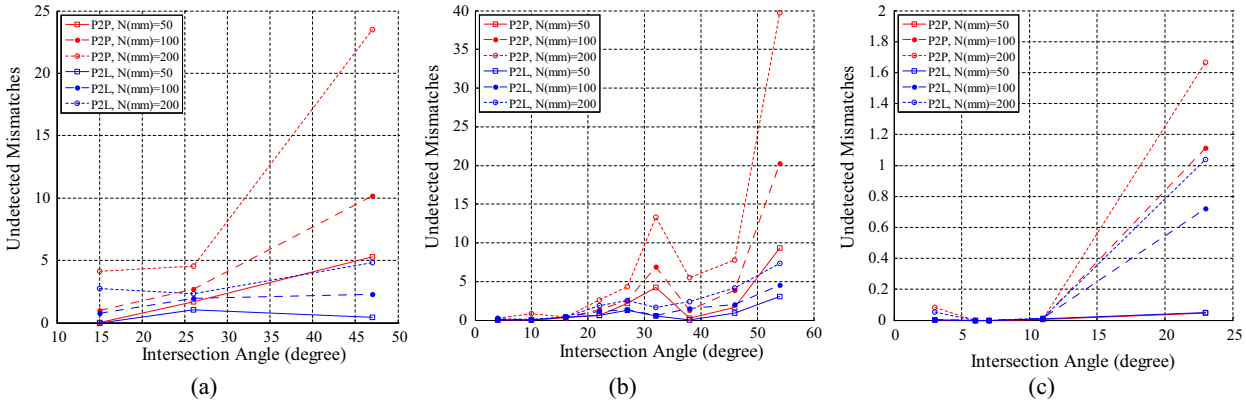

Fig. 8. The average number of undetected mismatches. N(mm) means the number of randomly generated mismatches. (a) Group I results. (b) Group II results. (c) Group III results.

$\overline{p'^{(l)}_{lower}p'^{(l)}_{upper}}$. Such matches cannot be eliminated since the exact elevation is unknown.

The area of the inlier-region can only reflect the theoretical ability to distinguish the mismatches. Virtual mismatches were simulated for each image pair, assuming the P2P method and the P2L method with $T_r = T^{(0.9)}_{(750,1000)}$, and the number of false positive (i.e., the unrecognized mismatches) were counted. For each image pair, the set of mismatches $\mathcal{M}_0$ was simulated three times (50, 100, and 200 mismatches). For each $\mathcal{M}_0$, the seed set $\mathcal{D}_j$ (from correct matches $\mathcal{C}_0$), the solution $\mathbf{A}_j$, and the number of false positive $N_{\mathcal{C}_j \cap \mathcal{M}_0}$ were computed 1000 times. Then the average number $E(N_{\mathcal{C}_j \cap \mathcal{M}_0})$ was computed, as shown in Fig. 8.

When using the P2L method, $E(N_{\mathcal{C}_j \cap \mathcal{M}_0})$ s were less than 5 in Group I, less than 10 in Group II, and less than 1.2 in Group III. Such results satisfied the requirements of the block adjustment. However, when using the P2P method, in image pair I-3, $E(N_{\mathcal{C}_j \cap \mathcal{M}_0})$ was nearly 25 when $N_{\mathcal{M}_0} = 200$, which was about 60% of $N_{\mathcal{C}_0}$; in image pair II-9, $E(N_{\mathcal{C}_j \cap \mathcal{M}_0})$ was nearly 40 when $N_{\mathcal{M}_0} = 200$, which was about the same with $N_{\mathcal{C}_0}$. These results did not meet the needs of the block adjustment. The results shown in Fig. 8(b) and (c) indicate that the elevation influence was much larger on high-resolution images. On lower resolution images, the mismatch detection result was much better.

3.3. Block adjustment tests

In order to evaluate the precision of the two methods, block adjustments using GCPs and independent check points (ICPs) were processed on image pair III-5 using the block adjustment procedure introduced in (Zhang et al., 2015). The block adjustment accuracy was measured by the horizontal mean squared error of the ICPs. The distribution of the GCPs and the ICPs are shown in Fig. 9. There were 27 GCPs on the left im In order to evaluate the precision of the two methods, block adjustments using GCPs and independent check points (ICPs) were processed on image pair III-5 using the block adjustment procedure introduced in (Zhang et al., 2015). The block adjustment accuracy was measured by

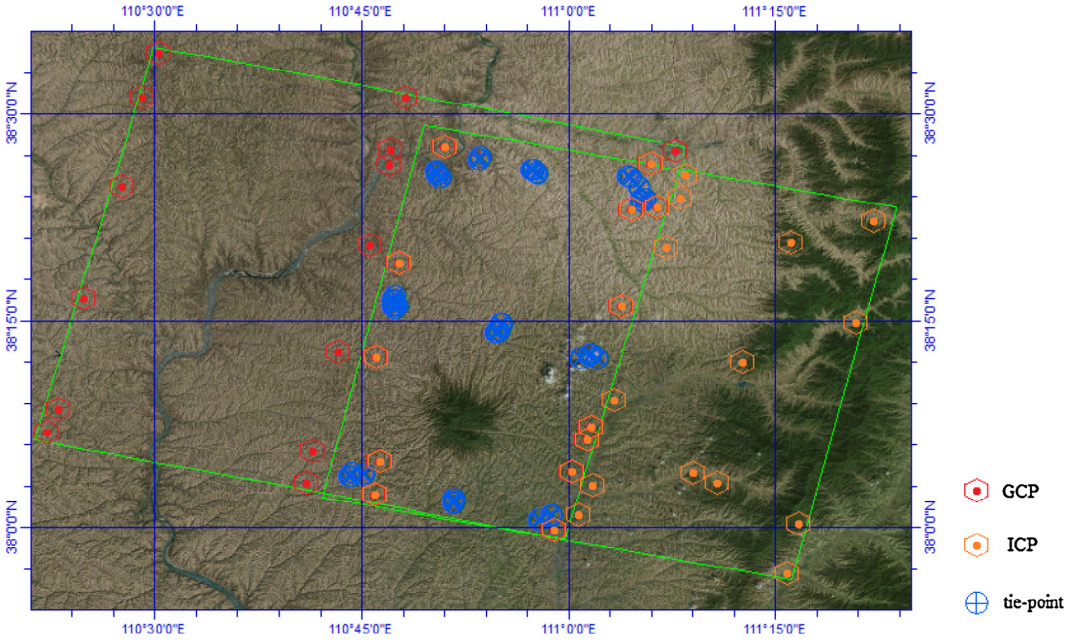

Fig. 9. The distribution of the GCPs, ICPs and tie-points on image pair III-5. On the overlapping area, each GCP corresponds to two image points. However, the ground points with their left image correspondences were treated as GCP records, and the ground points with their right image correspondences were treated as ICP records. The tie-points shown in the figure are only the manually-collect ones while the simulated mismatches are not shown.

the horizontal mean squared error of the ICPs. age and 27 ICPs on the right image. The elevations of the GCPs and ICPs were interpolated from the 30 m resolution SRTM-DEM. The tie-points included the 34 manually collected corresponding points and randomly simulated mismatches. The mismatch detection was processed before the block adjustment. For each number of the simulated mismatches (0, 100, and 200), the P2L method was processed with $T_r = 2$, while the P2P method was proceeded with $T_r = 2, 5,$ *and* 10. For each group of conditions (number of mismatches, detection methods, and the threshold T_r), the experiment was processed for 100 times to avoid the influence of random factors. As a result, the median numbers of the both false positive and the false negative among the results of the 100 mismatch detections were calculated, and the mean value of the horizontal RMSEs of ICPs of the 100 block adjustments was calculated.

The experiment results are listed in Table 3. When using the P2L method with $T_r = 2$, the mean false negative numbers (MFNs) were all less than 2. When using the P2P method with $T_r = 2$, nearly half of the manually collected tie-points were falsely deleted (the Med-FNs were 18, 15, and 15 respectively to 0, 100, and 200 mismatches). Only when using the P2P method with $T_r = 10$, were the Med-FNs less than 2. Meanwhile, the median false positive number (Med-FP) with 200 mismatches was 8. These results also illustrate that the P2L method can better distinguish the mismatches.

Before the block adjustment, the horizontal RMSE (XY-RMSE in Table 3) of the ICPs was 18.18 m. With the P2L method, the mean XY-RMSEs after the block adjustments were 1.40 m with no mismatches, 2.01 m with 100 mismatches, and 2.92 m with 200 mismatches. With the P2P method, the smallest XY-RMSE after the block adjustment was 1.61 m with no mismatches, 1.94 m with 100 mismatches, and 2.68 m with 200 mismatches, which had a weak advantage (compare the bolded data in Table 3). However, the smallest XY-RMSEs of the P2P method corresponded to different thresholds. Without or with 100 simulated mismatches, the most accurate results were achieved with $T_r = 10$ and with 200 simulated mismatches $T_r = 5$; and with an inappropriate T_r, the P2P method would exhibit much worse block adjustment accuracy than the P2L method.

This experiment showed that the P2P method, if well configured, can achieve similar or even better block adjustment accuracy compared to the non-configured P2L method. However, the configuration work is difficult and strongly depends on the experience of the processor, which is not suitable for the automatic data processing systems.

Table 3
The result of the mismatch detections and block adjustments on image pair III-5.

Conditions			Mismatch detection		Mean RMSE of ICPs (meter)		
$N_{\mathcal{M}_0}$	Method	T_r (pixel)	Med-FN	Med-FP	X	Y	XY
0	P2L	2	2	0	**1.10**	**0.87**	**1.40**
	P2P	2	18	0	3.84	1.08	3.99
		5	4	0	2.62	0.99	2.80
		10	0	0	**1.23**	**1.04**	**1.61**
100	P2L	2	0	1	**1.68**	**1.11**	**2.01**
	P2P	2	15	0	2.56	1.11	2.79
		5	3	1	2.35	**1.05**	2.57
		10	0	1	**1.50**	1.23	**1.94**
200	P2L	2	0	3	**2.61**	**1.32**	**2.92**
	P2P	2	15	0	2.96	1.15	3.18
		5	3	2	**2.47**	**1.05**	**2.68**
		10	0	8	2.98	2.24	3.73
Before block adjustment					6.98	16.79	18.18

Note: 1. Med-FN is the median number of false negative (i.e., the number of deleted correct matches).
2. Med-FP is the median number of false positive (i.e., the number of undetected mismatches).
3. X is the direction of longitude. Y is the direction of latitude. XY means error in horizontal.

4. Experiment with real matching points

In Section 3, it was demonstrated that the P2L method with a fixed threshold can work well in almost all conditions and can better distinguish the correct matches for image pairs having large intersection angles. While the P2P method does not have this property but is more computationally efficient than the P2L method, the P2L method may solve more than one affine model with each group of seeds (see Section 2.4). In this section, the real matched Harris key points were used to evaluate the performance and efficiency of the two methods. The matching was based on the NCCs within an 11×11 matching window. The matching point ($p^{(r)}$) was searched within the 100×100 searching window centered at the predicted location of the left point on the right image. When the number of matched points exceeded 1000, only 1000 points were randomly selected and involved in the experiment.

Table 4
The experiment results using automatic matched points.

Image pair	Matched points	Number of inliers (Min & Max)				Median iteration times		Median time costs (ms)	
		P2P		P2L		P2P	P2L	P2P	P2L
I-1	1000	671	795	850	872	8	5	0.6	79.7
I-2	1000	697	810	922	927	8	5	0.7	79.7
I-3	1000	456	508	903	906	37	5	2.9	88.5
II-1	1000	967	971	960	963	5	5	0.5	7.0
II-2	1000	885	921	899	905	5	5	0.5	29.5
II-3	1000	905	961	961	963	5	5	0.5	28.4
II-4	1000	838	934	982	984	5	5	0.5	78.2
II-5	1000	582	682	715	733	14	9	1.2	141.9
II-6	1000	337	360	801	926	108	5	5.9	82.6
II-7	1000	538	590	573	617	22	17	1.4	281.1
II-8	918	296	325	739	775	110	5	7.8	95.1
II-9	981	64	69	292	298	67188	398	3398.4	6687.9
III-1	1000	938	949	909	916	5	5	0.6	6.5
III-2	1000	986	990	955	958	5	5	0.7	6.7
III-3	1000	915	927	846	855	5	5	0.6	6.7
III-4	1000	939	947	905	910	5	5	0.5	6.6
III-5	1000	848	922	888	894	5	5	0.6	29.1

To imitate the automatic processing condition, the thresholds of both methods were fixed: 10 pixels for the P2P method and three pixels for the P2L method. For each image pair, the experiments were processed 100 times to avoid the results being influenced by random factors. The RANSAC process terminated when the number of iteration reached $N_{\max}$:

$$N_{\max} = \begin{cases} [\ln\alpha/\ln(1-\lambda^3)], & if\ [\ln\alpha/\ln(1-\lambda^3)] > 5 \\ 5, & if\ [\ln\alpha/\ln(1-\lambda^3)] \leqslant 5 \end{cases} \quad (10)$$

where $\lambda = \max_j\{N_{c_j}\}/N_S$; α is the tolerance of failure, which was set as 1% in our implementation. The experiment was conducted on a laptop computer, of which the CPU is Intel® Core™ i5-3210 M version with 2.50 GHz frequency. The implementation did not utilize parallel computing techniques. Table 4 lists the experimental results.

When dealing with about 1000 matches, the P2P method cost about 0.1 ms per iteration while the time-cost-per-iteration of the P2L method varied from 1 ms to 18 ms. From this aspect, it can be concluded that the P2P method was much more computationally efficient than the P2L method. However, with the image pairs I-1through I-3 and II-5 through II-9, the P2P method obtained obviously fewer inliers than the P2L method, which led to many more RANSAC iterations in order to satisfy the tolerance of failure. On II-9, which had the largest intersection angle, the P2P method obtained only 64 ~ 69 inliers and cost about 67,188 iterations on average, while the P2L method obtained 292 ~ 298 inliers and cost only about 398 iterations on average. The features on the image pairs having large intersection angles are usually faced with geometric deformation, which can lead to mismatches. When there are more than 50% mismatches, finding more correct matches can greatly reduce the needed number of RANSAC iterations. From this aspect, the P2L method was more computationally efficient when dealing with image-pairs having large intersection angles.

5. Conclusion

This paper introduced a practical RANSAC-based mismatch detection method, called the P2L method, for satellite image matching based on two key elements: (1) an easy-to-solve function model and (2) accurate response computation. Two conclusions were made from the experimental results in Section 3.1: (1) the image-space affine model of this new method that only needs three samples to provide a solution is capable of absorbing the relative orientation error; and (2) the distance between the point and the epipolar line segment reflects the real geometric rigidities of the matches instead of the mixture of the matching error and the propagation of the elevation error. As a result, one fixed threshold (e.g., five pixels) can be used in almost all situations with the P2L method. The experiments in Section 3.2 demonstrated the good mismatch detection of the P2L method in that it successfully detected over 95% of the mismatches (simulated within a 200×200 searching window around the predicted position) using samples selected from the correct matches and the minimum threshold at which 90% correct matches could be maintained. The experiment results in Section 3.3 demonstrated that good and stable block adjustment accuracy could be reached with the tie-points selected by the P2L method with a fixed threshold. The real data experiments in Section 4 demonstrated that the P2L method was effective in discriminating the correct matches and mismatches and reduced the RANSAC iterations by finding larger consensus sets.

In Section 3.2, it is shown that the P2L method could reduce the mismatches from about 80% to less than 20%. Thus, the block adjustment would not converge to an incorrect solution. When dealing with multi-view image matching points, the mismatch detection can be processed in two steps: (1) detect mismatches in each image pair using the P2L method and (2) sort and combine the result of the P2L method, then the block adjustment can be made on the images and multi-view matching points. The mismatches when the error direction was close to the epipolar line could not be detected by the P2L method because it utilizes the epipolar geometry. Such mismatches likely can be detected in the least-square-based block adjustment if there are more correspondences in other images; but if sufficient correspondences do not exist or the rays of the correspondences fall on the same space plane, it will be almost impossible to detect them using geometric constraint.

The P2L method will play an important role in processing large-scale satellite images automatically since it can obtain good results without being configured according to the properties of the data. However, there are still works to do in designing a better optimization strategy to solve the affine model in the P2L method with seed points, since the current RANSAC strategy is neither efficient nor sufficiently accurate. In addition, the problem remains of how to judge the correctness of the mismatch detection results when the consensus sets are too small. For example, if a consensus set of 50 is selected from 1000 matches and no larger consensus set is found, it may be made up of correct matches, or mismatches which happen to have good geometric consistency. Before checking them manually, both conditions happening is a possibility. How to avoid such manual work and make the automatic judgments accurately and efficiently is planned for future work.

Acknowledgements

The authors are graceful to Environmental Systems Research Institute for testing the approach with the IKONOS-2 and GeoEye-1 images. This work was supported by the National Science Fund for Excellent Young Scholars of China (Grant No. 41322010) and the National Natural Science Foundation of China (General Program, Grant No. 41571434).

References

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110, 346–359.
Breton, E., Bouillon, A., Gachet, R., Delussy, F., 2002. Pre-Flight and in-Flight Geometric Calibration of SPOT 5 HRG and HRS Images.
Chen, Q., Defrise, M., Deconinck, F., 1994. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. IEEE Trans. Pattern Anal. Mach. Intell. 16, 1156–1168.
Dowman, I.J., Michalis, P., 2003. Generic rigorous model for along track stereo satellite sensors. In: ISPRS Workshop on High Resolution Mapping from Space 2003, Hanover.
Fraser, C., Dial, G., Grodecki, J., 2006. Sensor orientation via RPCs. ISPRS J. Photogram. Remote Sens. 60, 182–194.
Fritsch, D., Stallmann, D., 2000. Rigorous Photogrammetric Processing of High Resolution Satellite Imagery. Universität Stuttgart, Fakultät Bauingenieur-und Vermessungswesen, Institut für Photogrammetrie.
Google, 2016. Google Maps <www.google.com/maps>.
Grodechi, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogramm. Eng. Rem. S 69, 59–68.
Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. South African J. Photogram. Remote Sens. Cartogr. 14, 175–187.
Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.
Hartmann, W., Havlena, M., Schindler, K., 2016. Recent developments in large-scale tie-point matching. Isprs J. Photogram. Remote Sens. 115, 47–62.
Helava, U.V., 1978. Digital correlation in photogrammetric instruments. Photogrammetria 34, 19–41.
Ke, Y., Sukthankar, R., 2004. PCA-SIFT: a more distinctive representation for local image descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 506–513.
Lhuillier, M., Quan, L., 2002. Match propagation for image-based modeling and rendering. IEEE Trans. Pattern Anal. Mach. Intell. 24, 1140–1146.
Ling, X., Zhang, Y., Xiong, J., Huang, X., Chen, Z., 2016. An Image Matching Algorithm Integrating Global SRTM and Image Segmentation for Multi-Source Satellite Imagery. 8.

Long, T., Jiao, W., He, G., Zhang, Z., 2016. A fast and reliable matching method for automated georeferencing of remotely-sensed imagery. Remote Sens-Basel 8.
Long, T., Jiao, W., He, G., Zhang, Z., Cheng, B., Wang, W., 2015. A generic framework for image rectification using multiple types of feature. ISPRS J. Photogram. Remote Sens. 102, 161–171.
Lowe, D.G., Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110.
Michalis, P., Dowman, I., 2008. A generic model for along-track stereo sensors using rigorous orbit mechanics. Photogram. Eng. Remote Sens. 74, 303–309.
Mikolajczyk, K., Schmid, C., 2013. A performance evaluation of local descriptors. IEEE Trans. on, p. 257.
Morel, J.M., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. Siam J. Imag. Sci. 2, 438–469.
Morgan, M., Kim, K.-O., Jeong, S., Habib, A., 2006. Epipolar resampling of space-borne linear array scanner scenes using parallel projection. Photogram. Eng. Remote Sens. 72, 1255–1263.
Mulawa, D., 2008. On-Orbit Geometric Calibration of the OrbView-3 High Resolution Imaging Satellite, 35.
Ozcanli, O.C., Dong, Y., Mundy, J.L., Webb, H., 2014. Automatic geo-location correction of satellite imagery. Computer Vision Pattern Recogn. Workshops, 1–15.
Poli, D., 2007. A rigorous model for spaceborne linear array sensors. Photogram. Eng. Remote Sens. 73, 187–196.
Sadeghian, S., Zoej, M.J.V., Delavar, M.R., Abootalebi, A., 2001. Precision rectification of high resolution satellite imagery without ephemeris data. Int. J. Appl. Earth Obs. Geoinf. 3, 366–371.
Salvi, J., Armangue, X., Pages, J., 2001. A survey addressing the fundamental matrix estimation problem. In: International Conference on Image Processing, 2001. Proceedings, pp. 17–22.
Sheng, Y., 2005. Theoretical analysis of the iterative photogrammetric method to determining ground coordinates from photo coordinates and a DEM. Photogram. Eng. Remote Sens. 71, 863–871.
Sun, Y., Zhao, L., Huang, S., Yan, L., Dissanayake, G., 2014. L2-SIFT: SIFT feature extraction and matching for large images in large-scale aerial photogrammetry. ISPRS J. Photogram. Remote Sens. 91, 1–16.
Takaku, J., Tadono, T., 2009. PRISM on-orbit geometric calibration and DSM performance. IEEE Trans. Geosci. Remote Sens. 47, 4060–4073.
Tao, C.V., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogramm. Eng. Rem. S 67, 1347–1358.
Teo, T.-A., Chen, L.-C., Liu, C.-L., Tung, Y.-C., Wu, W.-Y., 2010. DEM-aided block adjustment for satellite images with weak convergence geometry. Geosci. Remote Sens., IEEE Trans. 48, 1907–1918.
Torr, P.H.S., Murray, D.W., 1997. The development and comparison of robust methods for estimating the fundamental matrix. Int. J. Comput. Vision 24, 271–300.
Toutin, T., 2004. Review article: geometric processing of remote sensing images: models, algorithms and methods. Int. J. Remote Sens. 25, 1893–1924.
Toutin, T., 2006. Spatiotriangulation with multisensor HR stereo-images. Geosci. Remote Sens., IEEE Trans. 44, 456–462.
Wang, M., Hu, F., Li, J., 2011. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. ISPRS J. Photogram. Remote Sens. 66, 347–355.
Xiong, J., Zhang, Y., Zheng, M., Yuanxin, Y.E., 2013. An SRTM assisted image matching algorithm for long-strip satellite imagery. J. Remote Sens.
Ye, Y., Shan, J., 2014. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. ISPRS J. Photogram. Remote Sens. 90, 83–95.
Zakharov, I., Toutin, T., 2012. Subpixel image matching based on Fourier phase correlation for Radarsat-2 stereo-radargrammetry. Can. J. Remote. Sens. 38, 487–495.
Zhang, L., 2005. Automatic Digital Surface Model (DSM) Generation From Linear Array Images. Institute of Geodesy & Photogrammetry.
Zhang, Y., Lu, Y., Wang, L., Huang, X., 2012. A new approach on optimization of the rational function model of high-resolution satellite imagery. Geosci. Remote Sens., IEEE Trans. 50, 2758–2764.
Zhang, Y., Wan, Y., Huang, X., Ling, X., 2015. DEM-assisted RFM block adjustment of pushbroom nadir viewing HRS imagery. IEEE Trans. Geosci. Remote Sens. 54, 1–10.
Zhang, Y., Zheng, M., Xiong, J., Lu, Y., 2014. On-orbit geometric calibration of ZY-3 three-line array imagery with multistrip data sets. IEEE Trans. Geosci. Remote Sens. 52, 224–234.
Zitová, B., Flusser, J., 2003. Image registration methods: a survey. Image Vis. Comput. 21, 977–1000.