ISPRS Journal of Photogrammetry and Remote Sensing 153 (2019) 123–136

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

An a-contrario method of mismatch detection for two-view pushbroom satellite images

Yi Wan, Yongjun Zhang*, Xinyi Liu

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

ARTICLE INFO

Keywords:
Image matching
Mismatch detection
Remote sensing
Satellite image
a-contrario method

ABSTRACT

Mismatch detection is a key step in the geometric correction of satellite images. However, most RANSAC-based mismatch detection methods face two problems in practical application, i.e., how to preset the threshold when the apriori matching accuracy is not known and how to validate the correctness of the results when the proportions of true matches are very low. In this paper, we propose an *a-contrario* method named ORSA-SAT to remove the mismatches for two-view satellite images by finding the most meaningful set of matches. The formula first is defined to compute the geometric rigidity of a set of point matches according to the image match search area with the matching accuracy measured by the maximum point-to-epipolar-line distance. Then, the meaningfulness of a set is rated by a probabilistic criterion that estimates the number of false alarms (NFA), which indicates the expected times that a set can be found by chance from non-rigid and randomly distributed matched points. The criterion is a function of the quantity of point-matches and the geometric rigidity and is used in ORSA-SAT for comparing two sets. The true matches are collected by finding the most meaningful set; thus, no preset thresholds are needed to separate the true matches and the mismatches. Furthermore, the criterion also justifies the correctness of the sets obtained by ORSA-SAT since rigid sets rarely occur from mismatches. In this paper, we use both simulated data and real matched points on images captured by IKONOS-2, ZY-3, and Landsat-8 to demonstrate ORSA-SAT. The results of the simulated experiments show that both the precisions and the recalls were ensured above 80% in the correct results of ORSA-SAT even though there were over 90% mismatches originally.

1. Introduction

Since the launch of the Ikonos satellite in 1999, enormous technological progress has been made in high-resolution remote sensing image acquisition. However, despite the rapid development of direct positioning and orientation technology, the mosaicking errors of satellite images still remain even as the ground resolution continues to improve with the progress in sensor manufacturing and data storage. Past efforts to improve geo-referencing accuracy after image acquisition have studied and tested many methods with a rigorous sensor model (RSM) (Dowman and Michalis, 2003; Fritsch and Stallmann, 2000; Toutin, 2003) or a rational polynomial camera (RPC) model (Fraser et al., 2006; Grodechi and Dial, 2003; Tao and Hu, 2001; Teo et al., 2010; Toutin, 2004a, 2004b; Xiong and Zhang, 2009; Zhang et al., 2016; Zheng and Zhang, 2016). Most of these methods need reliable corresponding points to recover the geometric relationship of the satellite images and to improve their geo-referencing accuracy. In most cases, the automatically matched point correspondences include mismatches which will do harm to the geometric correction (Ling et al., 2016). In this paper, we introduce a new method that filters the mismatches as outliers and supplies a criterion to justify the correctness of the results, i.e., the sets of inliers.

Automatic image matching methods have been extensively studied in the past few decades, most of which can be classified as either feature-based methods or area-based methods. Feature-based methods have three main steps: feature extraction, feature description, and feature matching. The features that can be used for image matching include points, lines, line-segments, and junctions. Scale invariant scale transform (SIFT), introduced in (Lowe, 2004), is one of the most widely used feature matching methods. A large number of variants have been based on or inspired by SIFT, e.g., speeded up robust features (SURF) (Bay et al., 2008), PCA-SIFT (Ke and Sukthankar, 2004), and ASIFT (Morel and Yu, 2009). When matching satellite images, the differences in the optical properties of the sensors and the acquisition dates can result in significant spectral changes for the same locations. Thus, many variants have been developed based on SIFT to overcome the non-linear

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

https://doi.org/10.1016/j.isprsjprs.2019.04.020
Received 5 August 2018; Received in revised form 25 April 2019; Accepted 29 April 2019

spectral transform, e.g., SAR-SIFT (Dellinger et al., 2015), adaptive binning SIFT (AB-SIFT) (Sedaghat and Ebadi, 2015), support-line based method (Li et al., 2017), and the uniform competency (UC) local feature extraction method (Sedaghat and Mohammadi, 2018b). Area-based methods, on the other hand, use the intensity information to measure the similarities between two image windows; and by sliding the matching window, the pair of windows having the highest similarity scores will be considered as corresponding locations (Kasser and Egels, 2002). There are also area-based methods that use mutual information (Chen et al., 2003) or frequency domain information (Chen et al., 1994; Foroosh et al., 2002). The main drawback of the area-based methods is that they are sensitive to the changes caused by scale variants, rotations, or viewpoint changes. However, when matching satellite images, these geometric changes can be recovered by ortho-rectification of the images with their initial geo-referencing models and a digital elevation model (DEM). Thus, the area-based methods are widely used in satellite image matching since most of them are more efficient than the feature-based methods. Some studies, targeted on matching different types of geographic data like DEMs, SAR images and vector data, combines the feature matching strategy and the area or the phase information (Fan et al., 2018; Sedaghat and Naeini, 2018; Ye et al., 2018). In the experiments presented in this paper, the satellite images were matched with a normalized correlation coefficient (NCC) matcher. When only the highest NCC score was accepted, we obtained one-on-one point-matches. When all the NCC scores over a certain threshold were accepted, we obtained one-on-multiple point-matches. Both the one-on-one match sets and the sets of one-on-multiple matches were considered and analyzed by the proposed method.

The results of image matching inevitably will include mismatches. The matching of two satellite images acquired over some special landforms can be rather challenging and the proportion of mismatches may be very high. Those mismatches should be filtered to avoid disturbing the geometric correction. The current mismatch filtering strategies mainly fall in one of two categories: nonparametric methods and parametric methods. The nonparametric methods, which consider photometric constraints or geometric constraints or both, can be used in both rigid image matching and non-rigid image matching (Ma et al., 2015a, 2014; Torresani et al., 2008; Zhao et al., 2013; Zhou et al., 2016). Local geometric constraints are often used in nonparametric methods (Sedaghat and Mohammadi, 2018a). The parametric methods are based on a hypothesize-and-verify framework. First, it is hypothesized that the point-matches obey a parametric geometric relationship. Then, solving the parametric model is attempted with the matches. Finally, the optimal subset is found, which is the highest number of possible matches that accurately obey the parametric geometric model. Empirically, if the proportion of mismatches is less than 20% and the scales of the gross errors are limited, the mismatches can be eliminated by solving the parametric model with the least-square method (LSM) and a robust loss function. However, when the proportion of errors exceeds 20%, more robust methods must be used, such as RANSAC (Fischler and Bolles, 1981) and its variants (Chum and Matas, 2005; Ma et al., 2015b; Raguram et al., 2013; Torr and Zisserman, 2000).

In most cases, the correspondences of the satellite image points are searched within small search regions and thus the geo-location errors of even mismatches are only pixels or tens of pixels (Xiong et al., 2013). As a result, the nonparametric methods, which generally do not use strict geometric constraints, are not suitable for satellite images. An accurate and simple parametric model is needed for the parametric methods to accurately filter out such mismatches. The RANSAC-based methods usually need a simple parametric model; otherwise, the computation may be rather inefficient. For satellite images, the RSM and RFM models are too complicated. The most commonly used method is to use a digital elevation model (DEM) to map the points from one image onto another image-space in two steps: back-projection and reprojection. Then, with the corresponding points on the same image-space, a 2D affine transformation, which has only six parameters, can be used to model their geometric relationship. This model was originally used for RFM refinement with GCPs (Grodecki and Dial, 2003). The elevation errors in the process of back-projection caused by the error in the DEM and the initial geo-reference is propagated in the process of reprojection (Zhang et al., 2016). For high-resolution image-pairs that also have a large convergence angle, such error propagation should not be ignored since it is almost the same scale as the geo-location errors of the mismatches. (Wan and Zhang, 2017) introduced a P2L method that uses an epipolar-line-segment (ELS) constrained model to resolve this issue. Their model utilized the epipolar geometry of satellite images while considering the elevation uncertainty in the back-projection. Thus, the result of the process of back-projection and reprojection is an ELS rather than a point. After applying the affine transformation, the distances between the ELSs and the corresponding points are used as residual errors. This model has been demonstrated to be effective in the RANSAC method to filter out mismatches and avoid the influence of elevation error. However, this method still uses a fixed threshold of geometric accuracy to judge inliers and outliers. This threshold should be preset according to the apriori error of the geometric model. The apriori error is difficult to estimate especially when non-linear distortion exists.

In this paper, we offer a probabilistic definition for the rigidity of a set of point matches and propose an *a-contrario* method to remove mismatches by finding the most meaningful subset. The paradigm of *a-contrario* mismatch removal was initially proposed by (Moisan and Stival, 2004) for detecting rigid point matches between two frame images. The authors designed *a probabilistic criterion to detect the existence of a rigid motion between two sets of point matches that permits to decide whether these points are independent or if they are correlated by a rigid motion*. In the case of satellite images, a point-match can be judged correct if it belongs to a large subset of which all the point matches can be accurately correlated by an image-space affine transformation because the existence of such subsets is a large deviation from randomness. In the context of *a-contrario* theory and methods, the meaningfulness of a certain image structure is measured by estimating its NFAs. In (Moisan and Stival, 2004), the meaningfulness of a set of point-matches is defined as *the expected number of similar sets (that is, having the same size and at least the same rigidity) in a random uniform distribution of points*. The estimation of meaningfulness combines the number of points, rigidities, etc., in which the accuracy threshold is not needed. Such expectation-based (rather than probability-based) formulation for image structure detection was originally developed by Desolneux et al. (2000, 2007), whereby a structure that cannot happen by chance under the null-hypothesis will be judged meaningful. This theory is now widely used in detecting low-level image structures, such as alignments (Desolneux et al., 2000), line-segments (von Gioi et al., 2010), vanishing points (Almansa et al., 2003), histogram gaps (Delon et al., 2004, 2005), contrasted boundaries (Cao et al., 2005; Desolneux et al., 2001), rigid sets of point matches (Moisan et al., 2012; Moisan and Stival, 2004), junctions (Xia et al., 2014; Xue et al., 2018), etc.

There are two main contributions to the study presented in this paper. First, a new parametric method named ORSA-SAT to eliminate mismatches for satellite images was developed that does not require presetting the geometric threshold. Second, a criterion was established to justify the correctness of the results of ORSA-SAT. These two contributions are expected to decrease the amount of work needed and improve the automation level of the large-area geometric correction process of satellite images.

The remainder of this paper is organized as follows. Section 2 briefly introduces the geometric model of the P2L method. Section 3 defines rigidity and meaningfulness as they pertain to a set of point matches and introduces the *a-contrario* method of ORSA-SAT. Section 4 evaluates ORSA-SAT with simulated experiments. Section 5 evaluates the ORSA-SAT method with real matched points from a few very challenging matching tasks. Section 6 presents the conclusions of this paper.

2. Epipolar line-segment constraint

2.1. The P2L distance

The fundamental geometric relationship of photogrammetry is the collinear relationship of the object point, the perspective center, and the corresponding image location of the object point. The law of collinearity yields the epipolar geometry, i.e., the coplanarity of the image points corresponding to one object point, the two perspective centers, and the object point. For frame images, the epipolar geometry can be parametrically modeled in a very simple manner: $(x, y, 1)\mathbf{F}(x', y', 1)^{\mathrm{T}} = 0$, in which F is the 3×3 fundamental matrix, and $(x,y,1)$ and $(x',y',1)$ are two corresponding points. Epipolar geometry is useful in the parametric mismatch detection methods for frame images. However, for pushbroom satellite images, the epipolar geometry cannot be modeled parametrically in such a simple manner because of the dynamic perspective center. Wan and Zhang (2017) used the ELS rather than the epipolar line to constrain the relationship of the corresponding points with the aid of a DEM, which greatly simplified the parametric geometric model.

Denote $\mathcal{I}_l$ as the region of the left satellite images where the key-points and $\mathcal{I}_r$ as the region of right satellite images where the key-points are matched. Denote the matching point set as:

$$\mathcal{S} := \{(p_l, p_r)_i | i = 1, 2, \ldots n, p_l \in \mathcal{I}_l, p_r \in \mathcal{I}_r\} \tag{1}$$

where p_l s are the key-points; p_r s are the matches of key-points; and n is the number of matches. The p_l s can be back-projected onto the ground as P_l s through ray-tracing (Sheng, 2005, 2008) with a DEM and the geo-referencing model of $\mathcal{I}_l$. When considering the elevation error ∇H in a ray-tracing procedure, a segment is found on the left ray centered at P_l, which is denoted as L_l (see Fig. 1). Projecting a left ray onto the right view, an epipolar-curve is obtained, and by projecting the segment L_l s onto the right view, the epipolar curve-segments are obtained. Since the curve-segments are quite short and will cross only tens or hundreds of scanning-lines, there will not be a salient loss in accuracy when treating them as straight ELSs, denoted as l_l s. Thus, the point correspondences are transformed as point-to-line-segment correspondences, denoted as:

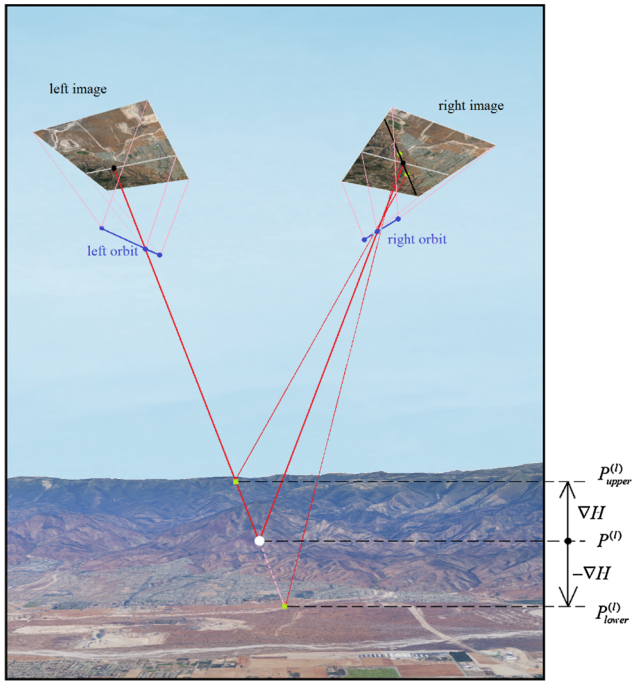

Fig. 1. The epipolar geometry of pushbroom satellite images with a limited elevation uncertainty (Wan and Zhang, 2017).

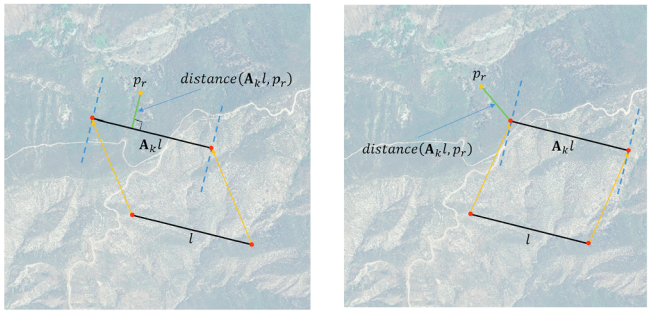

Fig. 2. The definition of the P2L distance.

$$\mathcal{S}_{P2L} := \{(l_l, p_r)_i | i = 1, 2, \ldots n, l_l \in \mathcal{I}_r, p_r \in \mathcal{I}_r\} \tag{2}$$

Note that when matching a digital ortho-map (DOM) with a satellite image to get GCPs, the DOM should be used as the left image and L_l is on the vertical direction.

Then, with an affine model that transforms l_l s towards p_r s, the geometric error of the i-th point-match are measured by the point-to-line-segment (P2L) distance, denoted as $d_{\mathbf{A}}(i) = distance(\mathbf{A}\cdot(l_l)_i, (p_r)_j)$. $\mathbf{A}\cdot l_l$ is a line-segment transformed from l_l and by the affine model $\mathbf{A}$. Using the P2L distance to measure the matching accuracy can effectively avoid the influence of the limited uncertainty in elevation. The drawbacks of this ELS constraint is that some mismatches may occur that happen to have small P2L distances. The definition of point-to-line-segment distance is illustrated in Fig. 2.

Then, with a threshold δ, the inliers can be collected by comparing the P2L distance and δ:

$$C(\mathbf{A}) = \{(p_l, p_r)_i \in \mathcal{S} | d_{\mathbf{A}}(i) \leqslant \delta\} \tag{3}$$

2.2. Estimating the affine transformation

An affine transformation includes six independent parameters, which is the solution of a 6×6 linear system or a $2n \times 6$ linear least-square minimization problem. At least three point-correspondences are required to recover the affine transformation. Since the limited uncertainty of the elevation is considered in the outlier detection of satellite image matches, only the correspondences between points and ELSs are considered. Thus, the candidates on the ELSs must be selected to make up the point correspondences and to solve the affine transformation.

Wan and Zhang (2017) used the equally-divided-points (EDPs) on the ELSs to make up the point-correspondences, and the number of EDPs was set according to the length of the ELS (see Table 1). As a result, in the RANSAC-based robust solvers, one seed set that includes three point-to-ELS correspondences may construct more than one 3-point-pair sets and their associated affine transformations. Then, the consensus set $C(\mathbf{A})$ of each affine transformation is constructed with the threshold δ. Finally, the optimal affine transformation $\overline{\mathbf{A}}$ is corresponded with the largest consensus set. If there are more than one largest consensus sets having equal element number, $\overline{\mathbf{A}}$ is corresponded with the one having the least value of the maximum P2L distance of inliers.

The RANSAC-based methods still suffer from the guessing element of setting an appropriate accuracy threshold δ because too many factors

Table 1
The number of EDPs according to the length of ELSs (Wan and Zhang, 2017).

Length of ELS (pixel)	0–5	5–20	20–60	> 60
Number of EDPs	1	3	5	7

are involved. Furthermore, it is not possible to judge whether the results are correct when the proportions of inliers are quite small, say less than 20%. Inspired by Moisan and Stival (2004), Section 3 proposes a criterion that combines the set number and the P2L distance to measure the meaningfulness of a subset and introduces a random sampling-based algorithm to detect rigid sub-sets from either one-on-one point matches or one-on-multiple point matches.

3. Rigid point-set detection

3.1. Measuring rigidity

In this section, the geometric rigidity of a set of two-view point-matches is measured from the probabilistic point of view. Before that step, however, it is necessary to know how to compute the probability of the occurrence of a point-match under the null-hypothesis. The null-hypothesis, denoted as $\mathcal{H}_0$, is that there is no meaningful structure in an image, e.g., a white noise image. Thus, the points matched on such images are meaningless. As a result, under $\mathcal{H}_0$, the matched locations on $\mathcal{I}_r$ of the key-points are randomly and uniformly distributed within their search regions.

The search region is a sub-region of $\mathcal{I}_r$ where p_r s are determined for key-point p_l s. When matching natural images, in most cases, the matching points are searched within the whole image region because information is lacking about the relative camera pose and the depth of the field. However, when matching satellite images with the aid of an initial geo-reference and a DEM, the search region can be limited to a much smaller range. Although the search region can be in any shape while matching is conducted, in this paper, a dilation of ELS is used as the search region since it considers the uncertainty in both the elevation and the relative orientation error. The search region of the i-th key-point is defined as:

$$\mathcal{I}_i^{srch} := dilate((l_l)_i, R_{srch}) \cap \mathcal{I}_r \tag{4}$$

where R_{srch} is the radium of dilation and is determined by the relative orientation error between the two views of the satellite images. Obviously, with a smaller search region, the two-view image-matching is more efficient and the proportions of mismatches are lower.

Define $\mathcal{I}_i^{inl}(\mathbf{A}, d)$ as the inlier region of the i-th point-match with the affine model $\mathbf{A}$ and the P2L distance d. The $\mathcal{I}_i^{inl}(\mathbf{A}, d)$ is on the right image space $\mathcal{I}_r$. The points within $\mathcal{I}_i^{inl}(\mathbf{A}, d)$ satisfy:

$$distance(\mathbf{A}\cdot(l_l)_i, p) \leqslant d \tag{5}$$

Obviously, the inlier region is a dilation of line-segment $\mathbf{A}\cdot(l_l)_i$ with a radium d:

$$\mathcal{I}_i^{inl}(\mathbf{A}, d) := dilate(\mathbf{A}\cdot(l_l)_i, d) \tag{6}$$

The definition of the search region and the inlier region is illustrated in Fig. 3.

Define an atom event $e_i(d)$ as the occurrence of point-matches $(p_l, p_r)_i$ of which the P2L distance $d_{\mathbf{A}}(i)$ is less than d. As is seen in Fig. 3, when $e_i(d)$ occurs, $(p_r)_i$ must be within both the search region $\mathcal{I}_i^{srch}$ and the inlier region $\mathcal{I}_i^{inl}(\mathbf{A}, d)$. Under the null-hypothesis, since the p_r s are randomly and uniformly distributed, the probability of $e_i(d)$ satisfies:

$$\begin{aligned} Prob(d_{\mathbf{A}}(i) \leqslant d|\mathcal{H}_0) &= area(\mathcal{I}_i^{inl}(\mathbf{A}, d) \cap \mathcal{I}_i^{srch})/area(\mathcal{I}_i^{srch}) \\ &\leqslant area(\mathcal{I}_i^{inl}(\mathbf{A}, d))/area(\mathcal{I}_i^{srch}) \end{aligned} \tag{7}$$

Define:

$$\begin{aligned} AreaRatio(i, d, \mathbf{A}) &:= area(\mathcal{I}_i^{inl}(\mathbf{A}, d))/area(\mathcal{I}_i^{srch}) \\ &= (2d\cdot length(\mathbf{A}\cdot(l_l)_i) + \pi d^2)/area(\mathcal{I}_i^{srch}) \end{aligned} \tag{8}$$

$AreaRatio(i,d,\mathbf{A})$ is an increasing function of variable d, when $area(\mathcal{I}_i^{srch})$ and $length(\mathbf{A}\cdot(l_l)_i)$ are the constants. In this paper, we chose to compute the value $AreaRatio$ $(i,d,\mathbf{A})$ rather than the rigorous probability to measure the accuracy of point-matches to avoid computing the area of an intersected region. If the point-matches of which the search regions cross the image border are ignored, the area of the search region can be computed by:

$$area(\mathcal{I}_i^{srch}) = 2R_{srch}\cdot length(\mathbf{A}\cdot(l_l)_i) + \pi R_{srch}^2 \tag{9}$$

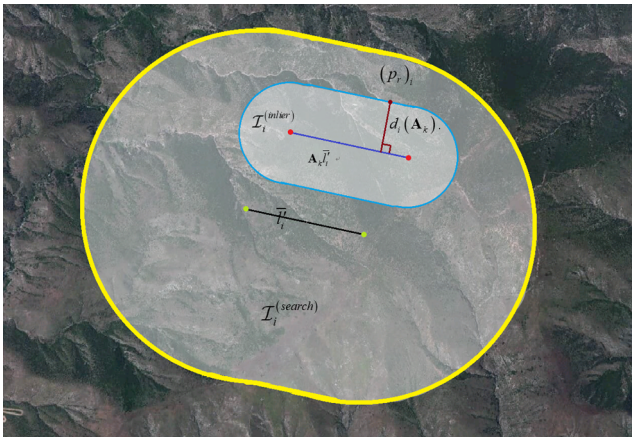

Fig. 3. The definition of the search region and the inlier region.

3.1.1. One-on-one point matches

For a set $\mathcal{S}$ of one-on-one point matches, define the $\mathbf{A}$-rigidity of its i-th point-match $(p_l, p_r)_i$ as:

$$\alpha_{\mathbf{A}}(\mathcal{S}, i) := AreaRatio(i, d_{\mathbf{A}}(i), \mathbf{A}) \tag{10}$$

Then the $\mathbf{A}$-rigidity of set $\mathcal{S}$ is defined as:

$$\alpha_{\mathbf{A}}(\mathcal{S}) := \max_{(p_l,p_r)_i \in \mathcal{S}} \alpha_{\mathbf{A}}(\mathcal{S}, i) \tag{11}$$

Using the maximum of $\alpha_{\mathbf{A}}(i)$ makes $\alpha_{\mathbf{A}}(\mathcal{S})$ very selective to outliers. $\alpha_{\mathbf{A}}(\mathcal{S})$ gives a probabilistic rigidity measurement to $\mathcal{S}$ under $\mathbf{A}$.

The global rigidity of a set of point matches should be defined with an affine model $\mathbf{A}$ that minimizes the $\alpha_{\mathbf{A}}(\mathcal{S})$s. However, finding such an $\mathbf{A}$ is difficult since an infinite number of 2D affine transformations exist. Here, a more computational process is adopted in which only the affine transformations determined by one of the possible 3-subsets of $\mathcal{S}$ are considered. Thus, the definition of the global rigidity of a set of point-matches is given as:

Definition 1. A set $\mathcal{S}$ of two-view satellite image matchings is α-rigid if an affine transformation $\mathbf{A}$ associated to one 3-subset of $\mathcal{S}$ exists and the following condition is satisfied:

$$\alpha_{\mathbf{A}}(\mathcal{S}) \leqslant \alpha \tag{12}$$

In other words, with any affine transformation $\mathbf{A}$s which can be resolved with a 3-subset of $\mathcal{S}$, the set $\mathcal{S}$ is $\alpha_{\mathbf{A}}(\mathcal{S})$-rigid. A conclusion, therefore, can be made that under null-hypothesis $\mathcal{H}_0$, the occurrence of an α-rigid satisfies:

$$Prob(\alpha_{\mathbf{A}}(\mathcal{S}) \leqslant \alpha|\mathcal{H}_0) \leqslant \alpha^{n-3} \tag{13}$$

Proof of (13): Since $\mathcal{S}$ is α-rigid, there exists an affine transformation $\mathbf{A}$ associated with a 3-subset of $\mathcal{S}$, under which $\alpha_{\mathbf{A}}(\mathcal{S}, i) \leqslant \alpha$ is satisfied with any $(p_l, p_r)_i \in \mathcal{S}'$. For the $(p_l, p_r)_i$ associated with $\mathbf{A}$, $\alpha_{\mathbf{A}}(\mathcal{S}, i) = 0$ and $Prob(\alpha_{\mathbf{A}}(\mathcal{S}, i) \leqslant \alpha|\mathcal{H}_0) = 1$. For the other $(n\text{-}3)$ point-matches, define $d(i,\alpha,\mathbf{A})$ as the solution of the following equation where i, $\mathbf{A}$, and α are constants:

$$AreaRatio(i, d, \mathbf{A}) = \alpha \tag{14}$$

According to (8), the function *AreaRatio*(*i*,*d*,**A**) is an increasing function with the variable *d*. Combining (7), (8), (10), and (14):

$$\begin{aligned} Prob(\alpha_{\mathbf{A}}(\mathcal{S}, i) \leqslant \alpha | \mathcal{H}_0) &= Prob(AreaRatio(i, d_{\mathbf{A}}(i), \mathbf{A}) \leqslant \alpha | \mathcal{H}_0) \\ &= Prob(d_{\mathbf{A}}(i) \leqslant d(i, \alpha, \mathbf{A}) | \mathcal{H}_0) \\ &\leqslant AreaRatio(i, d(i, \alpha, \mathbf{A}), \mathbf{A}) \\ &= \alpha \end{aligned} \tag{15}$$

Under the null-hypothesis, all the point-matches are independent and randomly distributed so:

$$\begin{aligned} Prob(\alpha_{\mathbf{A}}(\mathcal{S}) \leqslant \alpha | \mathcal{H}_0) &= \prod_{(p_l, p_r)_i \in \mathcal{S}'} Prob(\alpha_{\mathbf{A}}(\mathcal{S}, i) \leqslant \alpha | \mathcal{H}_0) \\ &\leqslant \alpha^{k-3} \end{aligned} \tag{16}$$

3.1.2. One-on-multiple point matches

In some cases, the true match of a key-point may be at the second or the third most similar location on the right image $\mathcal{I}_r$. If only the most similar locations to the key-points are considered, many true matches may be lost. With some special kinds of image features that are not easy to match, all the possible matches may be retrieved within the $\mathcal{I}_i^{srch}$ and then the correct matches can be selected by geometric constraints. Denote such a set of one-on-several one-on-multiple point-matches as:

$$\mathcal{M} := \{(p_l, \{(p_r)_j | j = 1, 2, ..., m_i\})_i \,|\, i = 1, 2, ..., n,\} \tag{17}$$

where n is the number of key-points; m_i is the number of point-matches of the i-th key-point. By the way, with the left key-points transformed onto the right image-space with ∇H as follows:

$$\mathcal{M}_{P2L} := \{(l_l, \{p_r\}_j)_i | (l_l)_i \in \mathcal{I}_r, (p_r)_{i,j} \in \mathcal{I}_r\} \tag{18}$$

Denote the P2L distance of the j-th point-match of the i-th key-point under the affine transformation **A** as:

$$d_{\mathbf{A}}(i, j) := distance(\mathbf{A} \cdot (l_l)_i, (p_r)_{i,j}) \tag{19}$$

and according to (7), we can easily prove that

$$Prob(d_{\mathbf{A}}(i, j) \leqslant d | \mathcal{H}_0) \leqslant AreaRatio(i, d, \mathbf{A}) \tag{20}$$

Then we should define $\alpha_{\mathbf{A}}(\mathcal{M}, i)$, the **A**-rigidity of the i-th one-on-multiple point-match in $\mathcal{M}$, and $\alpha_{\mathbf{A}}(\mathcal{M})$, the **A**-rigidity of $\mathcal{M}$. The two definitions should ensure that we can draw a similar conclusion as is in (13). Thus, $\alpha_{\mathbf{A}}(\mathcal{M}, i)$ is defined as m_i times of the minimum area-ratio

$$\alpha_{\mathbf{A}}(\mathcal{M}, i) := m_i \cdot AreaRatio(i, \min_j d_{\mathbf{A}}(i, j), \mathbf{A}) \tag{21}$$

because the following inequality is still satisfied as is in (15)

$$Prob(\alpha_{\mathbf{A}}(\mathcal{M}, i) \leqslant \alpha | \mathcal{H}_0) \leqslant \alpha \tag{22}$$

Proof of (22): According to the definition of $\alpha_{\mathbf{A}}(\mathcal{M}, i)$ in (21), and denoting $d(i, \alpha/m_i, \mathbf{A})$ as the solution of equation $AreaRatio(i, d, \mathbf{A}) = \alpha/m_i$, the possibility in (22) is equal to:

$$\begin{aligned} &Prob(\alpha_{\mathbf{A}}(\mathcal{M}, i) \leqslant \alpha | \mathcal{H}_0) \\ &\quad = Prob(AreaRatio(i, \min_j d_{\mathbf{A}}(i, j), \mathbf{A}) \leqslant \alpha/m_i \,\big|\, \mathcal{H}_0) \\ &\quad = Prob(\min_j d_{\mathbf{A}}(i, j) \leqslant d(i, \alpha/m_i, \mathbf{A}) \,\big|\, \mathcal{H}_0) \end{aligned} \tag{23}$$

The possibility that at least one of the $d_{\mathbf{A}}(i, j)$s is no-larger-than $d(i, \alpha/m_i, \mathbf{A})$ is equivalent to one subtracting the possibility that all the $d_{\mathbf{A}}(i, j)$s are larger than $d(i, \alpha/m_i, \mathbf{A})$. Combing (20), we have

$$\begin{aligned} &Prob(\min_j d_{\mathbf{A}}(i, j) \leqslant d(i, \alpha/m_i, \mathbf{A}) \,\big|\, \mathcal{H}_0) \\ &\quad = 1 - (Prob(d_{\mathbf{A}}(i, j) > d(i, \alpha/m_i, \mathbf{A}) | \mathcal{H}_0))^{m_i} \\ &\quad = 1 - (1 - Prob(d_{\mathbf{A}}(i, j) \leqslant d(i, \alpha/m_i, \mathbf{A}) | \mathcal{H}_0))^{m_i} \\ &\quad \leqslant 1 - (1 - AreaRatio(i, d(i, \alpha/m_i, \mathbf{A}), \mathbf{A}))^{m_i} \\ &\quad = 1 - (1 - \alpha/m_i)^{m_i} \end{aligned} \tag{24}$$

According to the inequality $(1 - \eta)^m \geqslant 1 - m \cdot \eta$ that is satisfied with $0 \leqslant \eta < 1$ and $m \in \mathbb{N}^*$

$$1 - (1 - \alpha/m_i)^{m_i} \leqslant 1 - (1 - m_i \cdot \alpha/m_i) = \alpha \tag{25}$$

Combining (23), (24), and (25), the inequality of (22) is proved.

Similar to definition 1, the definition of the global rigidity of $\mathcal{M}$ is determined. Before that process, it should be noted that the affine model **A** associated with a 3-subset of $\mathcal{M}$ is actually associated with a triplet of point-pairs selected from the 3-subset of $\mathcal{M}$. Within a 3-subset of $\mathcal{M}$, there are $m_{i(1)} \cdot m_{i(2)} \cdot m_{i(3)}$ numbers of triplets of point-pairs that can be selected.

Definition 2. A set $\mathcal{M}$ of two-view satellite image matchings is α-rigid if an affine transformation **A** associated to a 3-subset of $\mathcal{M}$ exists and the following condition is satisfied:

$$\alpha_{\mathbf{A}}(\mathcal{M}) \leqslant \alpha \tag{26}$$

According to (22), we can easily prove the following inequality with a similar manner in the proof of (13):

$$Prob(\alpha_{\mathbf{A}}(\mathcal{M}) \leqslant \alpha | \mathcal{H}_0) \leqslant \alpha^{n-3} \tag{27}$$

Obviously, when fixing all the m_is at 1, the set $\mathcal{S}$ of one-on-one point-matches is a special case of set $\mathcal{M}$, the Definition 2 degenerates to the Definition 1, and the inequality (27) degenerates to (13). Thus, the following analysis targeted to set $\mathcal{M}$ also considers the cases of $\mathcal{S}$.

3.2. Measuring meaningfulness

In the *a-contrario* methods, whether a structure can be detectable by human sight in an image is determined by the NFA of similar structures. The NFA can be estimated as:

$$NFA(E) = N_{occur} \cdot Prob_{occur}(E | \mathcal{H}_0) \tag{28}$$

When the condition $NFA(E) \leqslant \varepsilon$ is satisfied, event E is ε-meaningful. If an event is 1-meaningful, the associated image structure is detectable by human sight. In (28), E is the event when an image structure occurs; N_{occur} is the total number of potentially occurring similar events; and $Prob_{occur}(E|\mathcal{H}_0)$ is the probability of the occurrence of E under null-hypothesis $\mathcal{H}_0$. Accurate computation of the NFA is can be very complicated. A more frequently used method is designing a simpler formula, denoted as $\varepsilon(E)$. When condition $NFA(E) \leqslant \varepsilon(E)$ is always satisfied, event E is ε-meaningful when $\varepsilon(E) \leqslant \varepsilon$.

In this section, a formula is defined to estimate the NFA of the subsets of $\mathcal{M}$. Inspired by the theory in (Moisan and Stival, 2004), a proposition about a subset of satellite image matching set is as follows:

Proposition 1. Denote $\mathcal{M}$ as a set of n point matches of two-view satellite images. An α-rigid subset $\mathcal{M}' \subset \mathcal{M}$ of k point-matches is judged ε-meaningful if the following condition is satisfied:

$$\varepsilon(\alpha, n, k) := (n - 3) \cdot \binom{n}{k} \cdot \binom{k}{3} \cdot N_{set} \cdot N_{slt} \cdot \alpha^{k-3} \leqslant \varepsilon \tag{29}$$

The definition of $\varepsilon(\alpha,n,k)$ supplies a criterion to compare the meaningfulness of a subset which has a lot of point-matches but is less rigid with a subset which has fewer point-matches but is more rigid.

Proof of Proposition 1: In (29), $(n\text{-}3)$ is the number of choices of k; $\binom{n}{k}$ is the number of possible subsets with k point-matches from n; $\binom{k}{3}$

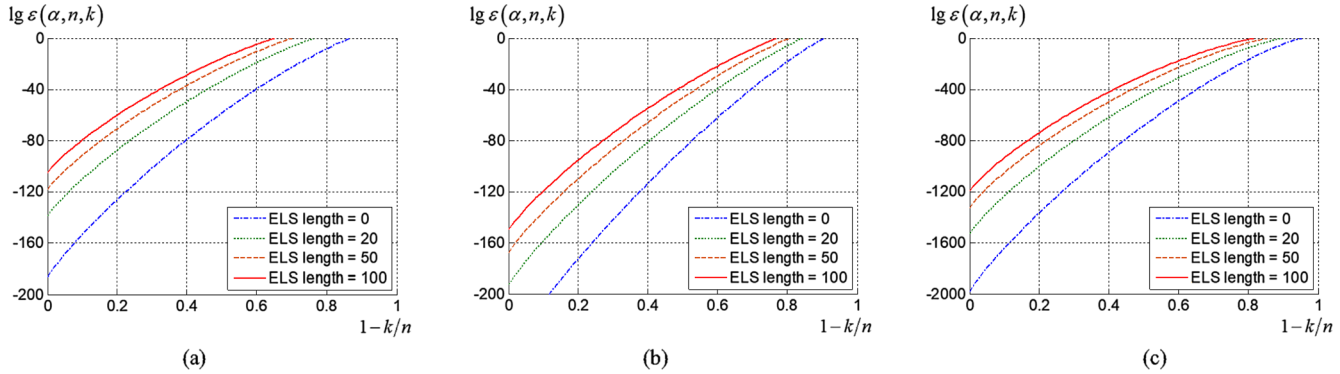

Fig. 4. The $\lg(\varepsilon_1) \sim (1 - k/n)$ curves of the one-on-one point-matches set, where $d = 5$ pixels; $N_{set} = 1$; the $length(\mathbf{A}\cdot(l_i))$ is set to 0, 20, 50, or 100 pixels; (a) $n = 100$, $R_{srch} = 50$ pixels, (b) $n = 100$, $R_{srch} = 100$ pixels, (c) $n = 1000$, $R_{srch} = 50$ pixels;

is the number of possible 3-subsets of $\mathcal{M}'$ used for recovering the affine transformations; N_{set} is the maximum number of sets of three point-pairs selected form the 3-subset of $\mathcal{M}'$; N_{slt} is the maximum number of affine transformations resolved by the sets of three point pairs. Thus:

$$N_{occur} \leqslant (n-3)\cdot\binom{n}{k}\cdot\binom{k}{3}\cdot N_{set}\cdot N_{slt} \tag{30}$$

Combining (27) and (30), proposition 1 is proved by:

$$\begin{aligned} NFA(\alpha_{\mathbf{A}}(\mathcal{M}') \leqslant \alpha | \mathcal{H}_0) &= N_{occur}\cdot Prob(\alpha_{\mathbf{A}}(\mathcal{M}') \leqslant \alpha | \mathcal{H}_0) \\ &\leqslant (n-3)\cdot\binom{n}{k}\cdot\binom{k}{3}\cdot N_{set}\cdot N_{slt}\cdot\alpha^{k-3} \\ &= \varepsilon(\alpha, n, k) \end{aligned} \tag{31}$$

In (29), N_{set} is computed by multiplying the three largest m_is, and N_{slt} is computed by multiplying the pixel lengths of the three longest ELSs, which is as follows:

$$\begin{aligned} N_{set} &= \prod_{t=1}^{3} \max(t)\{m_i\} \\ N_{slt} &= \prod_{t=1}^{3} \max(t)\{length(l_i)\} \end{aligned} \tag{32}$$

where $\max(t)\{m_i\}$ is the t-th largest number in the set $\{m_i\}$; and $\max(t)\{length(l_i)\}$ is the length of the t-th longest ELSs among the set $\mathcal{M}_{P2L}$.

Fig. 4 illustrates the $\lg(\varepsilon_1) \sim (1 - k/n)$ curves of set $\mathcal{S}$ of one-on-one point-matches. The maximum $(1 - k/n)$ of 1-meaningful sub-sets becomes lower with smaller n, larger d/R_{srch}, or longer $\mathbf{A}\cdot(l_i)$, which means that with a larger uncertainty of elevation or a larger convergence angle, the tolerance of the proportion of outliers becomes weaker.

3.3. Algorithms

3.3.1. Detect the most meaningful sub-sets

With the definition of $\varepsilon(\alpha,n,k)$, which measures the meaningfulness of the subsets of $\mathcal{M}$, the mismatches can be detected by finding the optimal sub-set that has the smallest $\varepsilon(\alpha,n,k)$ because the rigidity of true matches is far from the randomness under $\mathcal{H}_0$. First, the optimal subset of $\mathcal{M}$ under the affine transformation $\mathbf{A}$, denoted as $\overline{\mathcal{M}}(\mathbf{A})$, must be found. There are $(2^n\text{-}1\text{-}n\text{-}n(n\text{-}1)/2)$ numbers of sub-sets from $\mathcal{M}$ which have more than three elements, but not all of them need to be considered since the combination of very rigid point-matches and poorly rigid ones are meaningless. Sorting the point-matches according to their $\mathbf{A}$-rigidities and constructing the subsets from the first k point-matches, denoted as $\mathcal{M}(\mathbf{A}, k)$ first must be accomplished. Denote the $\mathbf{A}$-rigidity of $\mathcal{M}(\mathbf{A}, k)$ as $\alpha_{\mathbf{A}}(\mathcal{M}(\mathbf{A}, k))$. The optimal number of k and the corresponding optimal subset $\overline{\mathcal{M}}(\mathbf{A})$ are defined by:

$$\begin{aligned} \overline{k}(\mathbf{A}) &= \arg\min_{k=4,\ldots,n} \varepsilon(\alpha_{\mathbf{A}}(\mathcal{M}(\mathbf{A}, k)), n, k) \\ \overline{\mathcal{M}}(\mathbf{A}) &= \mathcal{M}(\mathbf{A}, \overline{k}(\mathbf{A})) \end{aligned} \tag{33}$$

The workflow of determining the optimal subset $\overline{\mathcal{M}}(\mathbf{A})$ under the affine model $\mathbf{A}$ is described in Algorithm 1 below:

Algorithm 1: get the optimal subset $\overline{\mathcal{M}}(\mathbf{A})$

Input data: $\mathcal{M}$; $\mathcal{M}_{P2L}$; $\mathbf{A}$
Output: $\overline{\mathcal{M}}(\mathbf{A})$; $\lg\bar{\varepsilon}(\mathbf{A})$.

Steps:
1. compute the $\alpha_{\mathbf{A}}(\mathcal{M}, i)$ s according to (21);
2. sort the matches according to $\alpha_{\mathbf{A}}(\mathcal{M}, i)$ s in ascending order;
3. $\lg\bar{\varepsilon}(\mathbf{A}) \leftarrow \infty$
4. **for** $k \leftarrow 4$ to n **do**
5. construct $\mathcal{M}(\mathbf{A}, k)$ with the first k point-matches;
6. compute $\lg\varepsilon(\alpha, n, k)$ according to ;
7. **if** $\lg\varepsilon(\alpha, n, k) < \lg\bar{\varepsilon}$ **do**
8. $\overline{\mathcal{M}}(\mathbf{A}) \leftarrow \mathcal{M}(\mathbf{A}, k)$; $\lg\bar{\varepsilon}(\mathbf{A}) \leftarrow \lg\varepsilon(\alpha, n, k)$;
9. **end_if**
10. **end_for**
11. **output** $\overline{\mathcal{M}}(\mathbf{A})$ and $\lg\bar{\varepsilon}(\mathbf{A})$

Fig. 5 illustrates the $\lg\varepsilon(\alpha_{\mathbf{A}}(\mathcal{S}(\mathbf{A}, k)), n, k)k^{\sim}$ curve, which is drawn with affine model $\mathbf{A}$ resolved by three true matches from set $\mathcal{S}$, which includes 50 artificial point-matches and 250 simulated mismatches. The procedure for constructing set $\mathcal{S}$ and the detailed information for the

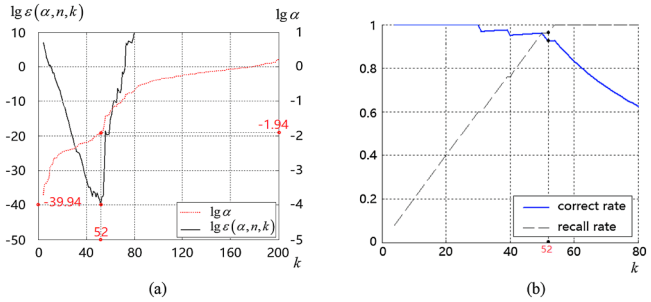

Fig. 5. (a) The $\lg\varepsilon \sim k$ curve and the $\lg\alpha \sim k$ curve. The lowest $\lg\varepsilon$ occurs at k = 52, meanwhile, $\lg\alpha$ is −1.94. (b) the precision $\sim k$ curve and the recall $\sim k$ curve. At the optimal $\overline{k}(\mathbf{A})$, 49 true matches out of 50 are involved in the subset of inliers.

tested satellite images is provided in Section 4. As shown in Fig. 5, the **A**-rigidity of all the true matches are limited and obviously smaller than the outliers. As a result, the ε declines and the α increases slowly when more true matches are included with an increasing k when k is under 52. However, when the k exceeds 52, the mismatches begin to be included. As a result, the increasing α becomes significant and $\varepsilon(\alpha,n,k)$ ceases the trend of declining and begin to rise.

After obtaining the optimal subset $\overline{\mathcal{M}}(\mathbf{A})$, the set of inliers are extracted as:

$$\overline{\mathcal{S}}(\mathbf{A}) = \{(p_l, (p_r)_{j(i,\mathbf{A})})_i \,|\, j(i,\mathbf{A}) = \arg\min_j d_{\mathbf{A}}(i,j)\} \tag{34}$$

The total number of 3-subsets of $\mathcal{M}$ is $n(n\text{-}1)(n\text{-}2)/6$, which means that considering all the possible 3-subsets is too time-consuming when there are hundreds or thousands of point-matches in $\mathcal{M}$. Thus, the random sampling method is used to test as many 3-subsets as possible to ensure that at least one is made up with true matches. The 3-subset will be used for solving affine transformations by the method described in Section 2.2. The workflow of the random sampling process is described in Algorithm 2 below:

Algorithm 2: random sampling process
Input data: $\mathcal{M}_{P2L}$; N_{set}; N_{slt}

Input coefficients: maximum iteration number $N_{\max}$.
Output: $\bar{\mathcal{S}}'$; $\lg\bar{\varepsilon}$; $\overline{\mathbf{A}}$; $\overline{\mathcal{A}}$.

Steps:
1. $\bar{\mathcal{S}}' \leftarrow \varnothing$; $\bar{\mathcal{M}}' \leftarrow \varnothing$; $\overline{\mathbf{A}} \leftarrow \mathbf{I}$; $\overline{\mathcal{A}} \leftarrow \varnothing$; $\lg\bar{\varepsilon} \leftarrow \infty$; $N_{iter} \leftarrow 0$;
2. **while** $N_{iter} < N_{max}$ **do**
3. sampling 3-subsets from $\mathcal{M}_{P2L}$;
4. resolve $\mathcal{A} = \{\mathbf{A}\}$ with the 3-subsets;
5. **for_each** $\mathbf{A} \in \mathcal{A}$
6. process **Algorithm 1** to get $\overline{\mathcal{M}}(\mathbf{A})$ and $\lg\bar{\varepsilon}(\mathbf{A})$;
7. **if** $\lg\bar{\varepsilon}(\mathbf{A}) < \lg\bar{\varepsilon}$ **do**
8. $\bar{\mathcal{M}}' \leftarrow \overline{\mathcal{M}}(\mathbf{A})$; $\lg\bar{\varepsilon} \leftarrow \lg\bar{\varepsilon}(\mathbf{A})$; $\overline{\mathbf{A}} \leftarrow \mathbf{A}$; $\overline{\mathcal{A}} \leftarrow \mathcal{A}$;
9. **end_if**
10. **end_for_each**
11. $N_{iter} \leftarrow N_{iter} + 1$
12. **end while**
13. generate $\bar{\mathcal{S}}'$ from $\bar{\mathcal{M}}'$ according to (34);
14. **output** $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$, $\overline{\mathbf{A}}$, and $\overline{\mathcal{A}}$;

Moisan and Stival (2004) recommended the following optimizing strategy when the proportion of inliers is very low: add another set of iterations in which the seeds are sampled only from $\bar{\mathcal{S}}'$ which is the most meaningful subset previously obtained since subsets $\bar{\mathcal{S}}'$ usually had larger proportions of true-matches than the original set $\mathcal{M}$ of point-matches. The additional set of iterations affords the opportunity to obtain 3-subsets comprised of all true-matches and thus more meaningful subsets. The optimizing workflow in Algorithm 3 is as follows.

Algorithm 3: additional random sampling process

Input data: $\mathcal{M}_{P2L}$; N_{set}; N_{slt}; the results of **Algorithm 2**: $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$, $\overline{\mathbf{A}}$, and $\overline{\mathcal{A}}$;
Input coefficients: maximum iteration number $N_{\max}$.

Output: $\bar{\mathcal{S}}'$; $\lg\bar{\varepsilon}$; $\overline{\mathbf{A}}$; $\overline{\mathcal{A}}$.
Steps:
1. $\bar{\mathcal{M}}' \leftarrow \varnothing$; $N_{iter} \leftarrow 0$;
2. **while** $N_{iter} < N_{max}/10$ **do**
3. sampling 3-subsets from $\bar{\mathcal{S}}'$;
4. resolve $\mathcal{A} = \{\mathbf{A}\}$ with the 3-subsets;
5. **for_each** $\mathbf{A} \in \mathcal{A}$
6. process **Algorithm-1** to get $\overline{\mathcal{M}}(\mathbf{A})$ and $\lg\bar{\varepsilon}(\mathbf{A})$;
7. **if** $\lg\bar{\varepsilon}(\mathbf{A}) < \lg\bar{\varepsilon}$ **do**
8. $\bar{\mathcal{M}}' \leftarrow \overline{\mathcal{M}}(\mathbf{A})$; $\lg\bar{\varepsilon} \leftarrow \lg\bar{\varepsilon}(\mathbf{A})$; $\overline{\mathbf{A}} \leftarrow \mathbf{A}$; $\overline{\mathcal{A}} \leftarrow \mathcal{A}$;
9. **end_if**
10. **end_for_each**
11. $N_{iter} \leftarrow N_{iter} + 1$
12. **end while**
13. generate $\bar{\mathcal{S}}'_{add}$ from $\bar{\mathcal{M}}'_{add}$ and $\overline{\mathbf{A}}$ according to (34);
14. **output** $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$, $\overline{\mathbf{A}}$, and $\overline{\mathcal{A}}$;

3.3.2. Adjust the uncertainty of the elevation

The variable ∇H, measuring the elevation uncertainty in the back-projection of image points, is determined by two factors: (1) the elevation error of the DEM, (2) the horizontal geo-reference error if the land is uneven. Zhang et al. (2016) introduced a method to estimate σ_H, the a priori elevation error of back-projection. To ensure the true matches within the search region $\mathcal{I}_i^{srch}$, ∇H is usually exaggerated during the image matching. We suggest using $3\sigma_H$ or even larger value as ∇H.

Since ∇H is exaggerated, the lengths of ELSs are exaggerated. As a result, the area of the inlier region, the geometric rigidity measurement, and the meaningfulness measurement $\varepsilon(\alpha,n,k)$ are all larger than what they should be when computed with an appropriate ∇H, and more undetected mismatches are present, for which the errors are located along the direction of the ELSs. The following steps were added to this process to deal with this problem. First, ∇H is gradually reduced from its initial value; then, all the ELSs are recalculated according to the adjusted ∇H. Finally, with each adjusted ∇H, an optimal affine model and its associated most meaningful subsets is obtained, and by comparing the meaningfulness, the optimal ∇H is obtained.

The workflow of the whole algorithm, named as ORSA-SAT, is as follows:

Algorithm 4: Optimized Random Sampling Algorithm for Satellite Image

Input data: the set $\mathcal{M}$; the initial geo-reference models of the two images; a DEM.
Input coefficients: the uncertainty of elevation ∇H; maximum iteration number $N_{\max}$.
Output: the optimal set of point-pairs $\bar{\mathcal{S}}'$.

Steps:
1. generate $\mathcal{M}_{P2L}$ according to ;
2. compute N_{set} and N_{slt} according to ;
3. process **Algorithm 2** to get $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$, $\overline{\mathbf{A}}$, and $\overline{\mathcal{A}}$.
4. **if** $size(\bar{\mathcal{S}}') < size(\mathcal{M})/2$
5. process **Algorithm 3** to optimize $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$, $\overline{\mathbf{A}}$, and $\overline{\mathcal{A}}$.
6. **end_if**
7. **for_each** η **in** {0. 9, 0. 8, 0. 7, ..., 0. 1, 0.0}
8. $\nabla H_{adj} \leftarrow \eta \cdot \nabla H$
9. generate $\mathcal{M}_{P2L}$ according to using ∇H_{adj};
10. **for_each** $\mathbf{A} \in \overline{\mathcal{A}}$
11. process **Algorithm-1** to get $\overline{\mathcal{M}}(\mathbf{A})$ and $\lg\bar{\varepsilon}(\mathbf{A})$;
12. **if** $\lg\bar{\varepsilon}(\mathbf{A}) < \lg\bar{\varepsilon}$ **do**
13. $\bar{\mathcal{M}}' \leftarrow \overline{\mathcal{M}}(\mathbf{A})$; $\lg\bar{\varepsilon} \leftarrow \lg\bar{\varepsilon}(\mathbf{A})$; $\overline{\mathbf{A}} \leftarrow \mathbf{A}$;
14. **end_if**
15. **end_for_each**
16. **end_for_each**
17. **if** $\bar{\varepsilon} < 1$
18. generate $\bar{\mathcal{S}}'$ from $\bar{\mathcal{M}}'$ with $\overline{\mathbf{A}}$ according to (34);
19. **else**
20. $\bar{\mathcal{S}}' \leftarrow \varnothing$
21. **end_if**
22. **output** $\bar{\mathcal{S}}'$, $\lg\bar{\varepsilon}$ and $\overline{\mathbf{A}}$;

4. Simulated experiments

4.1. Experimental data

In this section, two overlapped Ikonos-2 images with 1 m ground resolution and a 38-deg convergence angle are used to examine the ability of the ORSA-SAT to remove mismatches from a mixture of true matches and simulated mismatches without knowing the apriori matching accuracy of the true matches. The effect of adjusting the

Table 2
The information of the satellite images used in the simulated experiments.

Image	Conv. Ang. (deg)	Location	Size in pixel (width, height)	Acquisition date/time (GMT time)
Left	38	112.7E, 38.0N	11916, 27,640	02/23/02 03:27
Right		112.7E, 38.2N	14096, 33,588	10/05/05 03:44

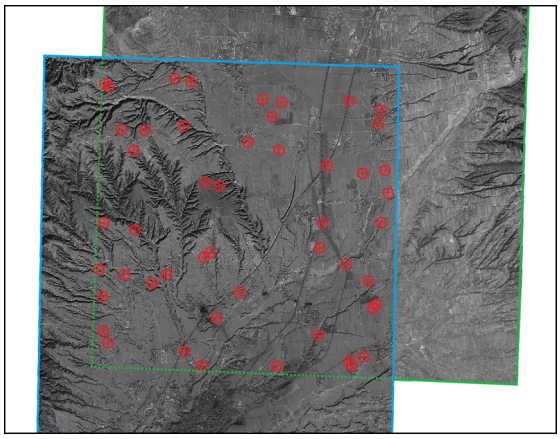

Fig. 6. The overlapping area of the two satellite images and the distribution of artificial matching points.

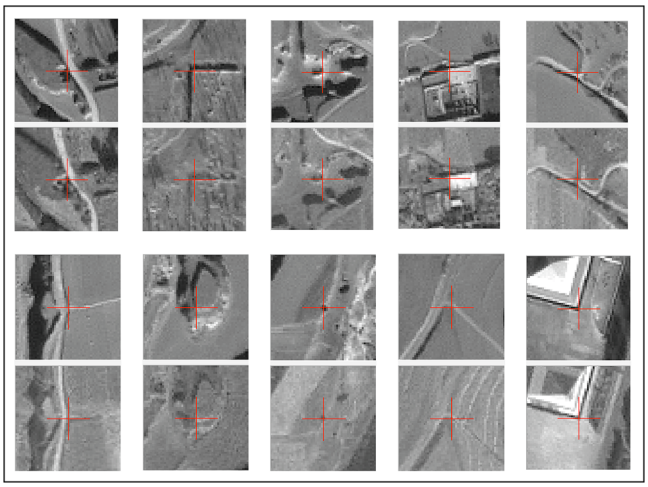

Fig. 7. Screenshots of some of the artificial matching points.

elevation uncertainty also was analyzed.

The two images were acquired in Shanxi province in China (see Table 2). The landscape is mainly flat plateaus and valleys. The elevation varies from about 900 m to about 1100 m. Fifty point-matches were artificially collected for which the geo-location accuracy was better than one pixel. The mismatches were randomly simulated, the key-points of which were randomly distributed within the overlapped area on the left image while the matches were randomly distributed within the search region $\mathcal{I}_i^{srch}$ of the key-points. The search region $\mathcal{I}_i^{srch}$ was the dilation of the i-th ELS and the dilation radius R_{srch} (see Fig. 3) was 30 pixels. The distribution of the artificial matching points is shown in Fig. 6. Some screenshots of the artificial true matches are shown in Fig. 7. The 90 m-resolution SRTM-DEM was used in the ORSA-SAT and the uncertainty of the back-projected elevation was set at 30 m.

4.2. Influence of the positional error of true matches

In this experiment, different types of positional errors were added to the geo-location of the matching points to simulate some typical errors in the optical pushbroom satellite imagery which are caused by lens distortion, oscillation, CCD installation errors, or other factors. The simulated errors included random error, periodic error, and piecewise polynomial error. There were six types of positional errors simulated on the matching points on the right image $\mathcal{I}_r$:

(1) Error type-1: no additional error was added.
(2) Error type-2: random error under three pixels in both the x- and y-directions.
(3) Error type-3: periodic error simulated by a trigonometric function with about 2 Hz frequency and 3-pixel amplitude in both the x- and y- directions.
(4) Error type-4: a mixture of type-2 (random) error and type-3 (periodic) error.
(5) Error type-5: a piecewise linear error under three pixels in both the x- and y- direction.
(6) Error type-6: a mixture of type-2 (random) error and type-5 (piecewise linear) error.

The number of simulated mismatches was increased from 0 to 950 so that the initial proportion of true matches, denoted as p, decreased from 100% to 5%. With a certain type of error and a certain number of mismatches, the process of data simulation and mismatch detection was conducted 100 times. Finally, the following values were computed for evaluating the ORSA-SAT method:

(1) The mean value of $\lg\varepsilon(\alpha,n,k)$s from 100 times of positional error simulation and mismatch detection.
(2) The empirical probability of successfully detecting a rigid set of inliers that has less-than-one $\varepsilon(\alpha,n,k)$.
(3) The average recalls. The mean recall rates were calculated with only valid results, i.e., the optimal subsets satisfying $\varepsilon(\alpha, n, k) \leqslant 1$.
(4) The average precisions. The average precision was also calculated with only valid results.

Fig. 8 illustrates the P2L distances of a set of point-matches, which included 30% simulated mismatches. The P2L distances were computed with the optimal affine model $\overline{\mathbf{A}}$ associated with the most meaningful subset. The trend of the P2L distances of the true matches reveals the properties of the simulated geo-location error. With error type-1 (see Fig. 8(a)), the maximum P2L distance of the inliers occurred at about two pixels, which is close to the apriori geo-location error of the artificially matched points. With error type-2, -3, and -5 (see Fig. 8(b), (c), and (e)), the maximum P2L distances of the inliers occurred at about five to six pixels, which are close to the mixture of the artificial matching error and either the random error or the periodic error or the piece-wise linear error. With error type-4 and -6 (see Fig. 8(d) and (f)), the maximum P2L distances of the inliers occurred at about eight to nine pixels, which are close to the mixture of the artificial matching error, the random error, and either the periodic error or the piece-wise linear error. The results demonstrate that the strategy of finding the most meaningful subset in ORSA-SAT can effectively filter out the mismatches from the true matches, regardless of what types or scales of apriori errors are in the geo-location.

Fig. 9 illustrates the statistic results of the simulation experiment. The solid lines correspond to using a fixed elevation uncertainty ∇H, which means that step 7 through 16 of algorithm-4 are not executed. The dotted lines correspond to using adjusted ∇H, which means that

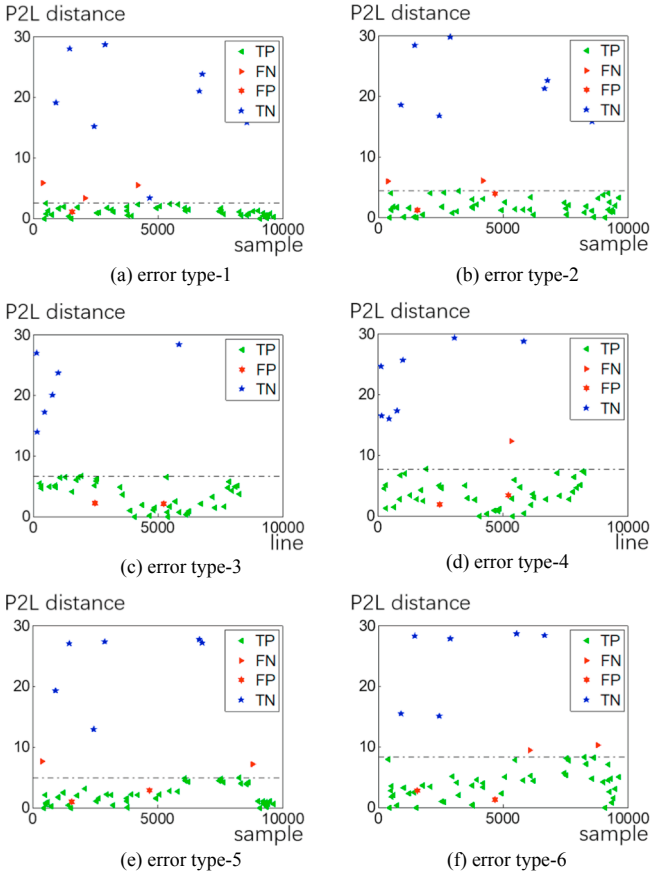

Fig. 8. The distribution of P2L distances of a set of point-matches containing 30% mismatches. The P2L distances were computed with the optimal affine model $\overline{\mathbf{A}}$ from the ORSA-SAT algorithm. Figure (a) through (f) correspond to the simulated positional error type-1 through type-6. The grey dash-dot lines indicate the maximum P2L distance of the inliers. Each marker corresponds to a point-match. TP means true positive; FN means false negative; FP means false positive; TN means true negative.

the whole algorithm-4 is executed. Obviously, when the ∇H was adjusted, the sets of inliers tended to have a lower $\varepsilon(\alpha,n,k)$s with any given proportion of mismatches; the rigid sets of inliers (that satisfy $\varepsilon(\alpha,n,k) < 1$) had a higher proportion of mismatches; and the precisions of the most meaningful subsets tended to be higher.

The statistic results demonstrate that the criterion of $\varepsilon(\alpha,n,k) < 1$ can be used to justify the correctness of the most meaningful subsets in the ORSA-SAT method because, as shown in the third and fourth columns, both the recall rates and the precisions of the results with the ∇H adjusted were higher than 80% when $\varepsilon(\alpha,n,k) < 1$ was satisfied, regardless of the error types simulated on the geo-location of matching points and how many mismatches were simulated in the original sets of point-matches.

The results also show that both the type and the scale of the simulated error on the matching points can influence the ability of ORSA-SAT to find the rigid sets of inliers. Although error type-2, 3, and 5 had similar scales, the $\varepsilon(\alpha,n,k)$s of the most meaningful subsets with error type-3 was obviously larger than those with error type-2 or 5. In conclusion, the systematic error, which varies continuously with the x- or y- coordinates, will do more harm to the mismatch detection than the discontinuously varying error.

4.3. Influence of the relative orientation error

The relative orientation accuracy of the two-view satellite images is determined by the geo-referencing model supplied by the vendors. When the relative orientation error is large, users will need large search regions to ensure that the correspondences of the key-points are within their search regions. However, more mismatches will be obtained and the running time will be long with large search regions.

In this experiment, we added horizontal movements to the left image introduced in Section 4.1 by modifying its RFM coefficients to test the ORSA-SAT algorithm under different relative orientation errors. The 50 artificial matches were still used as true matches, and when simulating mismatches, we still used the dilation of the ELS as the search region (see Fig. 3). Since the original RFM of the Ikonos image has about 5–10 m empirical geo-referencing error, the dilation radius R_{srch}, determined by the relative orientation accuracy, increased from 30 pixels to 60 pixels while we added other 0–30 m horizontal movements to the left image.

During the expansion of the search region, the ratio between the number of mismatches and the average area of the search regions was kept unchanged, which was closer to the reality than that keeping the number of mismatches unchanged. The reason is that with larger search region, a key-point that should have no correspondence is more probable to find a false match. Fig. 10(a) illustrates the variations of the number of mismatches and p (the initial proportion of true matches) according to the variation of R_{srch}. With each value of R_{srch}, the simulation and detection of mismatches were done for 100 times and the average $\lg\varepsilon(\alpha,n,k)$s, precisions and recalls were then calculated.

Fig. 10(b) and (c) show that there was no increasing or decreasing trend on the precisions and the recalls although the number of mismatches increased from 100 to 310 with the expansion of the relative orientation error and the search region. Even the $\lg\varepsilon(\alpha,n,k)$s were very stable. These results, so different with those in Section 4.2, demonstrated that the variation of the relative orientation error within a certain range has very little influence on the results of the ORSA-SAT algorithm. However, when the radius is too large, say exceeding 50 pixels, the matching process will need more running time and the proportion of mismatches will increase, which causes the ORSA-SAT algorithm to need more iterations of random samplings to get an accurate affine model. When the relative orientation is inaccurate, we suggest to down-sample the images and use a pyramid-matching strategy.

5. Experiments with real matched points

5.1. Data and procedure

In this section, five image-pairs were chosen and matched by the NCC matcher to obtain the automatic point-matches. The ORSA-SAT method and the P2L method were compared to determine their ability to detect the true matches from the automatic point-matches. Detailed information about the satellite images is provided in Table 3.

All the matching tasks were challenging. The images of group-1 and -2 were acquired over unstable landforms, which means the objects on the land may move slightly during the time intervals between the acquisition dates. The images of group-1 were acquired by the Ikonos-2 satellite over the Taklimakan Desert, which is the second largest flow desert in the world; and there was a four-year-interval between the acquisition times of the two images. The images of group-2 were acquired by the Ikonos-2 satellite over the Himalayas, where the snow-lines move with the variation of temperature in different seasons. The images of group-3 were acquired over northeastern China. One of the images was acquired in winter when the land was covered by heavy snows; the other image was acquired in summer when the land was covered by thick vegetation. Furthermore, the shadows of the buildings and trees were much longer in winter than in summer. In group-4, an Ikonos-2 image (1 m GSD) was matched on a reference image (2.5 m GSD), which was ortho-rectified with a ZY-3 panchromatic image and the SRTM-DEM (90 m GSD). The overlapped area was mostly farm land and villages in central China. Group-4 was the most challenging

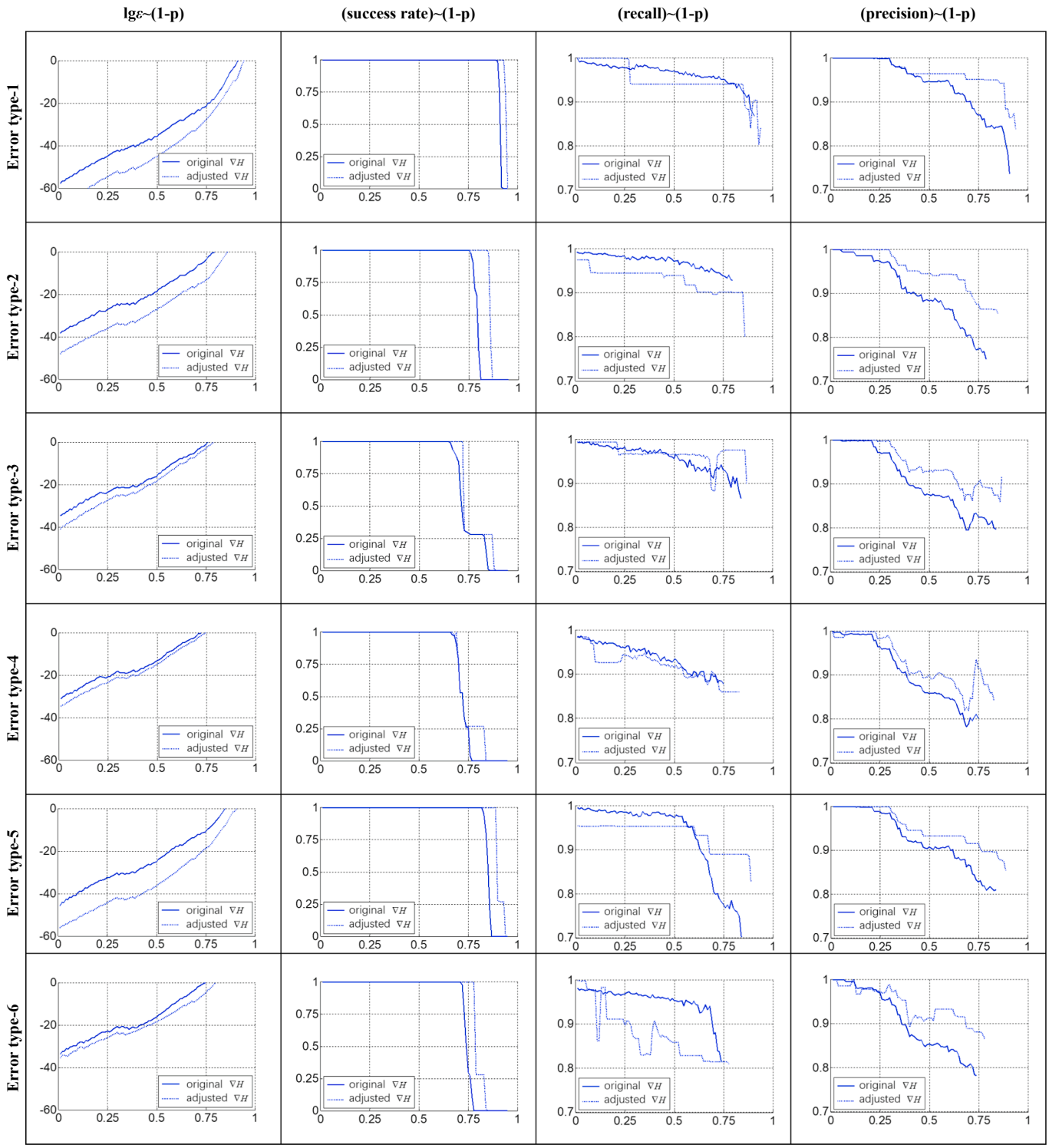

Fig. 9. The simulated experimental results. In each figure, the solid line indicates the result of ORSA-SAT with fixed elevation uncertainty ∇H, and the dotted line indicates the result with ∇H adjusted. In all the figures, the x-axis is (1-p), i.e., the rate of mismatches in the matching points. The figures in the first column show the curve of mean lg$\varepsilon(\alpha,n,k)$ and (1-p). The figures in the second column show the curve of success rate and (1-p). The figures in the third column show the curve of mean recall and (1-p). The figures in the fourth column show the curve of mean precision and (1-p). The figures in each row correspond to one type of simulated error.

because there was a ten-year-interval between the acquisition dates and the difference in the acquisition seasons. In group-5, an Ikonos-2 image was matched on a Landsat-8 panchromatic image. The Landsat-8 image was downloaded from the website of the U.S. Geological Survey (landsat.usgs.gov). Since there was a huge gap in the ground resolution, the Ikonos-2 image was down-sampled to 15 m GSD before matching.

The key-points were detected by the Harris corner feature detector and matched by the NCC-based matching method with an 11*11 matching window. With each image pair, the Harris key-points were detected from the first image and matched on the second image. The search region was computed with the initial RFMs and a 90 m-resolution SRTM-DEM while the elevation uncertainty ∇H was set at 30 m. The NCC-threshold δ_{NCC} was set at 0.8 and all the matching points having larger-than-δ_{NCC} NCC values remained in set $\mathcal{M}$ of the one-on-

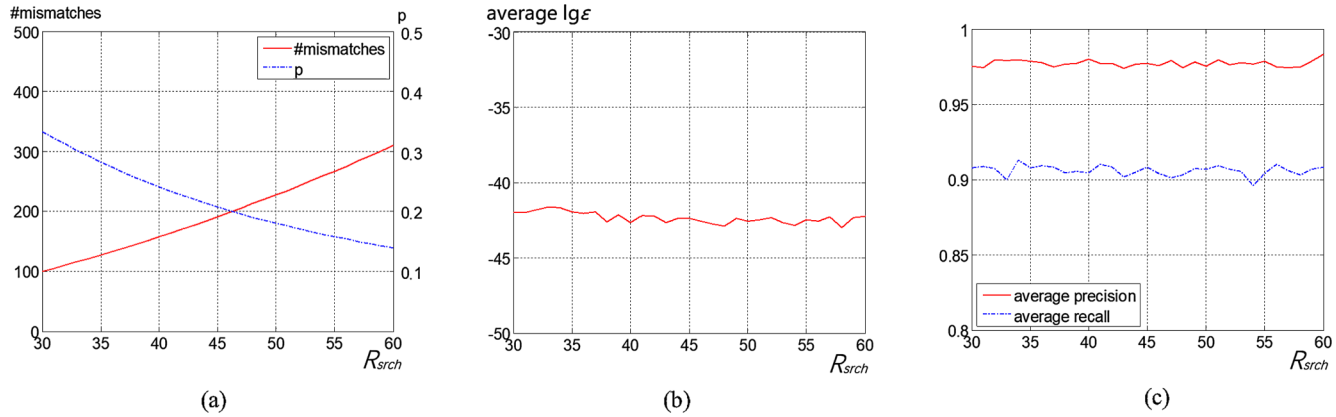

Fig. 10. The results of the ORSA-SAT with different search regions. (a) the (number of mismatch) ~ R_{srch} curve and the p ~ R_{srch} curve; (b) the (average lg$\varepsilon(\alpha,n,k)$) ~ R_{srch} curve; (c) the (average precision) ~ R_{srch} curve and the (average recall) ~ R_{srch} curve.

multiple point-matches.

After obtaining the automatic point-matches, they were validated artificially to obtain true matches. The point-matches, which were hard to recognize by the human eye or had a larger-than-2-pixel error were judged as invalid matches. In addition, all the point-matches lying on the sand dunes in group-1 or lying on the snowlines of the mountains in group-2 were judged invalid since it was not possible to judge how far the sand dunes or the snowlines moved during the time intervals between the acquisition dates. All the point-matches that were close to the shadows of trees or buildings in group-2 and -3 were judged false matches since the images were acquired in different season and the solar attitudes were so different. The numbers and proportions of true matches are listed in Table 4.

Attempts then were made to detect true matches using both the P2L algorithm and the ORSA-SAT. When using the P2L algorithm, the geometric accuracy threshold δ was fixed at three pixels. When using ORSA-SAT, two strategies were taken: the first strategy was to pick up the most similar matches from set $\mathcal{M}$ to populate set $\mathcal{S}$ of one-on-one matches, thus the geometric rigidity was computed with (11); the second strategy was to detect the rigid set of point-matches directly with set $\mathcal{M}$, thus the geometric rigidity of the set of point-matches was computed with (21). The experiments were conducted in a laptop having an Intel Core-i7-7700HQ CPU and 8 GB internal memory. No parallel computing strategy was used when implementing both the algorithms.

Finally, the sets of inliers obtained by the three automatic strategies were compared with the artificially obtained true matches so that the precision and the recall rate could be calculated to evaluate the three strategies.

5.2. Results of the image-pairs on unstable landforms

Table 5 lists the mismatch detection results of the image-pairs in group-1 and -2, which covered unstable landforms. In group-1, the results of all the three strategies had poor precisions, i.e., from 39.9% to 42.2%. In group-2, the precisions were also not good, i.e., from 61.1% to 69.2%. In both groups, the recalls of the results of the ORSA-SAT algorithm were better than those of the P2L method.

The reason for the low precisions is that the matching results of group-1 and -2 included three types of point-matches, the corresponding points on permanent objects (see Figs. 11(a) and 12(a)) which were artificially judged true matches, the corresponding points on movable objects (see Figs. 11(b) and 12(b)) which were artificially judged false matches, and mismatches. The second type of matches, although artificially judged false matches, did not have a random and uniform distribution. Therefore, the existence of both the first and the second types of matches can prove a *contrario* against the null-hypothesis $\mathcal{H}_0$ and the *a-contrario* based method cannot filter the second type of matches.

Table 4
The results of the automatic image match.

Group #	1	2	3	4	5
# all matches	488	570	347	792	114
# true matches	93	137	60	103	107
% true matches	19.1%	24.3%	17.3%	13.0%	93.9%

5.3. Results of the image-pairs on stable landforms

Table 6 lists the mismatch detection results of the image-pairs in group-3, -4, and -5, which covered stable landforms. To evaluate the running time and stability of both the algorithms, we processed with each strategy and each group for 100 times and calculated the maximum, the minimum, and the mean values of the precision, recall, number of iterations, and time cost.

Table 3
The information of the satellite images used in the experiments.

#Group	Satellite	GSD (m)	Topography and landcovers	Acquisition date	location
1	Ikonos-2	1.0	desert	2007-11-07	83.9E 38.1N
	Ikonos-2	1.0	desert	2011-04-01	83.9E 38.1N
2	Ikonos-2	1.0	mountain, rocks, glaciers	2009-11-09	83.9E 29.3N
	Ikonos-2	1.0	mountain, rocks, glaciers	2007-06-20	83.8E 29.3N
3	Ikonos-2	1.0	plain, farmland, forest, village	2004-09-20	127.5E 50.2N
	Ikonos-2	1.0	plain, forest, ice, snow, village	2003-12-09	127.4E 50.4N
4	Ziyuan-3 (ref)	2.5	plain, farmland, village	2013-06-13	110.6E 35.0N
	Ikonos-2	1.0	plain, farmland, village	2003-01-04	110.4E 34.7N
5	Landsat-8 (ref)	15.0	mountains, glaciers	2014-05-23	98.3E 30.3N
	Ikonos-2	1.0	mountains, glaciers	2009-04-07	98.5E 30.5N

Table 5
Mismatch detection results of Group-1 and −2.

Group #	1			2		
Algorithm	P2L	ORSA ($\mathcal{S}$)	ORSA ($\mathcal{M}$)	P2L	ORSA ($\mathcal{S}$)	ORSA ($\mathcal{M}$)
Precision	39.9%	42.2%	40.4%	64.1%	69.2%	61.1%
Recall	65.6%	87.1%	90.3%	61.3%	73.7%	82.5%
NFA ($\lg \bar{\varepsilon}$)	–	−95.7	−75.9	–	−81.7	−80.0
Max P2L dist.[1]	2.96	6.09	7.26	2.95	3.40	5.20

Note: 1. The unit of maximum P2L distance is the pixel of the right image.

In group-3, the precisions of the ORSA-SAT algorithm were obviously higher than those of the P2L algorithm, while the recalls were lower. With both set $\mathcal{S}$ and set $\mathcal{M}$, the precisions of the ORSA-SAT algorithm were about 80–87% and the recalls were about 80–88%. Meanwhile, for the P2L algorithm, the precisions were much lower at about 54–64% and the recalls were higher at about 86–98%. The P2L algorithm needed less running time than the ORSA-SAT algorithm. The running time of the P2L algorithm ranged from 2.39 s to 3.51 s, while that of the ORSA-SAT algorithm ranged from about 2.48 s to about 26.34 s.

In group-4, the precisions of the ORSA-SAT algorithm were still higher than those of the P2L algorithm, while the recalls were still lower. The ORSA-SAT algorithm with set $\mathcal{M}$ had slightly higher precisions than with set $\mathcal{S}$, i.e., they had an about 0.5% difference in the mean of precisions. Meanwhile, for the P2L algorithm, the precisions were about 78–87% and the recalls were about 86–98%. The P2L algorithm had no significant advantage in running time, but its numbers of iterations were much less than the ORSA-SAT algorithm. The fastest case of the ORSA-SAT algorithm cost only 26.43 s running time while the slowest cost 421.88 s.

In group-5, the precisions of the three strategies were all very high at about 96–100%. However, the recalls of the P2L algorithm (about 38–43%) were very poor comparing those of the ORSA-SAT algorithm (about 84–97%). Therefore, the numbers of iterations of the P2L algorithm were 57–80, much larger than those of the ORSA-SAT algorithm.

The maximum P2L distances of the inliers obtained by the ORSA-SAT algorithm can be treated as its adaptively set threshold, the means of which were 1.07 pixels and 0.97 pixel in group-3, 2.23 pixels and 2.30 pixels in group-4, and 18.52 pixels and 19.17 pixels in group-5. According to the GSD ratios between the image-pairs in Table 3, we can find that in all the three groups, the average maximum P2L distances of the inliers obtained by the ORSA-SAT algorithm were very close to one left-image-pixel. These results demonstrate that the a-*contrario* based method can accurately find out the scale of the a priori positional error of true matches and effectively pick them out.

5.4. Analysis

The results of group-1 and -2 in Section 5.2 show that the ORSA-SAT algorithm cannot well deal with the images covering unstable landforms. The corresponding points on movable objects, having 1–5 pixels positional errors, cannot be filtered by the *a-contrario* method. Thus, when dealing with these images, the P2L algorithm with a strict threshold may get more precise results.

The results of group-3, -4, and -5 in Section 5.3 demonstrate that when dealing with images covering stable landforms, the ORSA-SAT algorithm, having no preset thresholds, can effectively filter mismatches without knowing the apriori positional accuracy of true matches. In each group, the ORSA-SAT algorithm with both set $\mathcal{S}$ and set $\mathcal{M}$ had above 80% precisions in the 100 times of executions, which demonstrate the good stability of the algorithm. The results also show that there is no obvious difference between using one-on-one matches or one-on-multiple matches with the ORSA-SAT algorithm.

6. Summary

In this paper, we proposed an *a-contrario* method named ORSA-SAT to remove mismatches by finding the most meaningful subset from the set of point-matches of two-view satellite images. On the basis of the ELS constraining model introduced by Wan and Zhang (2017), the new method evaluates the meaningfulness of a set of point-matches by estimating the NFA under a certain null-hypothesis with the formula $\varepsilon(\alpha,n,k)$. When $\varepsilon(\alpha,n,k) < 1$ is satisfied with the most meaningful subset, the ORSA-SAT will judge the result to be reliable, otherwise, the

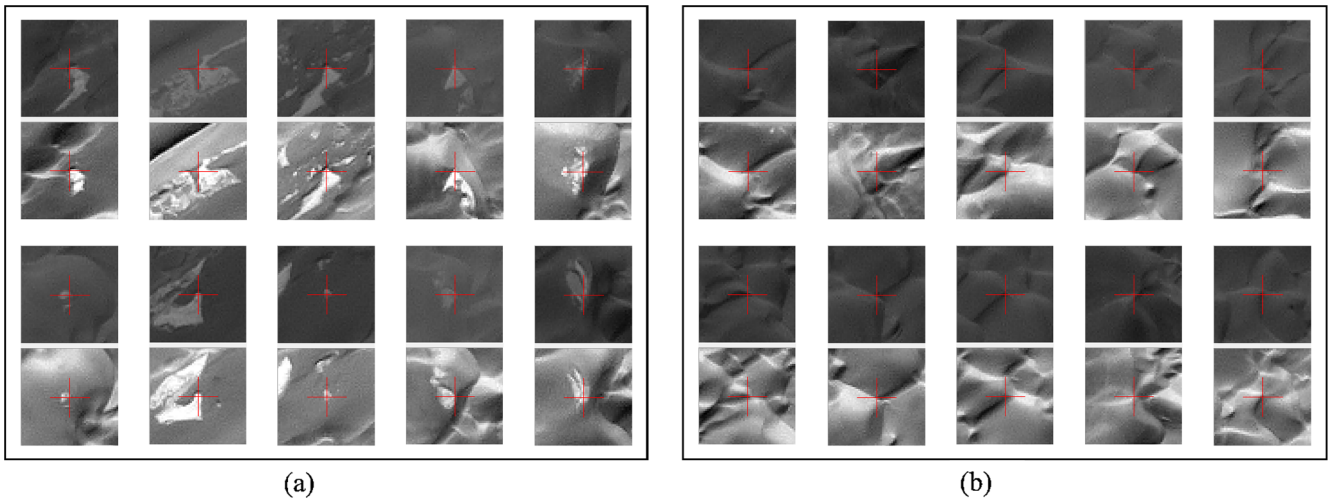

Fig. 11. Screenshots of inliers judged by ORSA-SAT($\mathcal{S}$) in group-1. The point matches in (a) were judged true artificially because they were on the stable landmarks, while the point matches in (b) were judged false artificially because they were on the sand dunes which are movable.

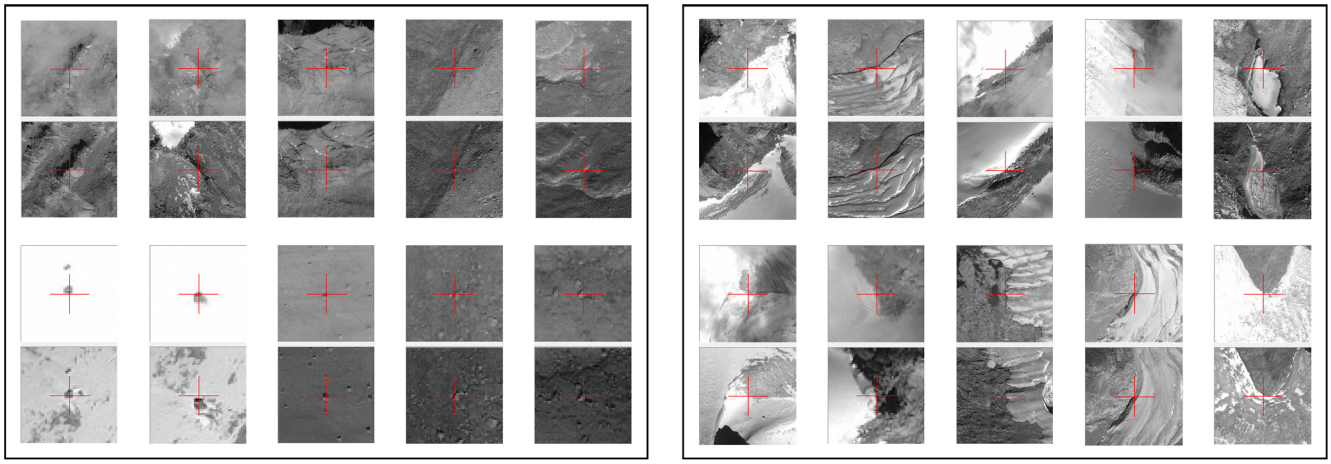

Fig. 12. Screenshots of inliers judged by ORSA-SAT(S) in group-2. The point matches in (a) were judged true artificially because they were on the stable landmarks such as rocks, while the point matches in (b) were judged false artificially because they were on the snowlines or glaciers which are movable.

Table 6
Mismatch detection results of Group-3, -4 and -5.

Group #		3			4			5		
Algorithm		P2L	ORSA (S)	ORSA ($\mathcal{M}$)	P2L	ORSA (S)	ORSA ($\mathcal{M}$)	P2L	ORSA (S)	ORSA ($\mathcal{M}$)
Precision	min	54.74%	**80.65%**	80.00%	78.16%	82.98%	**88.00%**	**97.62%**	97.06%	96.15%
	max	64.13%	**86.44%**	86.21%	87.23%	91.55%	**91.78%**	**100%**	99.00%	98.97%
	mean	59.48%	82.52%	**84.11%**	82.15%	89.63%	**90.15%**	97.88%	**97.98%**	97.38%
Recall	min	**86.67%**	83.33%	80.00%	**63.11%**	57.86%	55.34%	38.32%	84.11%	**85.98%**
	max	**98.33%**	88.33%	85.00%	**79.61%**	65.05%	65.05%	42.99%	**97.20%**	95.33%
	mean	**95.00%**	85.17%	82.83%	**72.52%**	60.78%	61.65%	40.93%	**92.48%**	91.92%
#Iterations	min	166	1085	1135	2127	7949	8697	57	5	5
	max	214	2387	2230	3396	11,909	12,770	80	5	5
	mean	184	1581	1730	2617	9957	10,644	64	5	5
Running time (s)	min	2.39	2.48	2.93	103.61	26.45	62.13	0.05	0.01	0.01
	max	3.51	18.76	26.34	153.8	168.15	421.88	0.09	0.03	0.03
	mean	2.89	11.61	8.02	127.36	44.79	109.7	0.07	0.01	0.02
Average NFA ($\lg \bar{\varepsilon}$)		–	−38	−38.6	–	−19.3	−17.3	–	−151.6	−145
Average Max P2L dist.[1]		2.95	1.07	0.97	2.96	2.23	2.30	2.95	18.52	19.17

Note: 1. The unit of maximum Point-to-ELS distance is the pixel of the right image.

ORSA-SAT will judge the set unreliable and will output an empty set to avoid the unreliable matches being involved in the geometric corrections or block adjustments.

The experimental results in Sections 4.2 and 5.3 have demonstrated the effect and the stability of the criterion $\varepsilon(a,n,k) < 1$ in justifying the correctness of the mismatch detection results because the precision kept above 80% with both simulated and real data. Also, the recalls kept above 80% with simulated data and the maximum P2L distances of inliers were close to one-pixel of the lower-resolution image with the real data. In Section 4.3, the ORSA-SAT algorithm was also tested with varying relative orientation error. It was shown that the precision, the recall, and even the meaningfulness measurement was almost not influenced by the relative orientation accuracy when the ratio between the number of mismatches and the area of search region is unchanged. The experimental results in Section 5.2 showed that the ORSA-SAT algorithm cannot filter the less accurate true matches on movable objects.

Combining the experimental results and analysis in this paper and in Wan and Zhang (2017), we have the following suggestion for choosing an appropriate mismatch detection algorithm:

(1) When mismatches do not obey the null-hypothesis $\mathcal{H}_0$, the *a-contrario* judgment with criterion $\varepsilon(a,n,k) < 1$ is not reliable and the ORSA-SAT algorithm should not be used. This condition happens when the images cover unstable landforms like deserts and glaciers or when some special constraints which can change the distribution of mismatches are used in image matching.
(2) The ORSA-SAT algorithm is a better choice when matching with DOMs for GCPs or when the geometric distortion is not clear or when there are too many images and the users want to ensure the correctness of the point-matches without checking them artificially.
(3) The P2L algorithm is a better choice when users believe that there is no local deformation or nonlinear distortion and want to ensure the residuals of inliers under a certain threshold.

The two major advantages of the ORSA-SAT algorithm, i.e., the automatic result validation and the self-adaptive threshold, make the mismatch detection software easier to use for less-professional users and save a lot of time spent on checking the results. However, the ORSA-SAT algorithm still uses the ELS-constraint model, which means it cannot filter the mismatches of which the error directions are close to the epipolar lines. This kind of mismatches can only be removed in the block adjustment using the multi-view geometric constraint.

Acknowledgements

The authors are grateful to Environmental Systems Research Institute for testing the approach with the IKONOS-2 and GeoEye-1

images. This work was supported by the National Key Research and Development Program of China, Grant No. 2018YFB0505003, and National Natural Science Foundation of China with project number 41871368 and 41571434.

References

Almansa, A., Desolneux, A., Vamech, S., 2003. Vanishing point detection without any a priori information. IEEE Trans. Pattern Anal. Mach. Intelligence 25, 502–507.

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Speeded-up robust features (SURF). Computer Vision Image Understand. 110, 346–359.

Cao, F., Musé, P., Sur, F., 2005. Extracting meaningful curves from images. J. Math. Imaging Vision 22, 159–181.

Chen, H.-M., Varshney, P.K., Arora, M.K., 2003. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. Geosci. Remote Sensing 41, 2445–2454.

Chen, Q.-S., Defrise, M., Deconinck, F., 1994. Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. IEEE Trans. Pattern Anal. Mach. Intelligence 12, 1156–1168.

Chum, O., Matas, J., 2005. Matching with PROSAC-progressive sample consensus. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, San Diego, CA, USA, pp. 220–226.

Dellinger, F., Delon, J., Gousseau, Y., Michel, J., Tupin, F., 2015. SAR-SIFT: a SIFT-like algorithm for SAR images. IEEE Trans. Geosci. Remote Sensing 53, 453–466.

Delon, J., Desolneux, A., Lisani, J., Petro, A., 2004. Histogram analysis and its applications to fast camera stabilization. In: International Workshop on Systems, Signals and Image Processing, Ambient Multimedia, pp. 431–434.

Delon, J., Desolneux, A., Lisani, J.L., Petro, A.B., 2005. Automatic color palette. In: IEEE International Conference on Image Processing, 2005, Genoa, Italy, II-706.

Desolneux, A., Moisan, L., Morel, J.-M., 2000. Meaningful alignments. Int. J. Computer Vision 40, 7–23.

Desolneux, A., Moisan, L., Morel, J.-M., 2001. Edge detection by Helmholtz principle. J. Math. Imaging Vision 14, 271–284.

Desolneux, A., Moisan, L., Morel, J.-M., 2007. From Gestalt Theory to Image Analysis: A Probabilistic Approach, vol. 34 Springer Science & Business Media.

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.

Fan, J.W., Wu, Y., Li, M., Liang, W.K., Cao, Y.C., 2018. SAR and optical image registration using nonlinear diffusion and phase congruency structural descriptor. IEEE Trans. Geosci. Remote Sensing 56, 5368–5379.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.

Foroosh, H., Zerubia, J.B., Berthod, M., 2002. Extension of phase correlation to subpixel registration. IEEE Trans. Image Process. 11, 188–200.

Fraser, C.S., Dial, G., Grodecki, J., 2006. Sensor orientation via RPCs. ISPRS J. Photogrammetry Remote Sensing 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.

Grodechi, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogrammetric Eng. Remote Sensing 69, 59–68.

Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. Photogrammetric Eng. Remote Sensing 69, 59–68.

Kasser, M., Egels, Y., 2002. Digital Photogrammetry. Taylor & Francis, London.

Ke, Y., Sukthankar, R., 2004. PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. 506–513.

Li, J., Hu, Q., Ai, M., Zhong, R., 2017. Robust feature matching via support-line voting and affine-invariant ratios. ISPRS J. Photogrammetry Remote Sensing 132, 61–76.

Ling, X., Zhang, Y.J., Xiong, J.X., Huang, X., Chen, Z.P., 2016. An image matching algorithm integrating global SRTM and image segmentation for multi-source satellite imagery. Remote Sensing 8, 672.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Computer Vision 60, 91–110.

Ma, J., Qiu, W., Zhao, J., Ma, Y., Yuille, A.L., Tu, Z., 2015a. Robust L2E estimation of transformation for non-rigid registration. IEEE Trans. Signal Process. 63, 1115–1129.

Ma, J., Zhao, J., Tian, J., Yuille, A.L., Tu, Z., 2014. Robust point matching via vector field consensus. IEEE Trans. Image Process. 23, 1706–1721.

Ma, J., Zhou, H., Zhao, J., Gao, Y., Jiang, J., Tian, J., 2015b. Robust feature matching for remote sensing image registration via locally linear transforming. IEEE Trans. Geosci. Remote Sensing 53, 6469–6481.

Moisan, L., Moulon, P., Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. Image Processing On Line 2, 56–73.

Moisan, L., Stival, B., 2004. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. Int. J. Computer Vision 57, 201–218.

Morel, J.M., Yu, G., 2009. ASIFT: a new framework for fully affine invariant image comparison. Siam J. Imaging Sci. 2, 438–469.

Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.-M., 2013. USAC: a universal framework for random sample consensus. IEEE Trans. Pattern Anal. Mach. Intelligence 35, 2022–2038.

Sedaghat, A., Ebadi, H., 2015. Remote sensing image matching based on adaptive binning SIFT descriptor. IEEE Trans. Geosci. Remote Sensing 53, 5283–5293.

Sedaghat, A., Mohammadi, N., 2018a. High-resolution image registration based on improved SURF detector and localized GTM. Int. J. Remote Sensing 1–26.

Sedaghat, A., Mohammadi, N., 2018b. Uniform competency-based local feature extraction for remote sensing images. ISPRS J. Photogrammetry Remote Sensing 135, 142–157.

Sedaghat, A., Naeini, A.A., 2018. DEM orientation based on local feature correspondence with global DEMs. GIScience Remote Sensing 55, 110–129.

Sheng, Y., 2005. Theoretical analysis of the iterative photogrammetric method to determining ground coordinates from photo coordinates and a DEM. Photogrammetric Eng. Remote Sensing 71, 863–871.

Sheng, Y., 2008. Modeling algorithm-induced errors in iterative mono-plotting process. Photogrammetric Eng. Remote Sensing 74, 1529–1537.

Tao, C.V., Hu, Y., 2001. A comprehensive study of the rational function model for photogrammetric processing. Photogrammetric Eng. Remote Sensing 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. IEEE Trans. Geosci. Remote Sensing 48, 1907–1918.

Torr, P.H., Zisserman, A., 2000. MLESAC: a new robust estimator with application to estimating image geometry. Computer Vision Image Understand. 78, 138–156.

Torresani, L., Kolmogorov, V., Rother, C., 2008. Feature correspondence via graph matching: models and global optimization. In: European Conference on Computer Vision. Springer, Berlin, Heidelberg, pp. 596–609.

Toutin, T., 2003. Block bundle adjustment of Landsat 7 ETM+ images over mountainous areas. Photogrammetric Eng. Remote Sensing 69, 1341–1349.

Toutin, T., 2004a. Review article: Geometric processing of remote sensing images: models, algorithms and methods. Int. J. Remote Sensing 25, 1893–1924.

Toutin, T., 2004b. Spatiotriangulation with multisensor VIR/SAR images. IEEE Trans. Geosci. Remote Sensing 42, 2096–2103.

von Gioi, R.G., Jakubowicz, J., Morel, J.-M., Randall, G., 2010. LSD: a fast line segment detector with a false detection control. IEEE Trans. Pattern Anal. Mach. Intelligence 32, 722–732.

Wan, Y., Zhang, Y., 2017. The P2L method of mismatch detection for push broom high-resolution satellite images. ISPRS J. Photogrammetry Remote Sensing 130, 317–328.

Xia, G.-S., Delon, J., Gousseau, Y., 2014. Accurate junction detection and characterization in natural images. Int. J. Computer Vision 106, 31–56.

Xiong, J., Zhang, Y., Zheng, M., Ye, Y., 2013. An SRTM assisted image matching algorithm for long-strip satellite imagery. J. Remote Sensing 17, 1103–1117.

Xiong, Z., Zhang, Y., 2009. A generic method for RPC refinement using ground control information. Photogrammetric Eng. Remote Sensing 75, 1083–1092.

Xue, N., Xia, G.-S., Bai, X., Zhang, L., Shen, W., 2018. Anisotropic-scale junction detection and matching for indoor images. IEEE Trans. Image Process. 27, 78–91.

Ye, Y.X., Shan, J., Hao, S.Y., Bruzzone, L., Qin, Y., 2018. A local phase based invariant feature for remote sensing image matching. ISPRS J. Photogrammetry Remote Sensing 142, 205–221.

Zhang, Y.J., Wan, Y., Huang, X.H., Ling, X., 2016. DEM-assisted RFM block adjustment of pushbroom nadir viewing HRS imagery. IEEE Trans. Geosci. Remote Sensing 54, 1025–1034.

Zhao, M., An, B., Wu, Y., Lin, C., 2013. Bi-SOGC: a graph matching approach based on bilateral KNN spatial orders around geometric centers for remote sensing image registration. IEEE Geosci. Remote Sensing Lett. 10, 1429–1433.

Zheng, M.T., Zhang, Y.J., 2016. DEM-aided bundle adjustment with multisource satellite imagery: ZY-3 and GF-1 in large areas. IEEE Geosci. Remote Sensing Lett. 13, 880–884.

Zhou, H., Ma, J., Yang, C., Sun, S., Liu, R., Zhao, J., 2016. Nonrigid feature matching for remote sensing images via probabilistic inference with global and local regularizations. IEEE Geosci. Remote Sensing Lett. 13, 374–378.