International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tres20

Concatenating wide-parallax satellite orthoimages for simplified regional mapping via utilizing line-point consistency

Ziming Tu, Xiubin Yang, Zongqiang Fu, Suining Gao, Ge Yang, Li Jiang, Mo Wu & Shaoen Wang

To cite this article: Ziming Tu, Xiubin Yang, Zongqiang Fu, Suining Gao, Ge Yang, Li Jiang, Mo Wu & Shaoen Wang (2023) Concatenating wide-parallax satellite orthoimages for simplified regional mapping via utilizing line-point consistency, International Journal of Remote Sensing, 44:16, 4857-4882, DOI: 10.1080/01431161.2023.2240033

To link to this article: https://doi.org/10.1080/01431161.2023.2240033

INTERNATIONAL JOURNAL OF REMOTE SENSING
2023, VOL. 44, NO. 16, 4857–4882
https://doi.org/10.1080/01431161.2023.2240033

Taylor & Francis
Taylor & Francis Group

Concatenating wide-parallax satellite orthoimages for simplified regional mapping via utilizing line-point consistency

Ziming Tu[a,b,c], Xiubin Yang[a,c], Zongqiang Fu[a,b,c], Suining Gao[a,b,c], Ge Yang[d,e], Li Jiang[f], Mo Wu[a,b,c] and Shaoen Wang[a,b,c]

[a]Changchun Institute of Optics, Fine mechanics and Physics, Chinese Academy of Sciences, Changchun, China; [b]Daheng College, University of Chinese Academy of Sciences, Beijing, China; [c]Key Laboratory of Space-Based Dynamics and Rapid Optical Imaging Technology, Chinese Academy of Sciences, Changchun, China; [d]Zhuhai Orbita Aerospace Science & Technology Co. LTD, Zhuhai, China; [e]Department of Mechanics, Tianjin University, Tianjin, China; [f]Physics Department, Changchun University of Science and Technology, Changchun, China

ABSTRACT

Block adjustment is a common way for achieving large-scale mapping from several ortho-rectified satellite images derived from wide-parallax pushbroom (WPP) imaging. However, block adjustment's accuracy can be affected by the amount of ground control points (GCPs), resolution of digital evaluation model (DEM) and unexpected and unpredictable imaging drift. In addition, block adjustment requires sensor model, which is hard to obtain and needs adjustment in real time. In view of these problems, this paper proposes a WPP blocks concatenation method for simplified regional mapping. First, considering distortions along different blocks, an adaptive-RPC model is proposed to ortho-rectified images. Then, a novel algorithm for stitching the ortho-rectified images is proposed, which consists of line-point consistency for matching neighbourhoods between two images, and a new energy function for suppressing artefact and distortions. Stitching experiments carried out on images from Orbita Satellite show that the stitching accuracy reaches 0.87 pixels, better than state-of-the-art methods, and absolute cumulative accuracy after stitching 10 orthoimages is only 3.99% higher than original one. Further simulation experiments with variable in satellite swing angles verify its robustness under complex scenes, while the estimation of cumulative error reiterates the necessity of performing orthorectification in prior. All results demonstrate that the proposed method can obtain accurate large regional mapping with distortions and artefact effectively suppressed.

ARTICLE HISTORY
Received 28 February 2023
Accepted 12 July 2023

KEYWORDS
Wide-parallax satellite image stitching; line-point consistency; energy function; orthorectification; simplified imaging

1. Introduction

Earth observation by remote sensing (RS) is an important way to obtain geospatial information, as is the fundamental to target recognition, comparisons and comprehensive analyses. However, due to the limited size of images, the above studies require more than one image

CONTACT Xiubin Yang yangxiubin@ciomp.ac.cn Changchun Institute of Optics, Fine mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

with partial overlap to cover the full study region. Block adjustment based on sensor model is a common method for achieving accurate regional mapping, mainly taking satellite attitude, camera installation matrix and trajectory parameters into account. This method has high accuracy, but the adjustment is complex and the generality is poor (Toutin 2003).

Nowadays, mainstream methods for block adjustment are mainly based on calculating rational polynomial coefficients (RPCs) (Fraser and Hanley 2005; Grodecki and Dial 2003; Wang et al. 2018; Zhang et al. 2015, 2016), which is a series of polynomial ratio defining the relationship between the image space and object space. It is easier to obtain the accurate image once it is downloaded from satellites and ortho-rectified, since RPCs are generated together with imagery on payloads (Tao and Hu 2001). Although the mentioned block adjustment methods are computationally practical and evolutionally efficient, there still exists two problems. First, block adjustment relies on the regular operation of sensor, which means a slight change in imaging model can probably lead to failure. Second, the accuracy of the mentioned block adjustment methods is usually affected by the number of GCPs and precision of DEM (Mayaux et al. 2002; Miyoshi et al. 2018). In view of this, a simplified yet effective stitching method is in need for performing wide-parallax image stitching to achieve larger regional mapping.

While block adjustment is still a widely deployed approach within photogrammetry community (Wang et al. 2018; Zhang et al. 2022; Zhou and Liu 2022), image stitching can provide an alternative solution for some scenarios (Zhu et al. 2023). It refers to the process of concatenating multiple images with overlapping areas into a composite one (Megha and Rajkumar 2021). However, it's challenging to stitch RS images due to the wide-parallax structures. Traditional single-feature–based matching methods often fail to generate high-accurate image, since the geometric projection cannot assure full alignment within the overlapping areas (Meng et al. 2013; Adel, Elmogy, and Elbakry 2015). An improvement is to introduce global homography (Gao et al. 2011), which is considered as a key to eliminating artefact, ghosting and distortions in overlapping areas (Brown and Lowe 2007). For example, divide RS image scenes into far-plane and ground-plane based on global homography (Xu et al. 2020). Since most point-based methods are not enough for calculating exact homography (Lowe 2004; Rublee et al. 2011), it is advisable to introduce line-pairs and line-point constraints (Jia et al. 2016, 2021). Regarding this, Zhang et al. constructed a line-band-descriptor (LBD) based on line-pairs' local characteristics, which has a good performance on high-texture differences images (Zhang and Koch 2013). Li et al. proposed a semi-automatic guidance strategy for generating seamlines on aerial photography using human-computer interaction constraints (Li et al. 2016). Improved studies focus on suppressing artefact and distortions in overlapping areas and deformation in non-overlapping ones. For example, As-Projective-As-Possible (APAP) proposed by Zaragoza et al. succeeded in controlling artefact in overlapping regions with local constraints to control warping but failed to control that in non-overlapping ones (Zaragoza et al. 2013). Liao and Li put forward an image-splicing method called Single-Perspective-Warps (SPW), assuming that wide-parallax stitching image is drawn from a main perspective (Liao and Nan 2020). It proved generally effective but sometimes invalid when images are warped due to unfixed viewpoints. More recent studies (Fan et al. 2022; Ye et al. 2022; Zhu et al. 2023) focus on proposing new feature descriptor, namely MABPC, SFOC, MALG, and CFOG (Ye et al. 2019), for more effective matching because traditional feature descriptors (e.g. SIFT, ASIFT, SURF) do not simultaneously consider the robustness of the nonlinear intensity and geometry variations

between different multi-modal RS images (Huang et al. 2014). However, most RS stitching models take narrow-swath RS (less than 2 km) images as example (Cui et al. 2021; Hui and Chen 2020; Xie et al. 2019; Xu et al. 2020), which means they may lead to high cumulative stitching errors within wide-parallax pushbroom (WPP) images. These existing models lack generality and accuracy. Except for the mentioned feature-based methods, the object-space oriented ones usually own high-precision but lack generality and portability (Jiang et al. 2017; Liu et al. 2022; Yeh and Tsai 2011). They are most likely to fail when the sensor model is unknown or time-variant, so they are beyond this study's interest. In summary, the mentioned methods above are not strictly applicable for stitching WPP blocks.

Main challenges of stitching WPP blocks include unknown sensor parameters (camera pose and trajectory), the severe geometric deformation along track and uncertain accumulative stitching errors. To cope with that, an orthorectification-based stitching method to achieve large regional mapping from a few WPP blocks via utilizing line-point consistency is proposed. Our method is inspired by several state-of-the-art algorithms (Jia et al. 2016, 2021; Liao and Nan 2020; Lin et al. 2015). It mainly consists of three parts. First, an adaptive-RPC model with GCPs and DEM is established to restore its original geometric feature. Without such step, there would be significant geometric distortion within each region, which would affect ground sampling distance (GSD) and cumulative stitching accuracy severely. Second, a novel principle, line-point consistency, is adopted to build neighbourhoods within each image where more line-pairs and point-pairs can be aligned. This process proves effective since more aligned feature pairs between RS images can lead to better stitching result than method with less aligned feature pairs (Fan et al. 2010). Third, a new energy function is designed to control serious distortion during warping, in which the weight function across the whole image is re-established and the connection between the projection of different regions is reinforced. Such improvements prove necessary since there are large overlaps between two WPP images, while the linear energy function in mentioned algorithms leads to grey value discontinuity near the edges. Compared with other RS stitching methods, our method can better utilize spatial information. For example, prior overlap based RS stitching model (Cui et al. 2021) only focuses on calculating the similarity (homography in our manuscript) of overlaps between two images, ignoring the extended deformation near edges. In contrast, a new energy function is designed to reinforce the homography in our method. Another example, UAV RS stitching model (Hui and Chen 2020) only focuses on point-alignment during stitching. It can be regarded as an update from the traditional common algorithm SIFT and ASIFT (Lowe 2004). Such single-point alignment proves not sufficient in RS images matching, thus highlighting the necessity of incorporating both point and line alignment. In summary, compared to state-of-the-art algorithms, ours introduces more conditional constraints, provides more evaluation metrics and proves more applicable for stitching WPP blocks. Main contributions of the proposed method can be summarized as follows:

(1) Compared with other RS stitching methods, our method can better utilize spatial information. Single-feature (e.g. point, line, local similarity) detection and alignment proves not enough for RS stitching, thus highlighting the proposed method of utilizing both point and line alignment.
(2) To the best of our knowledge, it is one of the first to leverage dual feature constraints of lines and points into stitching wide-parallax orthoimages for regional

mapping (not common RS images). Differences between our approach and existing RS stitching algorithms are discussed. The results demonstrated its potential application prospects within remote sensing community.

(3) Simulation experiments verify its superior and robust application prospects in complex scenes, such as different viewpoints caused by changes in swing angles. And the accumulative experiments further verify its capacity to produce accurate results due to stable GSD.

The rest of this paper is organized as follows. Section 2 introduces the proposed method in detail. Section 3 presents the study area and experiments, describes experimental results and proposes further discussions. Section 4 presents the conclusions.

2. Methodology

Method proposed in this paper consists of two main parts. First, adapt the RPC file with DEM and GCPs to restore the ortho-rectified image and improve accuracy. Second, apply the proposed model into stitching mentioned ortho-rectified images, which mainly consists of two steps, matching neighbourhoods across image based on line-point consistency and constructing a three-part energy function to preserve the global naturalness across image. Figure 1 illustrates the process outlined in this article.

Given input RS images $I_i (i = 1, 2, \ldots, n)$, ortho-rectify I_i to obtain orthoimages $I_i^{ot} (i = 1, 2, \ldots, n)$ via orthorectification module denoted as f. Improve the accuracy of I_i^{ot} by adopting DEM of corresponding region and setting up virtual GCPs according to section 2.1, and obtain the final orthoimages $I_i^o (i = 1, 2, \ldots, n)$ via stitching module denoted as F.

$$\begin{aligned} I_i^{ot} &= f(I_i, RPCs)(i = 1, 2, \ldots, n) \\ I_i^o &= \underset{\Pi}{\overset{n}{F}}\left(I_i^{ot}, DEM_i, GCPs_i^n\right) \end{aligned} \tag{1}$$

Adopt line-point consistency to divide neighbourhoods $\Omega_j^i (j = 1, 2, \ldots, s)$ within each orthoimage I_i^o according to section 2.2.1. Within each neighbourhood, build projection invariant based on detected lines l_x^j and points P_y^j (within Ω_j^i in I_i^o) to match specific regions from two image blocks I_i^o and I_{i+1}^o.

$$\text{Match}: CN\left(\Omega_j^i\right) = CN\left(\Omega_j^{i+1}\right)(j = 1, \ldots, s; i = 1, \ldots, n) \tag{2}$$

While such match only offers an approximate match within global area, further constraints are in need to control serious line-warping, as in section 2.2.2. To solve this problem, an energy function E composed of three components E_c, E_h, and E_n, is proposed, each described in section 2.2.2. Mesh grid $\hat{V}$ within the warped image $\hat{I}_i^w$ according to optimized E proves more visually natural and more accurate in terms of precision.

$$\mathrm{E}\left(\hat{V}\right) = \mathrm{E}_c\left(\hat{V}\right) + E_h\left(\hat{V}\right) + E_n\left(\hat{V}\right) \quad \hat{V} = argminE\left(\hat{V}\right) \quad \hat{I}_i^w = \underset{\hat{V}}{\mathbb{F}}\left(I_i^o, I_{i+1}^o\right) \tag{3}$$

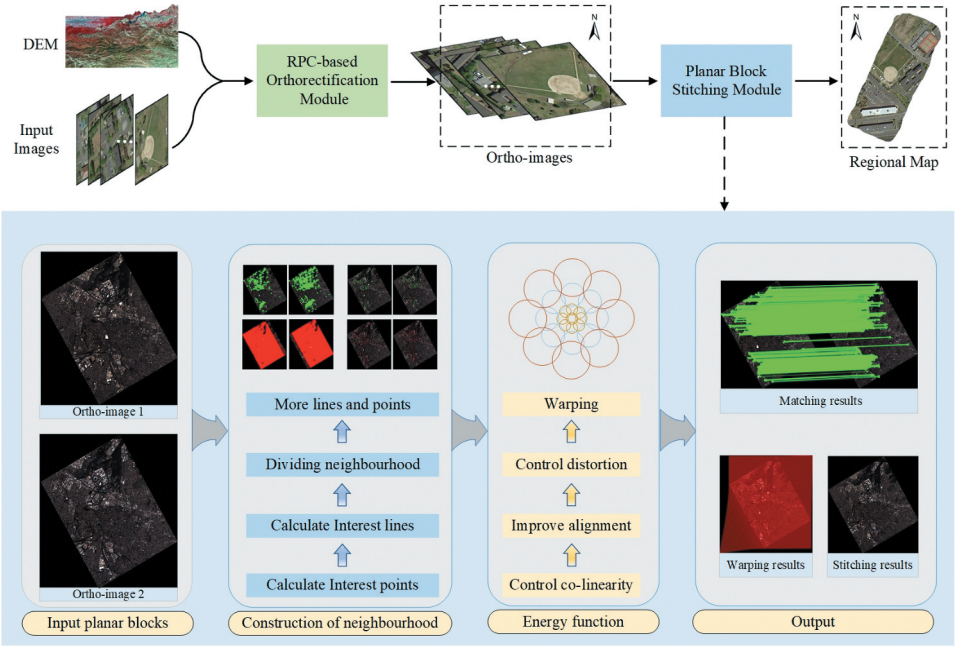


Figure 1. Flowchart of the proposed method. Specifically, planar block stitching module is illustrated in detail.

2.1. RPC-based orthorectification model

Orthorectification is a process that geo-corrects an image onto its upright map while restoring geometric structures. It is based on RPCs, which are generated together with payloads imaging. A WPP block without orthorectification has huge deformation across the whole image.

RPCs of High-Resolution Satellite Image (HRSI) is an important parameter for constructing rational function model (RFM), which sets up a connection between image coordinates and corresponding geographic ones without requiring sensor information. To ensure the calculation stability, RPCs regularize the image coordinates (I, s), latitude and longitude coordinates (B, L) and ellipsoid height H, so that the coordinates range between [−1,1]. Image regularized coordinates (I_n, s_n) corresponding to (I, s) and object regularized coordinates (U, V, W) corresponding to (B, L, H) are expressed as follows.

$$\begin{cases} \mathrm{I_n} = \frac{I - LineOff}{LineScale} \\ \mathrm{s_n} = \frac{s - SampleOff}{SampleScale} \end{cases} \tag{4}$$

$$\begin{cases} U = \frac{B - LonOff}{LonScale} \\ V = \frac{L - LatOff}{LatScale} \\ W = \frac{H - HeiOff}{HeiScale} \end{cases} \tag{5}$$

where *LineOff*, *SampleOff* are the offset values of image coordinates. *LineScale*, *SampleScale* indicates the scale value for image coordinates. (*LonOff*, *LatOff*, *HeiOff*) and

(*LonScale*, *LatScale*, *HeiScale*) represent the offset and scale values for object coordinates, respectively.

As shown in Figure 2, RPCs represent the coordinates of image points $P(l,s)$ as polynomial ratios with corresponding ground point's spatial coordinate $P^j_{(l,s)}\left(B^j_l, L^j_s, H^j\right)$ (where j denotes different elevation heights). RPCs do not require internal and external orientation elements. In order to enhance the stability of solution parameters, ground coordinates and image coordinates are normalized to [−1,1].

For each image, the relationship between the above two coordinates can be expressed as a polynomial ratio:

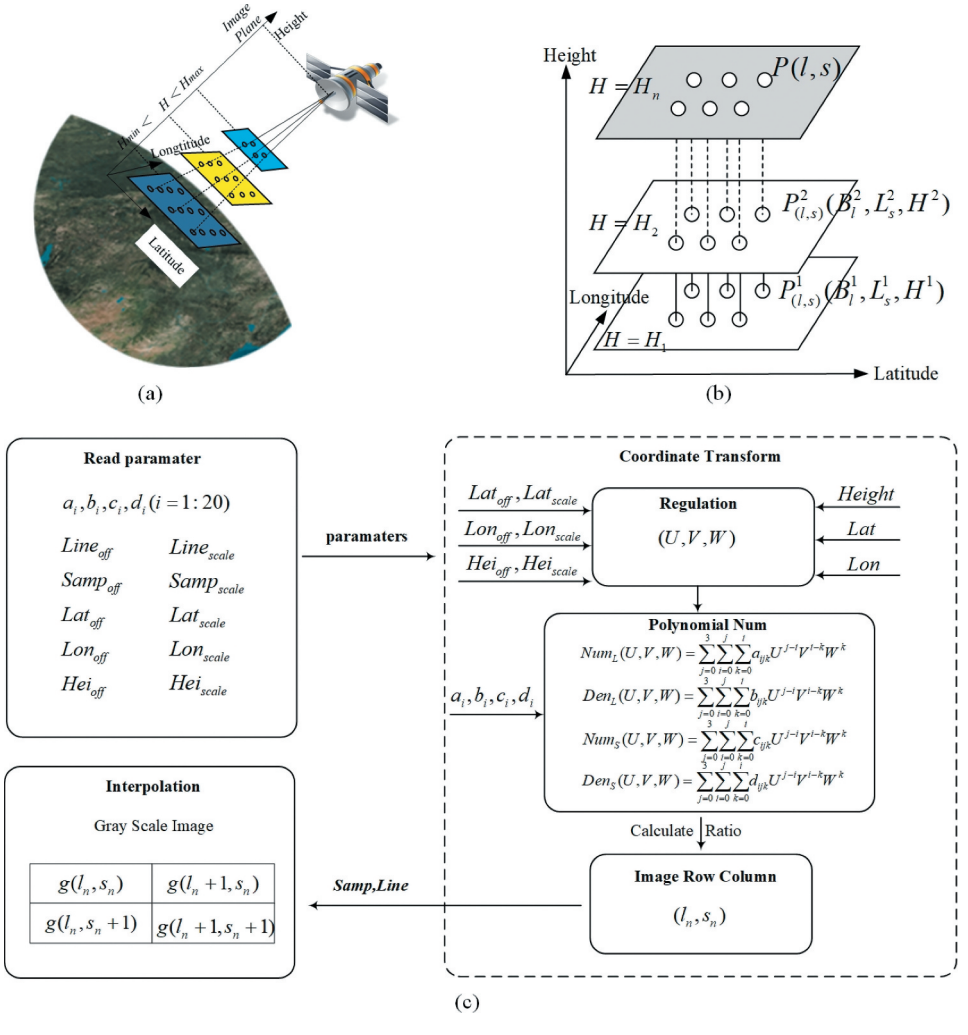


Figure 2. Construction of RFM. (a) Establishing RFM between satellite and Earth. (b) Hierarchical correspondence between object and image coordinates based on RPCs. (c) Flowchart showing the process of calculating RPCs.

$$\begin{cases} I_n = \frac{Num_L(U,V,W)}{Den_L(U,V,W)} \\ S_n = \frac{Num_S(U,V,W)}{Den_S(U,V,W)} \end{cases} \tag{6}$$

2.2. Planar block stitching model

After orthorectification, each image is restored as planar block with upright altitude and longitude coordinates. WPP images are wide in parallax, with large overlapping areas between two adjacent imaging. Among these areas, ensuring a high matching rate of feature pairs is the key to suppressing artefact and distortions therein. Therefore, this section introduces a new method of stitching. First, construct a more stable neighbourhood based on more detected line-pairs and point-pairs. Second, match them based on line-point consistency. Finally, establish an energy function to solve the distortion problems and finally apply that into image warping.

2.2.1. Preliminary processing based on line-point consistency

Since lines are the intersections of planes, assume areas divided by line segments are co-planar neighbourhoods of the image. Use line detection method LSD (Wang et al. 2008) to extract original lines, and then divide the neighbourhoods of the line segment into left one and right one according to gradient direction. This is because the points near the line segment suffering from the sharpest gradient change tend to be within the same planar while others do not. Gradient of a line-segment is defined as the average gradient of all points on it. Within the neighbourhoods constructed by the line-segment l, distance from any pixel to this line is less than $\alpha len(l)$, while that to the vertical bisector is less than $\beta len(l)$. To be noted that α and β are constants set based on image size, aiming at calculating a neighbourhood (size of $2\alpha l \times 2\beta l$) balancing effectiveness and efficiency.

Once constructing neighbourhoods, matching them is the pressing aim. The following operation constructed a line-point invariant to describe each neighbourhood. Calculated from cross ratio (Li and Tan 2010) among projective invariants, Characteristic Number (CN) is constructed as follows: let $\mathbb{P}^{\aleph}(\mathbb{A})$ be n-dimension projective space over field $\mathbb{A}$. $\{P_i\}_{i=1,2,\ldots,R}$ are interest points in $\mathbb{P}^{\aleph}(\mathbb{A})$, all of which construct a closed loop $(P_{R+1} = P_1)$. Distinctive points $\{Q_i^{(j)}\}_{j=1,2,\ldots,S}$ are distributed on line-segments $\{P_iP_{i+1}\}_{i=1,2,-1}$, which can be linearly described by endpoints P_i andP_{i+1} as $Q^{(j)} = a_i^{(j)}P_i + b_i^{(j)}P_{i+1}$, as shown in Figure 3. Assuming that $\mathcal{P} = \{P_i\}_{i=1,2,\ldots,R}$ and $\mathcal{Q} = \{Q_i^{(j)}\}_{i=1,2,\ldots,R}^{j=1,\ldots,S}$, CN between these points can be defined as

$$CN(\mathcal{P},\mathcal{Q}) = \prod_{R}^{i=1}\left(\prod_{S}^{j=1}\frac{a_i^{(j)}}{b_i^{(j)}}\right) \tag{7}$$

Obviously, CN can be described as $CN(l_j, P_1, P_2, P_3)$. The same calculation applies equally for the image from another viewpoint, as $CN(l_j', P_1', P_2', P_3')$. For matching neighbourhood with corresponding matching line-segments l_j and interest points P_i, it's easy to get

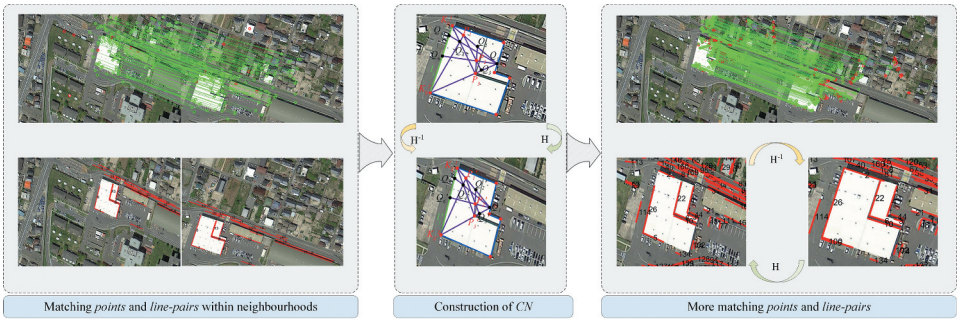


Figure 3. Neighbourhood match and homography of line-pairs and point-pairs within neighbourhoods (highlighted by blue lines). L_j, consisting of K_{lj}^1 and K_{lj}^2, and L_j', consisting of $K_{lj'}^1$ and $K_{lj'}^2$, are matching line-pairs. Red dots P_i and P_i' (i=1,2,3) represent interest points, while black dots Q_i (i=1,2,3,4,5) are the intersection points obtained by connection between K_{lj}^1, K_{lj}^2 and P_i. Obviously, Q_i is linearly related to P_i. Regions highlighted by the above elements are defined as neighbourhoods.

$$CN(l_j, P_i) = CN(l'_j, P'_i) \tag{8}$$

For a pair of regions in target and reference image, the similarity of neighbourhoods (part of the region) can be calculated by a series of *CN* values. Within the matching neighbourhoods (neighbourhoods with highest matching similarity), more interested points can be obtained by intersection of original extracted points, as shown in Figure 3. Assume that black dots Q_1 are the intersection of line $K^1 l_j, P_1$ and $K^2 l_j, P_3$, similarly for $Q_j (j = 2, \ldots, 6)$. It is necessary for an image with large parallax and rich features to conduct the step of extracting more interested points. In this way, more candidate points

are available for image warping.

Since the above invariant predicts the high possibility of interest points located in matching neighbourhoods, the homography (H) between two neighbourhoods can be calculated by matching points through Random Sample Consensus (RANSAC) (Fischler and Bolles 1981). For each line $l_{j_1} \in$ *neighbourhood M* in image 1 (I_1) and $l_{j_2} \in$ *matching neighbourhood M′* in image 2 (I_2), projection can be described as $l_j' = H \cdot l_j$, and $l_j = H^{-1} \cdot l_j'$. If both l_{j_1} and mapped point $l_{j_2}{}'$ (both in I_1) and mapped point $l_{j_1}{}'$ and l_{j_2} (both in I_2) can satisfy double-distance constraints proposed in (Fischler and Bolles 1981), then l_{j_1} and l_{j_2} are considered as matching line-pairs. In this way, more matching line-pairs can be detected. The whole process of extracting more matching lines can be summarized as **Algorithm 1**.

2.2.2. Precise processing based on energy function

The above homography only offers an approximate match within global area, while longer prominent line segment tends to lead to more serious line-warping. Therefore, it requires further constraints. To solve this problem, energy function is a reliable solution (Hao, Zhou, and Cai 2021; Jia et al. 2021; Li et al. 2016; Yuan et al. 2021). In the proposed method, a new energy function model is established considering the weight of both local (overlapping) areas and global areas. The proposed energy function E in (9) consists of three aspects, E_c,, E_h,, and E_n, aiming at preserving collinearity, improving line-point alignment and controlling mesh grid warping distortion, respectively.

Algorithm 1: Extraction of Line Pairs

Input: Two ortho-rectified images I_1^o and I_2^o

Use **SIFT** to detect points and **LSD** to detect lines
Divide coplanar neighbourhoods based on matching points and lines
Match neighbourhoods based on Characteristic Number (*CN*)
Use homography **H** to match lines within neighbourhoods
If (any matching neighbourhoods not used) **then**
    Calculate more line-pair candidates within Neighbourhoods
Else Select matching lines from above
End if

Output: matching lines l_i

$$E(\widehat{V}) = E_c(\widehat{V}) + E_h(\widehat{V}) + E_n(\widehat{V}) \tag{9}$$

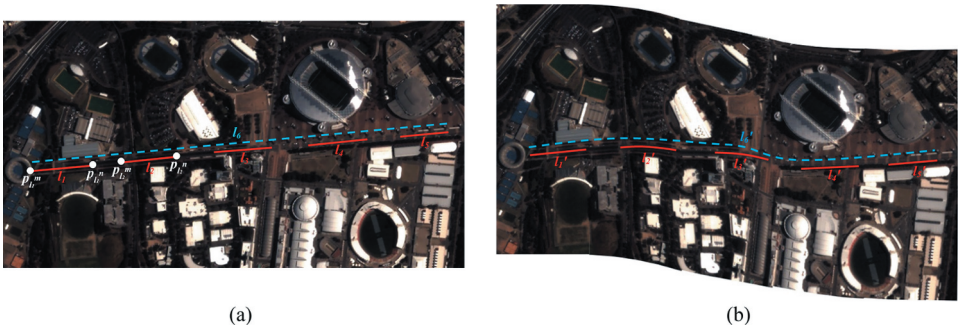
where $V = [x_1 y_1 \dots x_n y_n]^T$ and $\widehat{V} = [\widehat{x_1}\widehat{y_1} \dots \widehat{x_n}\widehat{y_n}]^T$ represent the coordinates vertex of mesh grid before and after warping. In $V \& \widehat{V} \in \mathbb{R}^{\not\in \bowtie}$, index for mesh grid ($n$) is set manually. Figure 5 presents the mesh grid warping by GSP, SPW and ours with same n.

Energy for preserving collinearity. As the first item, E_c in (10) consists of two parts, local linearity and global collinearity. Weight of either part, λ_{lo} and λ_{gl}, is set naturally.

$$E_c(\widehat{V}) = \lambda_{lo} E_{lo}(\widehat{V}) + \lambda_{gl} E_{gl}(\widehat{V}) \tag{10}$$

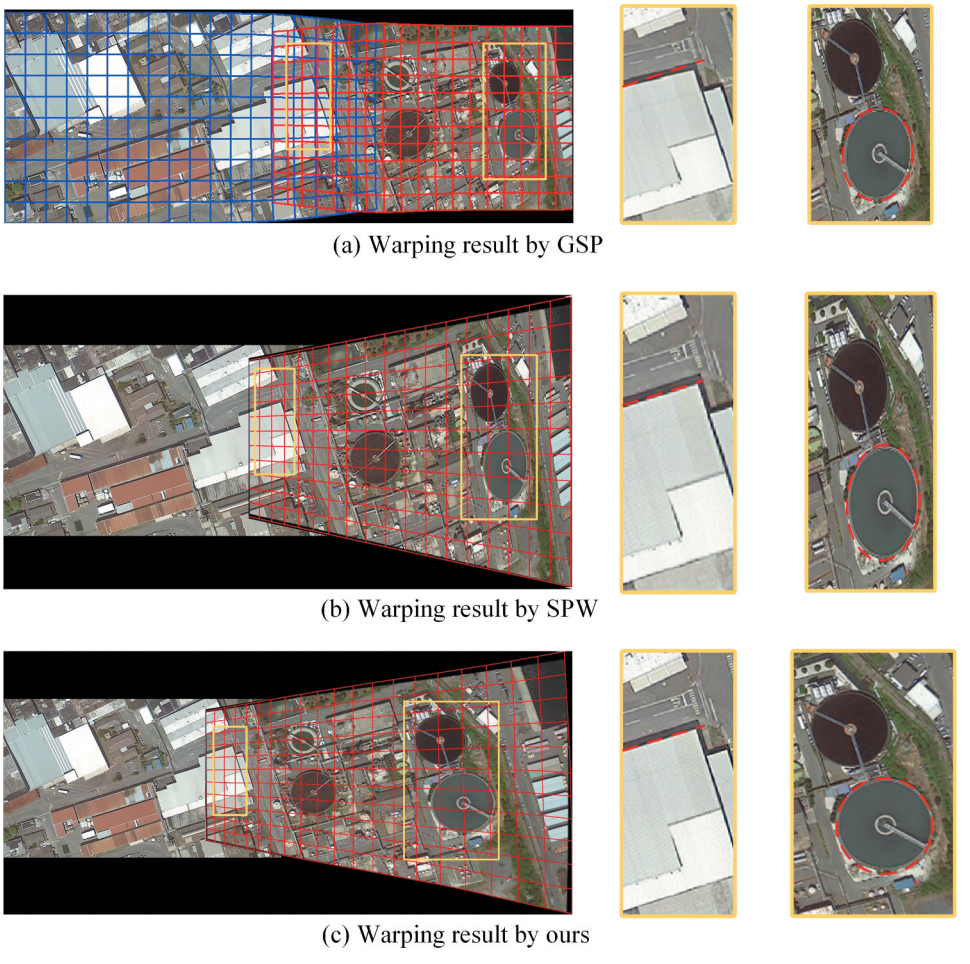
The region shown in Figure 4 is selected from overlapping areas within two orthoimages. Red plotted lines denote the local linear feature, while the blue denote the sum of local ones (in other word, global). Our goal is to preserve the linearity and collinearity of lines during warping. Linearity is used to describe the linear state within lines, while collinearity is to describe linear state between two or multiple line-pairs.

To achieve collinearity as shown in Figure 4(a), two rules are essential. First, distance between endpoints $p_{l_i}^m$ (on l_i) to another line-segment l_j and $p_{l_i}^n$ (on l_i) to l_j should be minimized (slope of l_i and l_j should be close to each other). Second, distance between two adjacent endpoints $p_{l_i}^n$ and $p_{l_{i+1}}^m$ of corresponding lines l_i and l_{i+1} should be minimized too. Given the above two rules, E_{lo} in (10) can be represented as the sum of proportional mentioned products as



(a) (b)

Figure 4. Collinearity in global areas and linearity in local areas according to E_c constraints. Result (a) with collinearity constraint; (b) without collinearity constraint.



(a) Warping result by GSP

(b) Warping result by SPW

(c) Warping result by ours

Figure 5. Comparison of mesh grid warping between GSP, SPW and the proposed method with same number of grid lines. Images are taken at a bay in Iwakuni, Japan. The base image (from the left) is selected from orthoimage, while the other is imaged with certain swing angle. Mesh grid warping by (a) GSP; (b) SPW; (c) ours.

$$E_{lo} = \arg\min_{E_{lo}} \sum_{ls=1}^{N_l} \left(dis\left\langle p_{l_s}^m, l_{ls+1} \right\rangle + dis\left\langle p_{l_s}^n, l_{ls+1} \right\rangle \right) = \sum_{ls=1}^{N_l} \left\| \tau\left(\widehat{\mathrm{p}_{\mathrm{ls}}^{\mathrm{m}}}\right) - \tau\left(\widehat{\mathrm{p}_{\mathrm{ls}}^{\mathrm{n}}}\right) \cdot \overrightarrow{\mathrm{n}_{\mathrm{ls}}} \right\|^2 \quad (11)$$

where $\overrightarrow{\mathrm{n}_{\mathrm{ls}}}$ is the normal vector of corresponding line *ls*. The Euclidean distance $dis\langle P, L\rangle$ denotes the discrepancy between endpoint P and the adjacent line segment L. $\tau(\widehat{\mathrm{P}})$ denotes the vector from original point to desired point P. As a matter of fact, the Euclidean norm of $\tau(\widehat{\mathrm{p}_{\mathrm{ls}}^{\mathrm{n}}}) \cdot \overrightarrow{\mathrm{n}_{\mathrm{ls}}}$ serves as a measure of the difference reflecting the disparity between the normal vector $\overrightarrow{\mathrm{n}_{\mathrm{ls}}}$ and the corresponding line segment *ls*. Likewise, for the mentioned local lines $\{l_{ls}\}(ls = 1, 2, \ldots, N_l)$, global constraints of E_{gl} in (10) can be obtained by incorporating desired point with Euclidean distance on adjacent line:

$$E_{gl} = \arg\min_{E_{gl}} \sum_{ls=1}^{N_l} \sum_{n=1}^{N'} dis\langle p_{ls}^n, p_{ls+1}^n \rangle = \sum_{ls=1}^{N_l} \sum_{n=1}^{N'-1} \left\| \tau\left(\widehat{p_{ls+1}^n}\right) - \tau\left(\widehat{p_{ls}^n}\right) \cdot \overrightarrow{\mathrm{n_{ls}}} \right\|^2 \tag{12}$$

For the lines $l_i (i = 1, 2, \ldots, 5)$ within the given region during image warping, there is no collinearity constraint applied as shown in Figure 4(b), while there is collinearity applied as shown in Figure 4(a). Red plotted lines are destroyed during warping without collinearity constraints. It proves necessary and effective to apply the proposed collinearity rules.

Energy for improving line-point alignment. E_h in (13) is designed to control point alignment and line alignment after warping, which means they should be coincident with themselves after warping.

$$E_h(\widehat{V}) = \lambda_p E_p(\widehat{V}) + \lambda_l E_l(\widehat{V}) \tag{13}$$

where point alignment element E_P, which is correspondences between $m_i(x_i, y_i) \in I$ and $m_i'(x_i', y_i') \in I'$, and line alignment element E_l are multiplied by corresponding weight λ_P and λ_l. Assume that matching point sets $\{p_i, p_i'\}_{i=1}^S$ and line sets $\{l_j, l_j'\}_{j=1}^T$ are the correspondences point sets and line sets from I and I'. Quality of point alignment E_P can be measured as

$$E_p(\hat{V}) = \sum_{i=1}^{N} ||m_i' - \varphi(\widehat{p_i})|| = ||M' - W_p \hat{V}||^2 \tag{14}$$

where M' is a series of coordinates of m_i' represented in $\mathbb{R}^{\nVdash\mathbb{N}}(N \neq n)$. $\mathrm{W_p} \in \mathbb{R}^{\nVdash\mathbb{N}\times\nVdash\times}$ is coefficient vector of coordinates of the bilinear interpolation of mesh point $\widehat{p_i}$. As for line alignment E_l, line l_{j_0} can be expressed as two endpoints $m_{j_0}^{start}$ and $m_{j_0}^{end}$ which is $l_{j_0} : a_{j_0}x + b_{j_0}y + c_{j_0} = 0$.

In this way, the quality of line alignment E_l can be measured as

$$E_l(\widehat{V}) = \sum_{j=1}^{M} \left| \left\langle \varphi\left(\widehat{m}_j^{start,end}\right), (a_j, b_j)^T \right\rangle + c_j \right|^2 = \| \mathrm{W_l}\widehat{V} + C\|^2 \tag{15}$$

where $\mathrm{W_l} \in R^{2M\times 2n}$ is the coefficient vector of inner results of bilinear interpolation of $m_j^{start,end}$ and $(a_j, b_j)^T$, similarly. Compared with alignment proposed in (Chen and Chuang 2016), functions in (14) and (15) prove more robust in controlling alignment.

Energy for controlling warping distortion. E_n in (16) defines the slope and size of grid lines to control distortion during image warping. The grid lines are regarded as intrinsic structures of the to-be-warped image. Distortion in this item is composed of two parts, rough projection E_{nr} and precise projection E_{np}. It is noteworthy that the grid lines generated by projection describe the intensity of image warping over the to-be-warped image, as shown in Figure 5.

$$E_n(\widehat{V}) = \lambda_{nr} E_{nr}(\widehat{V}) + \lambda_{np} E_{np}(\widehat{V}) \tag{16}$$

where rough projection element E_{nr} and precise projection E_{np} are multiplied by corresponding weight λ_{nr} and λ_{np}. In (16), rough projection E_{nr} is inspired by (Igarashi and Igarashi 2009), which can ensure a rough similarity transformation by

$$E_{nr}(\widehat{V}) = \sum_{i=1}^{N} \sum_{(u,v)\in L_i} \left\| (\widehat{v_v^i} - \widehat{v_u^i}) - D_{uv}^i (v_v^i - v_u^i) \right\|^2 \tag{17}$$

where v and $\widehat{v}$ denote the vertex of each point on edge lines (L_i) from image I_i of mesh grid before and after warping, respectively. D_{uv}^i represents the similarity transformation for edge (u, v) consisting of transform function c and s, which is

$$D_{uv}^i = \begin{bmatrix} c(e_{uv}^i) & s(e_{uv}^i) \\ -s(e_{uv}^i) & c(e_{uv}^i) \end{bmatrix} \tag{18}$$

where rotation matrix D_{uv}^i, given as a transformation that maps edge points to new positions, is calculated by $c(e_{uv}^i)$ and $s(e_{uv}^i)$ optimized in (Igarashi and Igarashi 2009).

After the rough projection E_{nr}, precise projection E_{np} is carried out. Different from (Chen and Chuang 2016; Jia et al. 2021; Liao and Nan 2020), ours reinforces the progressive connection between projection E_{nr} and E_{np} because overlapping areas within RS images are different from those in natural images in terms of overlap size and overall image size.

$$\begin{aligned} E_{np}(\widehat{V}) = \sum_{i=1}^{N} \sum_{e_j^i \in L_i} \Bigg\{ & W(e_j^i)^2 \left[\left(c(e_j^i) - s_i \cos\theta_i \right)^2 + \left(s(e_j^i) - s_i \sin\theta_i \right)^2 \right] \\ & - \lambda_i \left\| (\widehat{v}_j - \widehat{v}_i) - D_{ij}^i (v_j - v_i) \right\|^2 \Bigg\} \end{aligned} \tag{19}$$

where weight function $W(e_j^i)$ is conducted on image edge transformation. Given desired scale s_i and rotation angle θ_i for image I_i, the transformation is carried out for each edgee_j^i. Different from previous works, which establish $W(e_j^i)$ as linearly related to distance from desired point to edge, our study re-establishes the weight function $W(e_j^i)$ as cubic weight function. It makes the first item's weight decrease with distances further from the

overlapping areas. In comparison, the linear weight function adopted in (Liao and Nan 2020) results in uneven and discontinuous grey value at the edges, since there are large overlaps between WPP blocks. For overlapping areas, an anti-repeat factor λ_i is proposed to eliminate the impact of repeated warping from E_{nr}, while in areas far from overlapping areas, redesign the weight function and decrease λ_i. In this way, more weight is set for edge areas of the whole image, as these areas are lack of alignment from E_h.

Figure 5 presents the warping results between two RS images, in which the base image is selected from orthoimage over a bay in Iwakuni, while the other is imaged with certain swing. It aims at verifying the quality of mesh grid warping according to different energy functions adopted by GSP (Chen and Chuang 2016), SPW (Liao and Nan 2020) and ours, respectively, by estimating visual quality within the stitched image. Specifically, for the given same number of grid lines, three methods exhibit different performance in preserving linear structures and suppressing distortion. Two representative RS structures from overlapping region and global region, roof and oil tank are selected for presentation and highlighted with yellow bounding boxes. In GSP, linear structure (roof) within overlapping region is disrupted, while SPW and ours succeed in preserving linear structures. On the other hand, within the oil tank from global region, GSP and SPW both fail

to recover the circular shape of the oil tank, while our result is closer to the ground truth. It proves the effectiveness of the proposed E_n constraints.

Final warping. All mentioned constraints are quadratic. A reasonable reformation function can be solved by a sparse linear solver. Specifically, iteration methods are adopted in our sparse linear system to solve the energy minimization. The optimal mesh grid warping can be determined by

$$\widehat{V} = \arg\min_{\widehat{V}} E_c(\widehat{V}) + E_h(\widehat{V}) + E_n(\widehat{V}) \tag{20}$$

After the whole function is carried out, the warp between two images can merge towards a more precise way, with artefact and distortions effectively suppressed. The whole process of optimizing mesh grid warping can be summarized as **Algorithm 2**.

Algorithm 2: Optimization of mesh warping according to energy function

Input: Original mesh grid $\boldsymbol{V}$ linearly generated by RANSAC

Calculate rotation angle θ based on $\mathbf{H} \subseteq [3\times3]$ by RANSAC, and obtain rotation matrix $\mathbf{R} \subseteq [2\times2]$
Construct energy function in terms of collinearity, segment alignment and distortion:
  Construct collinearity based on rotated vertical lines v and horizontal lines u
    For $i = 1 \rightarrow num(u)$ **and** $j = 1 \rightarrow num(v)$ **do**
      For $x = lo$ and gl (local linearity and global linearity) **do**
        Calculate $E_x \subseteq (s\_ul+s\_vl, N_0)$ ($x = lo$ or gl) based on u and v
  Construct line and point alignment based on matching lines $\subseteq [m, 4]$ and matching points $\subseteq [2, n]$
    For $i = 1 \rightarrow m$ **and** $j = 1 \rightarrow n$ **do**
      Calculate energy for controlling line alignment $E_l \subseteq [s\_l, N_0]$
      Calculate energy for controlling point alignment $E_p \subseteq [s\_p, N_0]$
  Construct controlling-distortion based on homography $\mathbf{H}$ by RANSAC
      Calculate rough transformation $\mathbf{D}$ by desired scale factor s and rotation angle θ
      Project edge points v into $\widehat{v}$ within to-be-warped image I_1
      Calculate energy by rough projection $E_{nr} \subseteq [s\_nr, N_0]$ based on v, $\widehat{v}$, and $\mathbf{D}$
      Re-establish weight function cubic-correlated to distance to edge $W \propto (1 - (x/d)^3)^3$
      Calculate energy by precise projection $E_{np} \subseteq [s\_np, N_0]$
For all mentioned energy E_x and corresponding weight λ_x, total matrix $\mathbf{A}$ can be constructed as **do**
  $\mathbf{A} = [\sum E_x \cdot \lambda_x] = [s\_u+s\_v+s\_l+s\_p+s\_nr+s\_np, N_0] \subseteq [M_0, N_0]$
Construct ideal target $\mathbf{X}$=[**dir_p**, $\lambda_l \times$**dir_l**, 0, . . . , 0]$\subseteq[M_0, 1]$, where
  dir_p and **dir_l** store the location of a total of m_p matching points and m_l matching lines, as
    Location of matching points is **dir_p** = $[x_1, y_1, x_2, y_2, \ldots, x_{mp}, y_{mp}] \subseteq [2m_p, 1]$
    Location of matching lines is **dir_l** = $\left[-\frac{c_1}{d_1}, -\frac{c_1}{d_1}, -\frac{c_2}{d_2}, -\frac{c_2}{d_2}, \ldots, -\frac{c_{ml}}{d_{ml}}, -\frac{c_{ml}}{d_{ml}},\right] \subseteq [2ml, 1]$
      where lines are represented as
        $l_i : a_i x + b_i y + c_i = 0 (d_i = \sqrt{{a_i}^2 + {b_i}^2})$
Put linear equation system $\mathbf{A}\cdot\widehat{V} = \mathbf{X}$ into sparse linear solver
Obtain the optimized control vertex vector $\widehat{V}$ within I_1

Output: Optimized mesh grid

3. Experiments and discussion

3.1. Study area

Two study areas are used for verification in this paper, as shown in Figure 6. One is located at suburban Beijing, China (39°27′N-41°41′N, 114°58′E-117°47′E), as shown in Figure 6(a). Its ground resolution is 16 m, covered with panchromatic spectral bands by pushbroom GF-1 Satellite. The above area equivalent image solved by rigorous sensor model (RSM) is presented as Figure 6(b), while the corresponding 30 m-resolution DEM is exhibited as Figure 6(c). The RSM image in Figure 6(b) is solved as a reference ground truth of corresponding area by Landsat-8. Main geographical types in this area cover mountains, hills, plains, rivers, lakes, etc.

The other study area is imaged by Orbita hyperspectral satellite (OHS, denoted as Orbita), which is the first commercial hyperspectral satellite in China that completed launching and networking, with a ground resolution of 1 m. Imaging region of study area 2 is located at urban Sydney, Australia (33°45′S-33°54′S, 151°2′E-151°13′E), as shown in Figure 6(d). Given

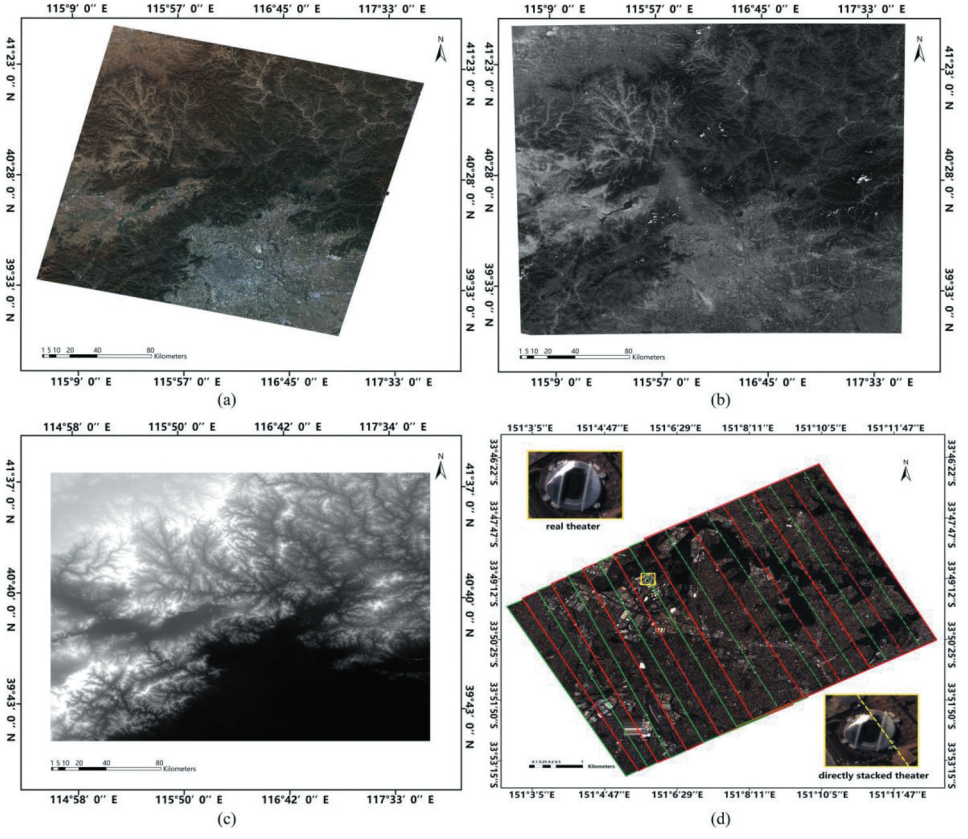


Figure 6. Study area and presentation of image data: (a) experimental image of study area 1 by GF-1; (b) imaging of study area 1 by Landsat-8, solved by RSM; (c) 30m-DEM of study area 1; (d) images from 1 to 10 of study area 2 by Orbita, all of which directly stacked together by ENVI software. The *n*-th red-and-green interlaced solid lines represent the start position of the *n*-th image, while the *n*-th dotted ones indicate the corresponding end position. Among it, the detail marked by the yellow box shows that the directly stacked theatre has obvious misalignment compared to the real theatre.

image plane by Orbita is 4096×3072, each imaging swath is about 4 km × 3 km. Main structures in this area cover theatres, harbours, lakes, civil buildings, featured with complex linear geometric details. Within Figure 6(d), image with red-and-green interlaced lines represents that it is formed by consecutively stacking 10 WPP images (mainly based on unrectified longitude and latitude coordinates). Due to the lack of geometric correction and accurate stitching, huge distortions and misalignments exist between the adjacent images, as shown in the yellow box. Accurate geometric details are hard to be extracted out on such images. Experiments are conducted on above areas.

3.2. Orthorectification of image

Conduct orthorectification following the flowchart is illustrated in Figure 2(c). In addition, to improve orthorectification accuracy, GCP distribution and global 30 m-DEM are adopted. The results are compared with the image by RSM shown in Figure 8(b). Specifically, virtual GCPs are distributed by crossroads at intersection of prescribed latitude and longitude lines.

For the evaluation of orthorectification, three metrics are adopted, root mean square error (RMSE), structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), respectively. First, RMSE for each GCP is calculated before and after orthorectification. Certain GCPs among 29 GCPs for comparative analysis are selected, in which the notable trend in RMSE variations is observable from 13th to 26th point. The corresponding residual (Res_i), which denotes the difference between the coordinate of GCP_i, before and after orthorectification, is exhibited in Table 1. RMSE in Table 1 denotes the RMSE of cumulative residuals as in (21). It indicates that RMSE tends to decrease with the increase in GCPs. Lower RMSE denotes better orthorectification result. Aside from

$$\left.\begin{aligned} \overrightarrow{Res_i} &= \overrightarrow{GCP_i(ortho)} - \overrightarrow{GCP_i(origin)} \\ \overrightarrow{Res_i} &= \overrightarrow{Res_{ix}} + \overrightarrow{Res_{iy}} \\ \overrightarrow{RMSE_i} &= \overrightarrow{RMSE_{ix}} + \overrightarrow{RMSE_{iy}} \end{aligned}\right\} RMSE_i = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\overrightarrow{|Res_i|}\right)^2} \tag{21}$$

Table 1. Relationship between number of GCPs and RMSE (study area 1).

	Residual of corresponding GCP (unit in pixel)		Control Point Cumulative Error: RMSE (unit in pixel)		
Number of GCPs	Res_x	Res_y	$RMSE_x$	$RMSE_y$	$RMSE$
13	−1.1126	0.5553	0.4574	0.2879	0.5400
14	0.3337	−0.2447	0.4516	0.2858	0.5342
15	0.2332	−0.3826	0.4416	0.2953	0.5311
16	0.2005	−0.5053	0.4321	0.3168	0.5358
17	0.1911	0.0847	0.4232	0.3079	0.5237
18	−0.0037	−0.4690	0.4116	0.3295	0.5274
19	0.0116	−0.2716	0.4005	0.3284	0.5179
20	0.0432	−0.1337	0.3905	0.3216	0.5058
21	0.0437	−0.0105	0.3816	0.3142	0.4942
22	0.0368	0.0958	0.3726	0.3074	0.4832
23	0.0495	−0.5042	0.3647	0.3211	0.4863
24	−0.1974	0.1168	0.3595	0.3153	0.4779
25	−0.2111	−0.0105	0.3542	0.3089	0.4700
26	−0.2616	−0.1200	0.3511	0.3042	0.4642

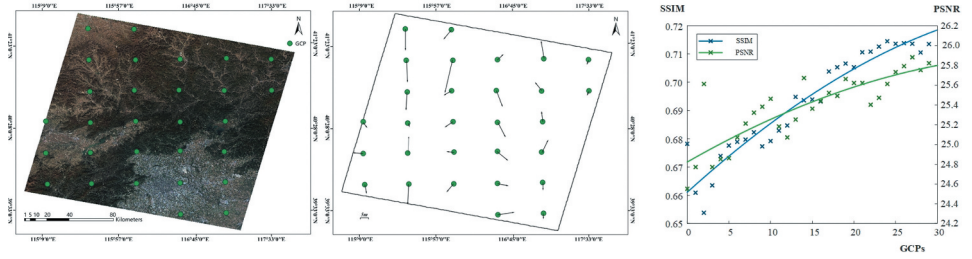


Figure 7. Distribution and residuals of virtual GCPs of study area 1. (a) Distribution of 29 virtual GCPs over study area 1; (b) Residuals of corresponding GCPs. (c) SSIM and PSNR according to different numbers of GCPs.

RMSE, which describes the cumulative deviation, SSIM and PSNR are also adopted to evaluate the accuracy of orthorectification. Select a subregion of 250×250 pixels of certain centre coordinates in the area shown in Figure 8(c) and compare it with the corresponding subregion imaged by RSM in Figure 8(b) in terms of SSIM and PSNR. According to (Tuna, Unal, and Sertel 2018), higher SSIM and PSNR indicate higher similarity between images. With variable in number of GCPs, comparison result shown in Figure 7(c) implies as the number of given control points increases, the similarity indicators between

orthoimage and RSM image increases, which indicates the specified region approaches the RSM imaging more closely. Figure 8(d)(j) display the differences between two pairs of orthoimages and non-orthoimages over study area 2. It is observable that the roof shape is different before and after orthorectification in Figure 8(f)(g), which confirms the necessity of orthorectification for recovering geometric structures.

3.3. Planar block stitching

In this section, both qualitive and quantitative experiments are carried out among stitching of RS images by Orbita and four groups of image sets covering four imaging modes, including swing angle at 0°, 10°, 20° and 27°. Set each grid size during warping as 40×40 for more refined merging. In the collinearity constraint function from energy function (10),

keep the λ_{lo} and λ_{gl} to 80 and 120, respectively. As parameters aiming at preserving the alignment and homography between lines and points in (13), λ_p and λ_l are kept as 2 and 8. Besides, parameters that construct mesh grid warping in (16), λ_{nr} and λ_{np} are set 80 and 120. Keeping all the mentioned parameters constant is conducive for proving the effectiveness of the proposed method. Visual results of the proposed stitching method between two input RS images by Orbita are shown as Figure 9, while the result of 10 consecutive images stitching over study area 2 by Orbita is exhibited in Figure 10. Each input image from Orbita is the same size as image plane, which is 4096×3072 pixels. For fair comparison, same orthoimages are used as input for the selected algorithms. Besides, all selected algorithms are validated due to their best performance provided by authors. As shown in Figure 9, edges of each image are displayed within the selected yellow box. It indicates that regional mapping area extends with two orthoimages effectively stitched.

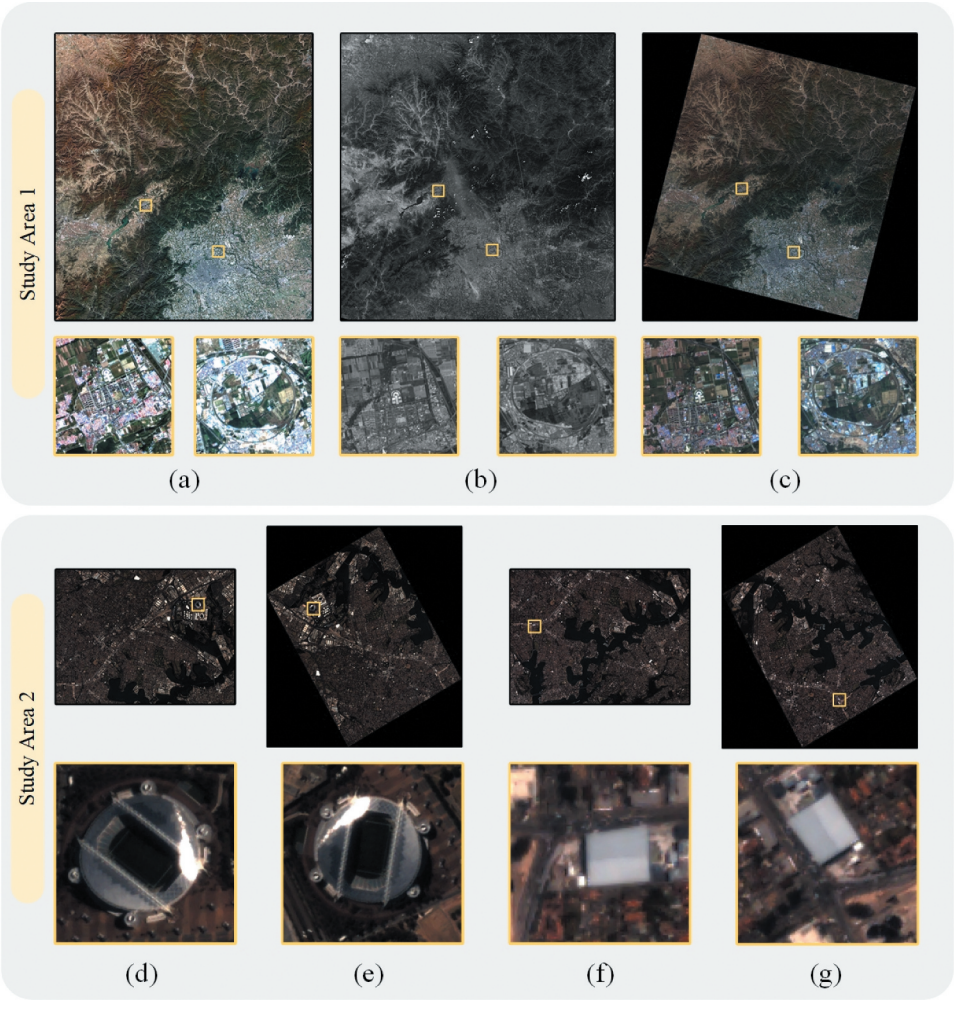


Figure 8. Orthorectification of both study areas. The first row is located at study area 1, while the second is study area 2. (a) Original image by GF-1; (b) RSM imaging by Landsat-8; (c) orthorectification with certain GCPs (variable); (d) original 1 by Orbita; (e) orthoimage 1; (f) original image 2 by Orbita; (g) ortho image 2 Yellow rectangles of same size and same centre geometric coordinates are enlarged to same size for visual comparison.

Qualitative evaluation. Due to the sensitivity of reflectance and variations in exposure levels, the appearance of certain objects in RS images can vary at different times, especially for objects with high reflectance (e.g. roofs, water surface). Therefore, two interested regions, building roof and swimming pool, are selected within the stitched image over study area 2 by six different methods, as shown in Figure 10. The interested regions, which are of same centre geometric coordinates and same regional size, vary significantly among stitched image by different methods. The building roof by SPW and APAP exhibits noticeable colour aberration, whereas AANAP shows pronounced linear misalignment, while by GSP and ENVI software displays structural deformation. Likewise, the swimming pool

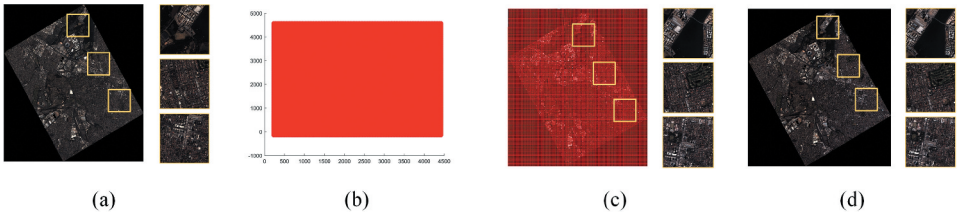


Figure 9. Results of two Orbita WPP orthoimages stitching based on the proposed method. (a) input base image (I_1^o). (b) Optimized mesh warping result according to energy function. (c) Image (I_2^o) warped according to mesh warping result $(I_{2\_warped}^o)$. (d) Stitching result $(I_1^o + I_{2\_warped}^o)$.

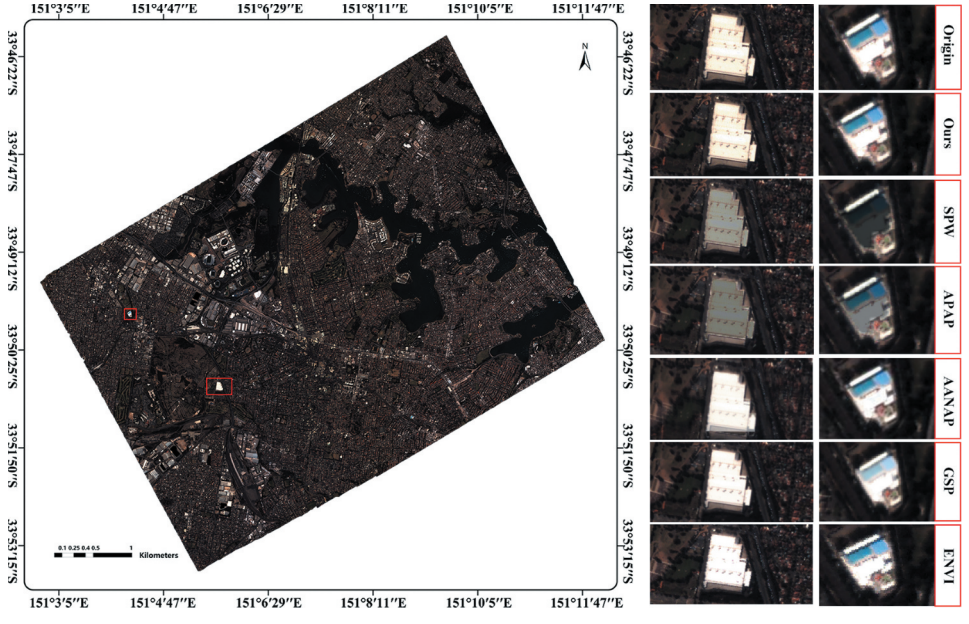


Figure 10. Results of concatenating ten WPP orthoimages over study area 2. Rectangles marked in red boxes exhibit building roof and swimming pool. On the right side displays the comparison of the red rectangle area within stitched image by SPW, APAP, AANAP, GSP, ENVI software, and the proposed method.

indicates similar results, in which ours preserves the original structure and colour tones most effectively in the cumulative stitching process. It proves that the proposed *CN* is remarkable for dividing matching regions, while the proposed energy function is effective for constraining the linear relationships during the matching process.

Quantitative evaluation. More quantitative evaluation is carried out in section 3.4.

3.4. Further discussion on generality and accuracy

To evaluate the robustness of proposed method, a simulation experiment is carried out on simulated images from Orbita with different swing angles (φ). Changes in swing angles

can result in local scale variations and different viewpoints across RS images. In view of this, the simulation experiment is aimed at verifying whether the images $I_j^s(j = 1, 2, 3)$, stitched by images $I_j(j = 1, 2, 3)$ with swing angles (φ_j) and image I_4 (based on another nadir different from I_j), can be restored into a unified viewpoint and preserve geometric structures or not. As shown in Figure 12, (a)(b)(c) are imaged based on the same nadir with different swing angles, 10°, 20° and 27°, corresponding to φ_i in Figure 11(a). The reason for selecting such swing angles is based on that too small swing angles cannot produce significant distortion (Phinn et al. 2002), while too large swing angles (Zhang et al. 2019) may result in decreased edge resolution, amplified swing angle caused

by severe atmospheric disturbances and capturing the sides of buildings rather than the roofs. Figure 12(d) is the adjacent image (different nadir from abc) without swing angle, corresponding to imaging with certain field of view (FOV) φ_P in Figure 11(a).

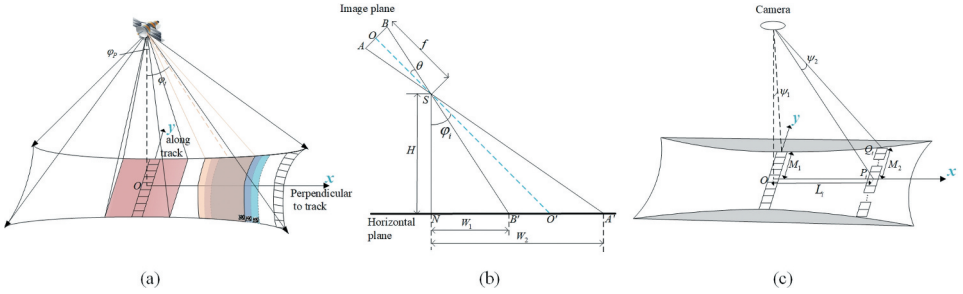


Figure 11. Diagram of pushbroom imaging with swing angles. (a) imaging with different swing angle φ_i compared to pushbroom imaging with Field of view (FOV) of φ_P (denoted by red area); (b) analysis along X-axis; (c) analysis along Y-axis.

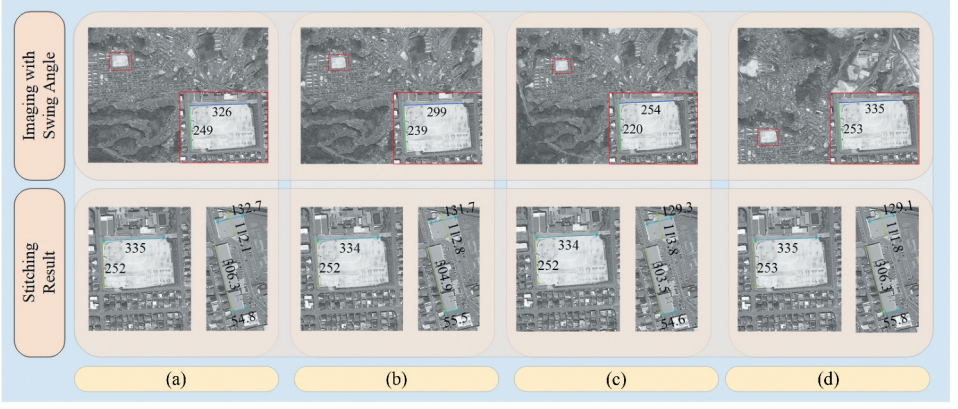


Figure 12. Simulated images with variable in swing angles and corresponding stitching results. Within the first row, (a)(b)(c) are imaged due to different swing angles based on the same nadir, while (d) is imaged based on another nadir without swing angle. The second row displays the geometric length of interested building within the stitched image. Within each image set, one building from centre (overlapping area) and the other from edge (non-overlapping area) are displayed. (a) swing angle of 10°; (b) swing angle of 20°; (c) swing angle of 27°; (d) no swing angle from another nadir. (unit in pixel).

Within the first row in Figure 12, area delimited in red rectangle in each image is enlarged to the same display size on the bottom right. Blue line and green line with annotated number in each enlarged rectangle represent length and width of the displayed building (unit in pixel). If the stitched image I_j^s can demonstrate visual naturalness and preserve linear structures across the whole image, it can verify robustness of the proposed method roughly. More quantitative experiments are carried out in section 3.4.1-3.4.3.

Given that the altitude of satellite Orbita (H) is 530 km, the Earth radius (R) is 6371 km, image plane $(m \times n)$ is the same as above 4096×3072 pixels, pixel size (a) is 4 μm and focal length (f) is 0.5 m; four experiments with different swing angles including that in section 3.3 are conducted.

In the optical imaging diagram in Figure 11, ground optical axis centre O' varies with H and φ. Resolution of nadir is $N = aHf^{-1}$. Half angular FOV alongside X and Y direction (θ, ψ) can be solved as $(\arctan(na2f^{-1}), \arctan(ma2f^{-1}))$. Width of imaging area (W) can be calculated as distance from right edge (which is A' to N, $W_2 = HN^{-1}\tan(\varphi + \theta) + 1)$ to left edge (which is B' to N, $W_1 = HN^{-1}\tan(\varphi - \theta)$), as shown in Figure 11(b). Likewise, half height of imaging area (alongside the vertical paper as shown in Figure 11(b)) near the nadir point is $M_1 = 2\tan\psi_1 N^{-1}\sqrt{(W_1 \times N)^2 + H^2}$, while beyond nadir is $M_2 = 2\tan\psi_2 N^{-1}\sqrt{(W_2 \times N)^2 + H^2} + 1$, as shown in Figure 11(c).

Given the above simulation conditions, further experiments are proceeded where the non-deformation adjacent image I_4 is adopted as base image, and the images with swing angle $I_j (j = 1, 2, 3)$ are adopted as the target (to-be-warped image). By comparing linear structures in the stitched image $I_j^s (j = 1, 2, 3)$, we can distinguish the difference visually and evaluate stitching quality. Furthermore, we conduct quantitative evaluation on both lines and points by adopting three indicators, line-pairs extraction and matching level, point-pairs fusion error (RMSE) and cumulative error, respectively. All input orthoimages are the same size as satellite image plane in section 3.3 (4096×3072).

3.4.1. Line extraction capability

The first indicator evaluates the matching level between line-pairs. Within the above image sets, compare the line-matching method proposed in section 2.2 and LP (Fan et al. 2010; Fan, Wu, and Hu 2012). Results show that the proposed method is robust in extracting line-pairs and matching them in various swing angle imaging.

Results in Table 2 consists of two parts. Within either column of LP or proposed method, number in the first column shows line-pairs that are matched correctly, while number in the second column denotes all matching line-pairs. The existence of correct-matching is attributed to that a line (l_x) in image 1 (I_1) could possibly be matched to different lines $\{l_{y\_n}\}(n = 1, 2, \ldots)$ in image 2 (I_2), meanwhile only one line $l_{y\_c}(c \in n)$ in I_2

Table 2. Comparison of line-pairs matching accuracy between LP and ours.

Dataset	Proposed Method	LP
Swing Angle of 0°	(**776**/**793**, **97.9%**)	(189/195, 96.9%)
Swing Angle of 10°	(**254**/**270**, **94.1%**)	(244/262, 93.1%)
Swing Angle of 20°	(**149**/**158**, **94.3%**)	(136/150, 90.7%)
Swing Angle of 27°	(**73**/**81**, **90.1%**)	(53/59, 89.8%)

is the correct matching one for I_x. The last column is the accuracy ratio between the two numbers ahead. Within Table 2, the proposed method reaches a matching accuracy higher than 90% in all swing angles.

Besides, our line extraction and matching strategy outperforms LP significantly with an increase in swing angle. It verifies the effectiveness of extracting more line-pairs by constructing neighbourhoods based on *CN*, and indicates that the proposed method can obtain higher matching accuracy and more extracted line-pairs with an increase in scene complexity.

In stitching two orthoimages from section 3.3 (denoted as swing angle of 0° in Tables 2 and 3), line-pairs extracted by ours reach 4.07 times of that by LP. The total matching number is important, as more matching candidates can bring higher robustness for image warping (Fan et al. 2010).

Although the number of matching line-pairs suffers from a sharp decline with an increase in swing angle, as shown in Table 2, the results verify that our line matching strategy still owns a better performance in both number of extracted line-pairs and matching accuracy.

Within interested building from stitched images $I_j^s (j = 1, 2, 3)$ and ground truth I_4^{GT} shown in the second row of Figure 12, buildings extracted from I_j^s, as shown in the second row in Figure 12(a)(b)(c) are geometrically accurate compared to the ground truth I_4^{GT}, as shown in the second row of Figure 12(d). Within the left main building (from the centre of overlapping area), maximum error in both length and width is less than 1 pixel, lower than 0.40%. Within the right long building (from image edge), maximum error in length is 0.92% (from c), while the maximum error in width is 2.15% (from c). It proves stable and capable of preserving structures across the whole image with different swing angles.

3.4.2. Root mean square error (RMSE)

The second indicator describes the alignment level of matched point-pairs. Compare the RMSE of APAP (Zaragoza et al. 2013), AANAP (Lin et al. 2015), GSP (Chen and Chuang 2016), SPW (Liao and Nan 2020) and ours operated on the same dataset, as shown in Table 3. The smaller RMSE is the better the fusion that is carried out. In experiment in section 3.3, among all methods conducted on RS images, the proposed method gains the lowest error with RMSE of 0.87, lower than that by SPW. Besides, variance of our RMSE is 0.007, 84% lower than SPW. Though RMSE tends to increase with a boost in swing angle, ours still elicits the best performance among the selected algorithms.

3.4.3. Cumulative stitching accuracy

It is essential to estimate cumulative accuracy in both absolute and pixel level because the single-feature based RS stitching algorithms mentioned in section 1 tend to result in high cumulative errors. Absolute accuracy (R_a^j) denotes the errors gradually accumulated and amplified during stitching of multiple images, unit in m. Itis the product result of cumulative stitching accuracy ($RMSE_j$, denoted as R_p^j, unit in pixel) and ground sampling

Table 3. Comparison of RMSE in different image sets.

Dataset	APAP	AANAP	GSP	SPW	Ours
Swing Angle of 0°	6.70	1.39	1.97	0.88	**0.87**
Swing Angle of 10°	4.56	4.91	2.42	1.15	**0.89**
Swing Angle of 20°	4.72	5.01	2.36	1.27	**0.94**
Swing Angle of 27°	8.96	4.28	4.30	1.48	**1.09**

resolution (GSD^j), where j denotes the stitched image after j-th stitching. To calculate GSD^j within the j-th stitched image, a common way is to divide the j-th stitched image into several sub-areas $(\Omega_i, i = 1, \ldots, n)$ and calculate the ratio of the number of pixels (m_x) to the actual distance ($\mathscr{L}_x$, unit in m) (x denotes a total number of $\mathscr{L}_x$ within Ω_i) on the ground in each sub-area, as

$$GSD^j = \frac{1}{n}\sum_{\Omega_i=1}^{\Omega_n}\sum_{x=1}^{l_i^{\Omega_i}}\frac{\mathcal{L}_x}{m_x} \tag{22}$$

In this section, denote cumulative accuracy (R_p^n) of a region as the average RMSE of each sub-region (Dennison and Roberts 2003), which is

$$R_p^n = \sqrt{\frac{\sum_{i=1}^{N}\left(P_i' - P_i\right)^2}{N}} = \frac{1}{n}\sum_{j=1}^{n} RMSE_j = \frac{1}{9}\sum_{j=1}^{9}\sqrt{\frac{\sum_{i=1}^{N_j}\left(P_i^{j\prime} - P_i^j\right)^2}{N_j}} \tag{23}$$

Given original GSD^0 of ortho image 1 (I_1^o), also known as nadir resolution, as 1 m. It's calculated according to pixel size (a), focal length (f), and orbital altitude (H). When GSD^j and R_p^j are derived from (22) and (23), R_a^j can be obtained by incorporating GSD^j and R_p^j.

$$\left.\begin{aligned} GSD^0 &= \tfrac{a}{f} \times H = 1m \\ GSD^j &= \tfrac{1}{n} \times \sum_{\Omega_i=1}^{\Omega_n} \tfrac{L_l}{m_j}\left(\Omega_n \subseteq I_j\right) \\ R_p^j &= RMSE^j \end{aligned}\right\} R_a^j = R_p^j \times GSD^j (j = 1, \ldots, 9) \tag{24}$$

As a result, Figure 13 presents the cumulative errors during the stitching process, including three forms of errors, GSD (unit in m), stitching accuracy (unit in pixel), and absolute accuracy (unit in m), respectively. X-axis plots the first to ninth stitching,

denoted as 1–9 in X-axis. To verify the necessity of ortho-rectification, a comparison is conducted between cumulative errors of orthoimages (indicated in green line) and un-orthoimages (indicated in blue line). As shown in Figure 13(a), GSD of un-orthoimages accumulates significant errors with an increasing number of stitching, 9.81% higher than that of orthoimages. Moreover, variance of GSD within orthoimages is 99% lower than that in un-orthoimages, which further verifies the stability of GSD by our stitching method. Although cumulative stitching error of un-orthoimages is slightly better than

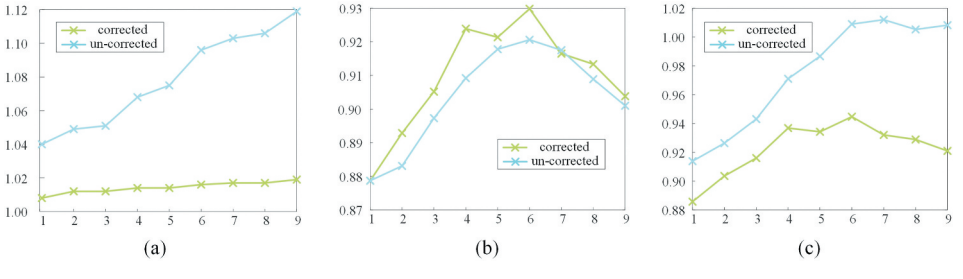


Figure 13. Cumulative error among multiple stitching. Green line denotes error of orthoimages, while blue line denotes error of un-orthoimages. (a) GSD (unit in m); (b) cumulative stitching error (unit in pixel); (c) cumulative absolute error (unit in m).

that of orthoimages as shown in Figure 13(b), the accuracy of orthoimages still outperforms un-orthoimages by 9.47% in terms of absolute error as shown in Figure 13(c). This is due to the fact that interpolation (pixel resampling) during orthorectification process may disrupt some linear structures, resulting in a higher contrast of matching feature-pairs in the un-orthoimages compared to orthoimages. As a result, the un-orthoimages may have better stitching performance at the pixel level. In summary, Figure 13 reiterates two key points, the necessity of orthorectification before stitching and the accuracy of the proposed stitching algorithm.

4. Conclusions

This paper proposes an orthorectification-based stitching method to obtain large regional mapping by stitching WPP blocks. It mainly consists of two parts, calculation of RPCs iteration to ortho-rectify the original images and preservation of linear structures within stitching WPP images based on line-point consistency. Compared with existing RS stitching method, our method exhibits higher generality and can prevent high accumulative errors. It requires for less sensor information and simplifies the computational complexity. Simulation experiments prove it has superior and robust application prospects in complex scenes such as different viewpoints caused by changes in swing angles. And the accumulative experiments prove it's applicable to produce accurate results due to stable GSD. Hopefully, it can provide a novel and simplified idea for achieving regional mapping within the remote sensing community.

Although the proposed method exhibits superior adaption to swing angle differences and robustness during multiple stitching, it is still sensitive to non-linear intensity because point-feature detection methods adopted in our method, SIFT and SURF, are based on first-order-gradient. Accordingly, our future research will include the exploration of this limitation more deeply. For example, it is of great significance to establish second-order gradient-based methods for improved point detection, such as SFOC and CFOG.

Acknowledgements

The authors are sincerely grateful for the constructive comments and suggestions of the manuscript reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Natural Science Foundation of Jilin Province under Grant 20210101099JC.

ORCID

Ziming Tu http://orcid.org/0000-0002-4241-5410

References

Adel, E., M. Elmogy, and H. Elbakry. 2015. "Image Stitching System Based on ORB Feature-Based Technique and Compensation Blending." *International Journal of Advanced Computer Science & Applications* 6 (9). https://doi.org/10.14569/ijacsa.2015.060907.

Brown, M., and D. G. Lowe. 2007. "Automatic Panoramic Image Stitching Using Invariant Features." *International Journal of Computer Vision* 74 (1): 59–73. https://doi.org/10.1007/s11263-006-0002-3.

Chen, Y. S., and Y. Y. Chuang. 2016. Natural Image Stitching with the Global Similarity Prior. Paper presented at the 14th European Conference on Computer Vision (ECCV), Amsterdam, NETHERLANDS, Oct 08-16.

Cui, J., M. Liu, Z. Zhang, S. Yang, and J. Ning. 2021. "Robust UAV Thermal Infrared Remote Sensing Images Stitching via Overlap-Prior-Based Global Similarity Prior Model." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14:270–282. https://doi.org/10.1109/jstars.2020.3032011.

Dennison, P. E., and D. A. Roberts. 2003. "Endmember Selection for Multiple Endmember Spectral Mixture Analysis Using Endmember Average RMSE." *Remote Sensing of Environment* 87 (2–3): 123–135. https://doi.org/10.1016/s0034-4257(03)00135-4.

Fan, B., F. C. Wu, Z. Y. Hu, and Ieee. 2010. Line Matching Leveraged by Point Correspondences. Paper presented at the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, Jun 13-18.

Fan, B., F. C. Wu, and Z. Y. Hu. 2012. "Robust Line Matching Through Line-Point Invariants." *Pattern Recognition* 45 (2): 794–805. https://doi.org/10.1016/j.patcog.2011.08.004.

Fan, J., Y. Ye, J. Li, G. Liu, and Y. Li. 2022. "A Novel Multiscale Adaptive Binning Phase Congruency Feature for SAR and Optical Image Registration." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–16. https://doi.org/10.1109/tgrs.2022.3206804.

Fischler, M. A., and R. C. Bolles. 1981. "RANDOM SAMPLE CONSENSUS - A PARADIGM for MODEL-FITTING with APPLICATIONS to IMAGE-ANALYSIS and AUTOMATED CARTOGRAPHY." *Communications of the Acm* 24 (6): 381–395. https://doi.org/10.1145/358669.358692.

Fraser, C. S., and H. B. Hanley. 2005. "Bias-Compensated RPCs for Sensor Orientation of High-Resolution Satellite Imagery." *Photogrammetric Engineering and Remote Sensing* 71 (8): 909–915. https://doi.org/10.14358/pers.71.8.909.

Gao, J. H., S. J. Kim, M. S. Brown, and Ieee. 2011. Constructing Image Panoramas Using Dual-Homography Warping. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, Jun 20-25.

Grodecki, J., and G. Dial. 2003. "Block Adjustment of High-Resolution Satellite Images Described by Rational Polynomials." *Photogrammetric Engineering and Remote Sensing* 69 (1): 59–68. https://doi.org/10.14358/pers.69.1.59.

Hao, M., M. C. Zhou, and L. P. Cai. 2021. "An Improved Graph-Cut-Based Unsupervised Change Detection Method for Multispectral Remote Sensing Images." *International Journal of Remote Sensing* 42 (11): 4005–4022. https://doi.org/10.1080/01431161.2021.1881182.

Huang, D., C. Zhu, Y. H. Wang, and L. M. Chen. 2014. "HSOG: A Novel Local Image Descriptor Based on Histograms of the Second-Order Gradients." *IEEE Transactions on Image Processing* 23 (11): 4680–4695. https://doi.org/10.1109/tip.2014.2353814.

Hui, W., and J. Chen. 2020. *UAV Remote Sensing Image Stitching*, 352–360. Springer International Publishing. https://doi.org/10.1007/978-3-030-32456-8_38.

Igarashi, T., and Y. Igarashi. 2009. "Implementing As-Rigid-As-Possible Shape Manipulation and Surface Flattening." *Journal of Graphics GPU and Game Tools* 14 (1): 17–30. https://doi.org/10.1080/2151237X.2009.10129273.

Jia, Q., X. Gao, X. Fan, Z. Luo, L. Haojie, and Z. Chen. 2016. Novel Coplanar Line-Points Invariants for Robust Line Matching Across Views. Paper presented at the 14th European Conference on Computer Vision (ECCV), Amsterdam, NETHERLANDS, OCT 08-16, 2016.

Jia, Q., Z. Li, X. Fan, H. Zhao, S. Teng, X. Ye, and L. Jan Latecki. 2021. Leveraging Line-Point Consistence to Preserve Structures for Wide Parallax Image Stitching. Paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, Nashville, TN, USA.

Jiang, Y., K. Xu, R. Zhao, G. Zhang, K. Cheng, and P. Zhou. 2017. "Stitching Images of Dual-Cameras Onboard Satellite." *ISPRS Journal of Photogrammetry and Remote Sensing* 128:274–286. https://doi.org/10.1016/j.isprsjprs.2017.03.018.

Liao, T., and L. Nan. 2020. "Single-Perspective Warps in Natural Image Stitching." *IEEE Transactions on Image Processing* 29:724–735. https://doi.org/10.1109/tip.2019.2934344.

Lin, C. C., S. U. Pankanti, K. N. Ramamurthy, A. Y. Aravkin, and Ieee. 2015. Adaptive As-Natural-As-Possible Image Stitching. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, Jun 07-12.

Li, L. L., and C. L. Tan. 2010. "Recognizing Planar Symbols with Severe Perspective Deformation." *Ieee Transactions on Pattern Analysis and Machine Intelligence* 32 (4): 755–762. https://doi.org/10.1109/tpami.2009.196.

Liu, S., F. Lin, X. Tong, H. Zhang, H. Lin, H. Xie, Y. Zhen, and S. Zheng. 2022. "Dynamic PSF-Based Jitter Compensation and Quality Improvement for Push-Broom Optical Images Considering Terrain Relief and the TDI Effect." *Applied Optics* 61 (16): 4655–4662. https://doi.org/10.1364/AO.453163 .

Li, L., J. Yao, L. Xiaohu, T. Jinge, and J. Shan. 2016. "Optimal Seamline Detection for Multiple Image Mosaicking via Graph Cuts." *ISPRS Journal of Photogrammetry and Remote Sensing* 113:1–16. https://doi.org/10.1016/j.isprsjprs.2015.12.007.

Lowe, D. G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60 (2): 91–110. https://doi.org/10.1023/b:visi.0000029664.99615.94.

Mayaux, P., G. F. De Grandi, Y. Rauste, M. Simard, and S. Saatchi. 2002. "Large-Scale Vegetation Maps Derived from the Combined L-Band GRFM and C-Band CAMP Wide Area Radar Mosaics of Central Africa." *International Journal of Remote Sensing* 23 (7): 1261–1282. https://doi.org/10.1080/01431160110092894.

Megha, V., and K. K. Rajkumar. 2021. Automatic Satellite Image Stitching Based on Speeded Up Robust Feature. Paper presented at the 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, Gujarat, India.

Meng, W., S. Zhu, B. Zhu, and S. Bian. 2013. "The research of TDI-CCDs imagery stitching using information mending algorithm." In *ISPDI 2013 - Fifth International Symposium on Photoelectronic Detection and Imaging*. Vol. 8908. Beijing, China: SPIE.

Miyoshi, G. T., N. N. Imai, A. M. G. Tommaselli, E. Honkavaara, R. Nasi, and E. A. S. Moriya. 2018. "Radiometric Block Adjustment of Hyperspectral Image Blocks in the Brazilian Environment." *International Journal of Remote Sensing* 39 (15–16): 4910–4930. https://doi.org/10.1080/01431161.2018.1425570 .

Phinn, S., M. Stanford, P. Scarth, A. T. Murray, and P. T. Shyy. 2002. "Monitoring the Composition of Urban Environments Based on the Vegetation-Impervious Surface-Soil (VIS) Model by Subpixel Analysis Techniques." *International Journal of Remote Sensing* 23 (20): 4131–4153. https://doi.org/10.1080/01431160110114998.

Rublee, E., V. Rabaud, K. Konolige, G. Bradski, and Ieee. 2011. ORB: An Efficient Alternative to SIFT or SURF. Paper presented at the IEEE International Conference on Computer Vision (ICCV), Barcelona, SPAIN, Nov 06-13.

Tao, C. V., and Y. Hu. 2001. "Use of the Rational Function Model for Image Rectification." *Canadian Journal of Remote Sensing* 27 (6): 593–602. https://doi.org/10.1080/07038992.2001.10854900.

Toutin, T. 2003. "Block Bundle Adjustment of Ikonos In-Track Images." *International Journal of Remote Sensing* 24 (4): 851–857. https://doi.org/10.1080/0143116021000009877.

Tuna, C., G. Unal, and E. Sertel. 2018. "Single-Frame Super Resolution of Remote-Sensing Images by Convolutional Neural Networks." *International Journal of Remote Sensing* 39 (8): 2463–2479. https://doi.org/10.1080/01431161.2018.1425561 .

Wang, L., S. Y. You, U. Neumann, and Ieee. 2008. SUPPORTING RANGE and SEGMENT-BASED HYSTERESIS THRESHOLDING in EDGE DETECTION. Paper presented at the 15th IEEE International Conference on Image Processing (ICIP 2008), San Diego, CA, Oct 12-15.

Wang, T. Y., G. Zhang, D. R. Li, R. S. Zhao, M. J. Deng, T. Zhu, and L. Yu. 2018. "Planar Block Adjustment and Orthorectification of Chinese Spaceborne SAR YG-5 Imagery Based on RPC." *International Journal of Remote Sensing* 39 (3): 640–654. https://doi.org/10.1080/01431161.2017.1388938.

Xie, R. P., J. M. Tu, J. Yao, M. H. Xia, and S. W. Li. 2019. "A Robust Projection Plane Selection Strategy for UAV Image Stitching." *International Journal of Remote Sensing* 40 (8): 3118–3138. https://doi.org/10.1080/01431161.2018.1539273 .

Xu, Q., L. B. Luo, J. Chen, W. P. Gong, D. H. Guo, and Ieee. 2020. UAV IMAGE MOSAICING BASED MULTI-REGION LOCAL PROJECTION DEFORMATION. Paper presented at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Electr Network, Waikoloa, HI, USA, Sep 26Oct 02.

Ye, Y. X., L. Bruzzone, J. Shan, F. Bovolo, and Q. Zhu. 2019. "Fast and Robust Matching for Multimodal Remote Sensing Image Registration." *IEEE Transactions on Geoscience and Remote Sensing* 57 (11): 9059–9070. https://doi.org/10.1109/tgrs.2019.2924684.

Yeh, C. K., and V. J. D. Tsai. 2011. SELF-CALIBRATED DIRECT GEO-REFERENCING of AIRBORNE PUSHBROOM HYPERSPECTRAL IMAGES. Paper presented at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, CANADA, Jul 24-29.

Ye, Y., B. Zhu, T. Tang, C. Yang, X. Qizhi, and G. Zhang. 2022. "A Robust Multimodal Remote Sensing Image Registration Method and System Using Steerable Filters with First- and Second-Order Gradients." *ISPRS Journal of Photogrammetry and Remote Sensing* 188:331–350. https://doi.org/10.1016/j.isprsjprs.2022.04.011.

Yuan, X., Y. Zheng, W. Zhao, S. Jiongming, and W. Jianzhai. 2021. "Image Stitching Method by Multi-Feature Constrained Alignment and Colour Adjustment." *IET Image Processing* 15 (7): 1499–1507. https://doi.org/10.1049/ipr2.12120.

Zaragoza, J., T. J. Chin, M. S. Brown, D. Suter, and Ieee. 2013. As-Projective-As-Possible Image Stitching with Moving DLT. Paper presented at the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, Jun 23-28.

Zhang, X. S., R. T. Feng, X. H. Li, H. F. Shen, and Z. X. Yuan. 2022. "Block Adjustment-Based Radiometric Normalization by Considering Global and Local Differences." *Ieee Geoscience and Remote Sensing Letters* 19:5. https://doi.org/10.1109/lgrs.2020.3031398.

Zhang, L. L., and R. Koch. 2013. "An Efficient and Robust Line Segment Matching Approach Based on LBD Descriptor and Pairwise Geometric Consistency." *Journal of Visual Communication and Image Representation* 24 (7): 794–805. https://doi.org/10.1016/j.jvcir.2013.05.006 .

Zhang, Y. J., Y. Wan, X. H. Huang, and X. Ling. 2016. "DEM-Assisted RFM Block Adjustment of Pushbroom Nadir Viewing HRS Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 54 (2): 1025–1034. https://doi.org/10.1109/tgrs.2015.2472498.

Zhang, D., L. Yuan, S. Wang, H. Yu, C. Zhang, H. Daogang, G. Han, J. Wang, and Y. Wang. 2019. "Wide Swath and High Resolution Airborne HyperSpectral Imaging System and Flight Validation." *Sensors* 19 (7): 1667. https://doi.org/10.3390/s19071667.

Zhang, Y. J., M. T. Zheng, X. D. Xiong, and J. X. Xiong. 2015. "Multistrip Bundle Block Adjustment of ZY-3 Satellite Imagery by Rigorous Sensor Model without Ground Control Point." *Ieee Geoscience and Remote Sensing Letters* 12 (4): 865–869. https://doi.org/10.1109/lgrs.2014.2365210.

Zhou, G., and X. Liu. 2022. "Orthorectification Model for Extra-Length Linear Array Imagery." *IEEE Transactions on Geoscience and Remote Sensing* 60:1–10. https://doi.org/10.1109/tgrs.2022.3223911.

Zhu, B., C. Yang, J. Dai, J. Fan, Y. Qin, and Y. Ye. 2023. "R_2FD_2: Fast and Robust Matching of Multimodal Remote Sensing Images via Repeatable Feature Detector and Rotation-Invariant Feature Descriptor." *IEEE Transactions on Geoscience and Remote Sensing* 61:1–15. https://doi.org/10.1109/tgrs.2023.3264610.