

Article

An SRTM-Aided Epipolar Resampling Method for Multi-Source High-Resolution Satellite Stereo Observation

Jingwen Hu ¹, Gui-Song Xia ²  and Hong Sun ^{1,*}¹ Electronic Information School, Wuhan University, Wuhan 430072, China; hujingwen@whu.edu.cn² State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan 430079, China; guisong.xia@whu.edu.cn

* Correspondence: hongsun@whu.edu.cn

Received: 14 February 2019; Accepted: 17 March 2019; Published: 21 March 2019



Abstract: Binocular stereo observation with multi-source satellite images used to be challenging and impractical, but is now a valuable research issue with the introduction of powerful deep-learning-based stereo matching approaches. However, epipolar resampling, which is critical for binocular stereo observation, has rarely been studied with multi-source satellite images. The main problem is that, under the multi-source stereo mode, the epipolar-line-direction (ELD) at an image location may vary when computed with different elevations. Thus, a novel SRTM (Shuttle Radar Topography Mission)-aided approach is proposed, where a point is transformed from the original image-space to the epipolar image-space through a global rotation, followed by a block-wise homography transformation. The global rotation transfers the ELDs at the center of the overlapping area to the x -axis, and then block-wise transformation shifts the ELDs of all grid-points to the x -axis and eliminates the y -disparities between the virtual corresponding points. Experiments with both single-source and multi-source stereo images showed that the proposed method is obviously more accurate than the previous methods that do not use SRTM. Moreover, with some of the multi-source image pairs, only the proposed method ensured the y -disparities remained within ± 1 pixel.

Keywords: epipolar resampling; high-resolution satellite image; stereo observation; SRTM; dense image-matching; RFM; geometric transformation

1. Introduction

Stereo observation has long been interested in computer vision, photogrammetry, and remote sensing. Conventional stereo observation depends on binocular stereo matching in image-space [1] or multi-view stereo matching in object-space [2,3]. Recently, even a monocular view can be used to reconstruct depth information [4–6]. However, binocular stereo matching is still the most widely used technique in industry, since its complexity is controllable and its results are stable. As the pre-step of binocular stereo observation, epipolar resampling eliminates the y -disparities between conjugate points, so that a left-image-point along the same x -axis can be matched on the right image. Afterwards, the x -disparity map is obtained to build the depth maps or the 3D models. Epipolar resampling greatly improves computational efficiency by facilitating the image matching task to be considered a one-dimensional correspondence-labeling task. Thus, many valuable stereo matching approaches have been inspired, among which the most widely used is semi-global matching (SGM) [7]. The SGM method and its variants are now widely used in aerial/UAV photogrammetry and along-track satellite stereo observation.

With the rapid development of high-resolution remote sensing technologies, more and more users are choosing cheaper and more flexible high-resolution along-track satellite stereo images to

replace the conventional aerial images when producing 3D digital terrain models (DTM). Currently, the stereo satellite images are mainly collected by two types of systems: multi-line-array sensors, which collect images from different viewing angles simultaneously (e.g., ZY-3 and TH-1, from China), and single-line-array sensors, which observe one place two times with different pitch angles (e.g., IKONOS, GeoEye-1, WorldView-1/2/3, and Pleiades). The along-track stereo images acquired by either of the two systems have time intervals of only seconds, which means the radiative conditions and the land-cover are almost the same on different views. The latest along-track stereo satellite imagery has been demonstrated to have very high imaging quality and very good accuracy in both the horizontal and altitude directions [8,9].

In addition to the along-track stereo satellite images, the images collected from different tracks can also be used for cross-track stereo observation as long as they have large enough convergence angles ($>10^\circ$). Cross-track stereo observation is very useful in producing high-relative-accuracy digital ortho-maps (DOM), because the high-resolution digital elevation model (DEM) is inaccessible to most users since it is either too expensive or restricted by the government. With inaccurate DEMs, the DOMs produced by overlapping satellite images with large convergence angles will have poor mosaicking accuracy. The mosaicking error is randomly distributed and can only be eliminated with more accurate elevation. However, unlike along-track stereo observation, cross-track stereo observation is very challenging. The images from different tracks may be acquired in different years or seasons, which have different landcovers and different radiative properties. Meanwhile, some stereo image pairs are acquired by multi-sensors. Multi-source satellite images also have different geometric properties. As a result, the SGM-based methods can hardly obtain satisfying results, and therefore cross-track or multi-source satellite stereo observation has rarely been applied before. Recently, the introduction of powerful CNN-based stereo matching approaches [10,11] has brought new hope to this area. As the pre-step of the stereo observation, epipolar resampling of multi-source satellite stereo observations should be carried out. Currently, there is no research showing that an epipolar resampling method can work well with multi-source satellite stereo images, especially when their track directions are quite different.

In this paper, a novel SRTM-aided epipolar resampling method, using a global rotation followed by a block-wise homography transformation, is proposed, and is demonstrated with both along-track and cross-track as well as single-source and multi-source stereo observations. The proposed method uses the pre-corrected rational function model (RFM) and the elevations supplied by SRTM to compute accurate epipolar-line-directions (ELDs). The experiments used 13 pairs of stereo images acquired by ZY-3, GF-1, and GF-2 from China; IKONOS-2, WorldView-2, and WorldView-3 from Digital Globe Inc. of the USA; and Pleiades-HR 1a and Pleiades-HR 1b from Europe.

2. Related Works

In this section, we briefly review the existing studies on the model of epipolarity between binocular linear pushbroom satellite images and the existing approaches of epipolar resampling.

The epipolar geometry of frame images can be modeled in a very simple manner through the fundamental matrix F , which can be recovered with only eight conjugate points, or even fewer under different conditions [12,13]. The epipolar geometry of frame images is based on the coplanarity of the two perspective centers, two conjugate image points, and the corresponding object-space point. The intersections of the epipolar plane and the two image planes result in two straightforward corresponding epipolar lines. Thus, epipolar resampling of frame images can also be implemented with very simple geometric transformations.

Unlike the frame imagery, the pushbroom satellite imagery has a very complicated rigorous epipolar relationship. The pushbroom images are formed by stitching the one-dimensional scanning lines captured as the sensor moves. Each scanning line has a unique perspective center and a unique camera attitude, since the satellite is moving in an elliptical orbit with varying velocity and is slowly rotating to keep the sensors facing the ground [14]. It has been pointed out that, when modeling the exterior orientation parameters as second-order polynomials, the epipolar curve can be approximated

as a straight line in a small range but not for the entire pushbroom image. Moreover, different points on the projection of a light ray may correspond to different epipolar curves, which implies that there is actually no conjugate epipolar curve on pushbroom images [15].

When undertaking epipolar resampling for pushbroom satellite images, the projection model is usually simplified. Some works [16,17] have used linear approximation with the pushbroom imaging model, which assumes that, over the acquisition duration of one image scene, the sensor is moving in a straight line with a constant velocity and the attitude of the sensor is fixed. With the approximation, the projection model can be recovered with several ground control points (GCPs), and the epipolar curves can be represented as hyperbola curves. Some works [18,19] have used the even simplified parallel projection model for high-resolution satellite imagery with a very narrow field-of-view (FOV). The parallel-projected image can be transformed from the original perspective-projected image through a simple image-space rectification. Under the parallel projection, the epipolar curves are straight lines and have unique correspondences. Thus, epipolar-resampling with the parallel-projection occurs in a similar manner to the frame images. Some researchers have found that the ELD is much more stable on what they called the projection reference plane than on the image plane. The projection reference plane is a virtual horizontal plane in object space [20,21]. In these works, the images were firstly ortho-rectified onto the projection reference plane, and then transformed with a 2D affine model that rotates the images to make the ELD parallel to the x -axis. The georeferencing models used in these methods should be pre-corrected to eliminate the convergence error of conjugate points. Either the rigorous sensor model [22] or the rational function model (RFM) [23] can be used as the projection model. In recent years, some works have directly used the pre-corrected RFM as the projection model, without making any simplification when modeling the epipolar relationship. In other works [24,25], the authors used 2D polynomial functions calculated with conjugate image points to transform the epipolar curves into straight lines. In another paper [26], the authors used block-wise transformation to transform the epipolar curves into straight lines.

The above-mentioned methods make some assumptions to simplify the pushbroom projection model based on the property of high-resolution satellite imagery as having narrow FOVs. Most of the methods have been demonstrated to be effective with along-track stereo images. Wang et al. [21] also demonstrated their method with cross-track SPOT-5-HRG images, which have intervals of several days between the acquisition dates. However, multi-source satellite stereo image pairs are much more complicated to deal with [27–32]. None of the methods have been demonstrated to be effective with multi-source imagery, and there is a lack of analysis on the accuracy loss of their assumptions, especially for those assuming that a unified ELD exists and the epipolar curves are straight. In reality, the ELD changes not only with different image locations, but also with different back-projection elevations at the same image location. Moreover, the angles between the orbits of different satellites may be large, and thus the simplifications may be inaccurate. To deal with these problems, a new SRTM-aided method is proposed in this article, and is demonstrated to be effective with both single-source and multi-source stereo satellite images.

3. Epipolar Geometry of Pushbroom Imagery

In this section, we briefly explain why medium-accuracy elevation should be used in computing the ELD for pushbroom images, and describe the process of obtaining the ELD and the corresponding epipolar-line-segment for an image-space location with pre-corrected RFMs.

3.1. Property of Epipolar-Line-Direction

Define $(\mathbf{p}, \mathbf{p}')$ as a pair of conjugate points on two-view pushbroom stereo images, where \mathbf{p} is on the left image I and \mathbf{p}' is on the right image I' . Define \mathbf{P} as the corresponding object-space point of \mathbf{p} and \mathbf{p}' . The ELD at the position of \mathbf{p} , denoted as $\overrightarrow{\text{ed}}(\mathbf{p})$, is defined as the direction of the left-image-projection of a small segment, which is on the light of \mathbf{p}' and crosses the object-space point \mathbf{P} . When the two-view images are relatively well oriented, the point \mathbf{p}' will be lying on the

right-image-projection of the light of \mathbf{p} . For frame images, the ELD $\vec{ed}(\mathbf{p})$ can be obtained without knowing the actual location of \mathbf{p}' because the coplanarity conveyed by the perspective projection ensures the $\vec{ed}(\mathbf{p})$ is not changed with different \mathbf{p}' 's on the right-image-projection of the light of \mathbf{p} . However, a similar property is not established with two-view pushbroom images. Figure 1 illustrates four typical types of geometric relationships of pushbroom stereo images. For the convenience of analysis, it is assumed that the satellites are moving in straight lines with constant velocities, and that the attitudes of the sensors are not changed during the acquisitions.

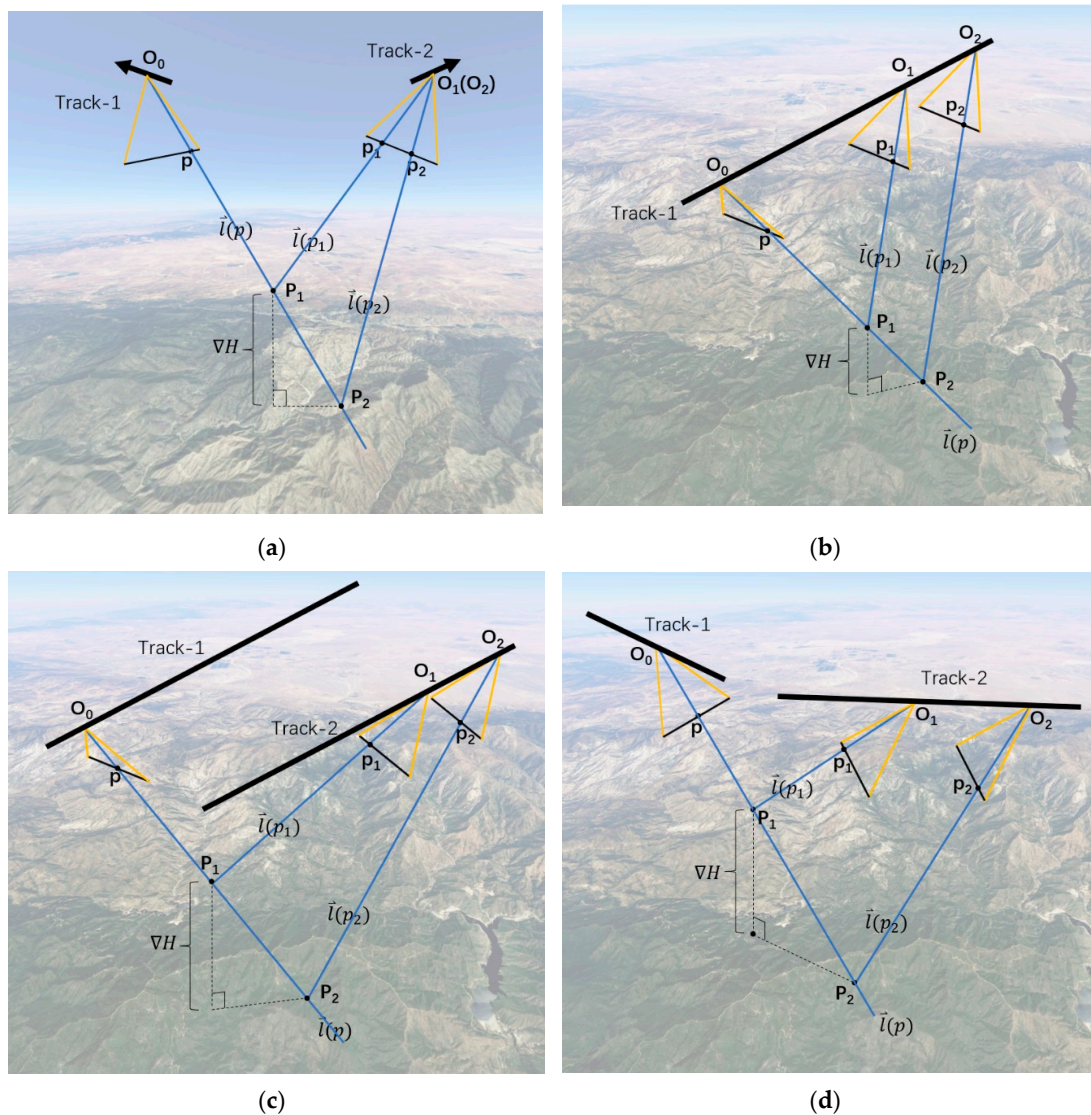


Figure 1. Four typical types of geometric relationships of pushbroom stereo images: (a) cross-track coplanar stereo; (b) along-track stereo; (c) cross-track non-coplanar stereo; and (d) cross-multi-directional-track stereo.

The condition in Figure 1a is called cross-track coplanar stereo, where Track-1 and Track-2 are approximately parallel, and the scanning planes of the two tracks are approximately parallel. As a result, different points on the light of \mathbf{p} , denoted as $\vec{l}(p)$, correspond to the same perspective center. Thus, the points \mathbf{O}_0 , \mathbf{p} , \mathbf{P}_1 , \mathbf{P}_2 , \mathbf{p}_1 , \mathbf{p}_2 , and \mathbf{O}_1 (\mathbf{O}_2) are in a same plane, and the ELD is fixed at the direction of the x -axis on both images. The condition in Figure 1b is called along-track stereo, where only one track exists and is lying on the plane formed by points \mathbf{O}_0 , \mathbf{p} , \mathbf{P}_1 , \mathbf{P}_2 , \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{O}_1 , and \mathbf{O}_2 . As a result, the ELD is fixed at the direction of the y -axis in the image-space. Apparently,

under the cross-track coplanar stereo condition and the along-track stereo condition, the ELD will not be influenced by different corresponding points, or, in other words, by different heights. This phenomenon explains why methods which do not consider the influence of different heights have still achieved relatively good results with some of the cross-track stereo images, and all of the along-track stereo images. Stereo observation with images captured by satellites that obtain nadir images with a fixed pitch angle (e.g., multi-track ZY3-NAD, GF-1, or GF-2) follows the cross-track coplanar stereo condition. The along-track stereo condition corresponds to the most widely used along-track stereo images captured by multi-line-array sensors (e.g., ZY3-TLC and TH1-TLC), or a single-line-array sensor with different pitch angles (e.g., IKONOS stereo, GeoEye stereo, and WorldView stereo).

Under the conditions shown in Figure 1c,d, coplanarity no longer established, thus the ELD will be influenced by different heights and there are no conjugate epipolar curves. Condition (c) is called the cross-track non-coplanar stereo, which usually happens for image pairs made up with a forward-view and a backward-view captured on different tracks (e.g., stereo observation with a ZY3 forward-view and a ZY-3 nadir-view in different tracks). Condition (d) is called the cross-multi-directional-track (CMDT) stereo, which is a special case of the cross-track non-coplanar condition. The CMDT condition usually happens for multi-source stereo images. In the experimental section of this paper, it is shown that there exist about 20 degree angles between the tracks of a Chinese mapping satellite image (e.g., ZY3, GF1, or GF2) and a high-resolution satellite image provided by DigitalGlobe, Inc. (e.g., IKONOS or WorldView-2). For better analysis, the image pairs following the CMDT condition are tested.

In summary, the objective of epipolar resampling for frame images—that is, transforming the conjugate epipolar-curves to the same row—can never be achieved under the cross-track non-coplanar stereo condition or CMDT stereo condition of pushbroom images, since the conjugate epipolar-curves do not exist [15]. However, for the purpose of digital terrain model generating, we only need to focus on the ELD computed with the elevations near the ground. Many public data sources can supply medium-accuracy elevation information, such as the elevation information supplied by the Shuttle Radar Topography Mission (SRTM), which has a better-than-100 m accuracy in most areas and covers land all over the world within the latitude range from 54°S to 60°N [33]. As a result, in the generic epipolar resampling method proposed in this paper, an accurate ELD is determined with the elevation supplied by the SRTM.

3.2. Obtaining Epipolar-Line-Directions

In this section, we introduce the geometric process used in the proposed epipolar resampling approach. When the stereo satellite images I (the left image) and I' (the right image) have been relative-oriented accurately, the lights of the two conjugate points \mathbf{p} and \mathbf{p}' will intersect somewhere in the object-space. As a result, the epipolar-curve $\mathbf{ec}(\mathbf{p})$ corresponding to the left-image-point \mathbf{p} can be easily obtained by projecting the light of \mathbf{p} onto the right image-space, as $\mathbf{ec}(\mathbf{p})$ will cross the point \mathbf{p}' . Similarly, the epipolar-curve $\mathbf{ec}(\mathbf{p}')$ which crosses the point \mathbf{p} can be obtained by projecting the light of \mathbf{p}' .

The light of \mathbf{p} or \mathbf{p}' can be obtained by the process of backprojection with the RFM model [23]. The RFM model projects the geographic coordinates (L, P, H) to the image-space coordinates (x, y) with two fractions of three-order polynomials, where L is latitude; P is longitude; H is height; x is the column of the image (the sample coordinate); and y is the row of the image (the number of scanning lines). When fixing the height H , the RFM can also be used to calculate the horizontal coordinates (latitude and longitude) with the image-space coordinates (x, y) through solving a set of non-linear equations, which is known as the backprojection process. For the convenience of presentation, the process of projection is denoted as:

$$(L, P, H) \xrightarrow{proj} (x, y) \quad (1)$$

and the process of backprojection is denoted as:

$$(x, y, H) \xrightarrow{backproj} (L, P) \quad (2)$$

Before obtaining the ELDs, we need to find the corresponding points. Some approaches [24,25] directly use the matched tie-points. However, this may not be a good choice for production since the resampling accuracy will be influenced by the matching error and the spatial distribution of the tie-points. In the proposed approach, we chose to use the pre-corrected RFMs and the SRTM to obtain virtual corresponding points. Although the SRTM has tens of meters of elevational error, the ELDs are quite accurate. Because the satellites are relatively stable carriers, the curvature of the epipolar curves on the high-resolution satellite images are very small, and treating a small segment of an epipolar curve as a straight line-segment will not cause the loss of any accuracy. The corresponding point of \mathbf{p} , denoted as \mathbf{p}' , is determined in two steps. First, backproject the point \mathbf{p} onto the elevation plane of SRTM through an iteration process [34] (which is called the ray-tracing process in computer graphics). Thus, the ground point \mathbf{P} and its elevation H are obtained. Second, project the ground point \mathbf{P} onto the right image \mathbf{I}' and, thus, the corresponding point \mathbf{p}' is obtained. The iterative process of backprojecting the \mathbf{p} onto the elevation plane is denoted as:

$$(x, y, SRTM) \xrightarrow{\text{ray-tracing}} (L, P, H) \tag{3}$$

The process of obtaining the corresponding point with the aid of SRTM is denoted as:

$$(\mathbf{p}, SRTM) \xrightarrow{\text{ray-tracing}} \mathbf{P} \xrightarrow{\text{proj}} \mathbf{p}' \tag{4}$$

Then, the ELD $\overrightarrow{\text{ed}}(\mathbf{p})$ at the location of \mathbf{p} is obtained by projecting a small segment of the light (denoted as $\overline{\mathbf{P}'_-\mathbf{P}'_+}$) corresponding to \mathbf{p}' around the ground point \mathbf{P} onto the left image \mathbf{I} (see Figure 2). The segment $\overline{\mathbf{P}'_-\mathbf{P}'_+}$ is obtained through the backprojection process (Equation (2)) on \mathbf{p}' with the heights $(H-\nabla H)$ and $(H+\nabla H)$. ∇H , the height interval between point \mathbf{P} and \mathbf{P}'_- or \mathbf{P}'_+ , can be set as 20–100 m. Conversely, the ELD $\overrightarrow{\text{ed}}(\mathbf{p}')$ at the location of \mathbf{p}' is obtained by similar steps to the point \mathbf{p} . The process of obtaining the ELD of \mathbf{p} is denoted as:

$$(\mathbf{p}, \mathbf{p}', \mathbf{P}, \nabla H) \xrightarrow{\text{backproj}} \overline{\mathbf{P}'_-\mathbf{P}'_+} \xrightarrow{\text{proj}} \overrightarrow{\text{ed}}(\mathbf{p}) \tag{5}$$

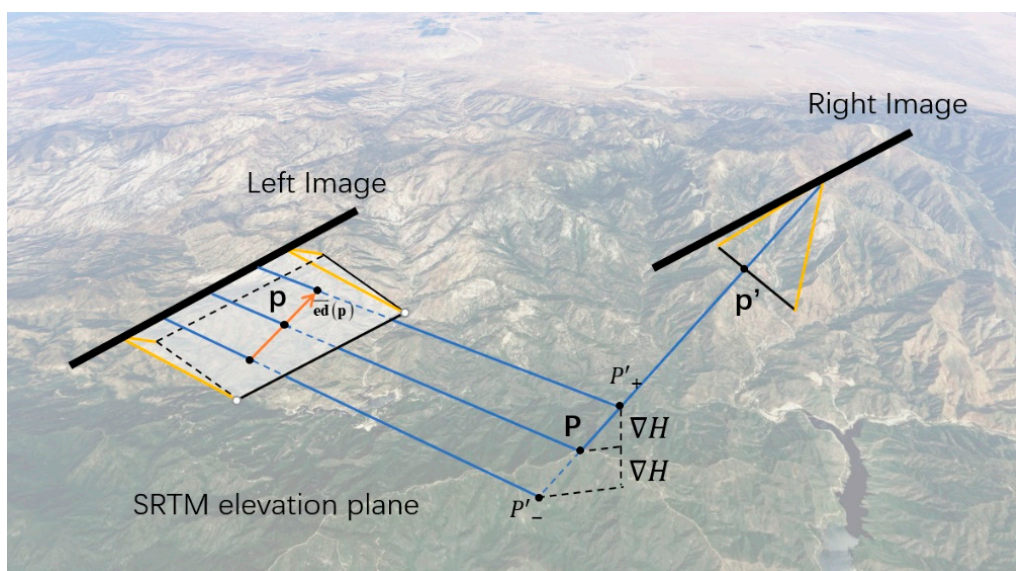


Figure 2. The process of obtaining the corresponding point and the ELD with the aid of SRTM.

4. Proposed Approach

The proposed approach was designed with two objectives in mind: (1) Eliminate the y -disparities between the image points and their correspondences obtained through the process in Equation (4).

(2) Transform the ELDs obtained through the process in Equation (5) at all image locations to the positive direction of the x -axis.

4.1. Workflow

The proposed approach has three steps:

1. Prepare the global rotation matrix and the block-wise homography transformation models.
2. Evaluate the prepared transformation model through virtual corresponding points.
3. Perform epipolar resampling.

In the proposed approach, a point either on the left image or on the right image needs two transformations from the original image-space to the epipolar image-space (see Figure 3). First, the point in the original image-space is rotated into the rotated image-space. In the rotated image space, only the ELD at the center of the overlapping area is ensured to be in the direction of the x -axis. Second, a block-wise homography transformation is made to transform the point from the rotated image-space to the epipolar image-space. Conversely, a point is transformed from the epipolar image-space by a reverse block-wise homography transformation, followed by a reverse global rotation.

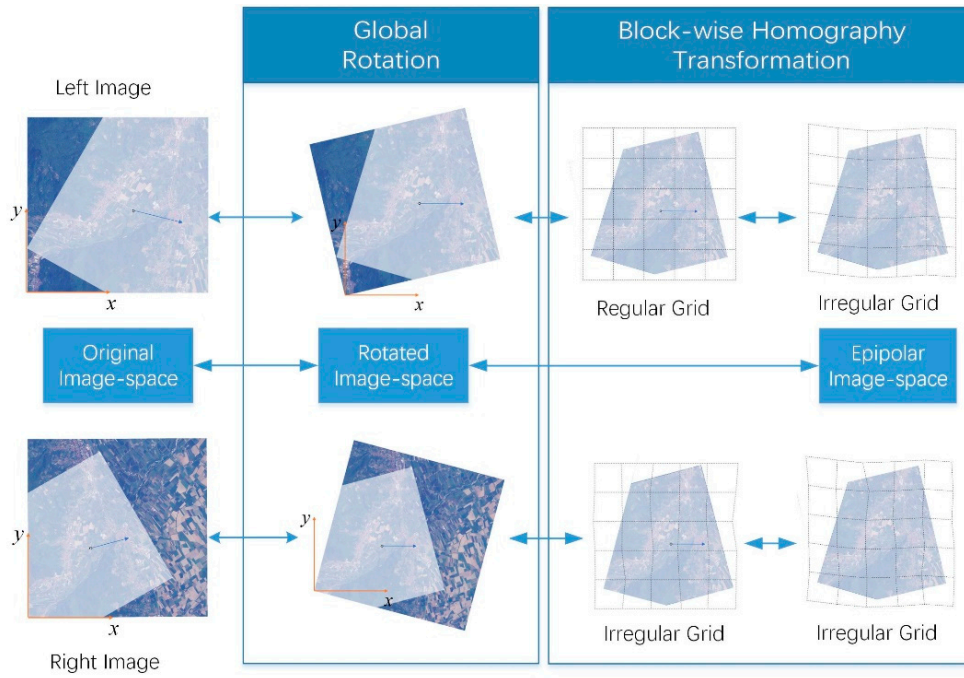


Figure 3. The process of transforming points between the original image-space and the epipolar image-space in the proposed approach.

The rotation matrix can be easily constructed according to the ELD computed with the process described in Section 3.1. Denote $\theta_c^{(org)}$ and $\theta'_c^{(org)}$ as the angle between the x -axis and the ELD at the center of the overlapping area on the left image and the right image, then the global rotation matrixes of the left image \mathbf{I} and the right image \mathbf{I}' are:

$$\mathbf{R} = [\mathbf{r} | \mathbf{t}] = \begin{bmatrix} \cos \theta_c^{(org)} & \sin \theta_c^{(org)} & t_x \\ -\sin \theta_c^{(org)} & \cos \theta_c^{(org)} & t_y \end{bmatrix} \quad (6)$$

$$\mathbf{R}' = [\mathbf{r}' | \mathbf{t}'] = \begin{bmatrix} \cos \theta'_c^{(org)} & \sin \theta'_c^{(org)} & t'_x \\ -\sin \theta'_c^{(org)} & \cos \theta'_c^{(org)} & t'_y \end{bmatrix}$$

where the translating parameters \mathbf{t} and \mathbf{t}' are included to ensure the entirety of the rotated overlapping areas are in a proper image-coordinate for resampling.

After the global rotation, we need to obtain the corresponding grid locations at the rotated image-space and the epipolar image-space, which is the most complicated part of the proposed approach. The workflow is illustrated in Figure 4. After that, the homography transformation within a block can be easily solved with its four grid-point-correspondences.

Finally, when the transformation models are prepared, the y -disparities of the virtual corresponding points—which are transformed to the epipolar image-space—are used to examine the accuracy of the models before image resampling.

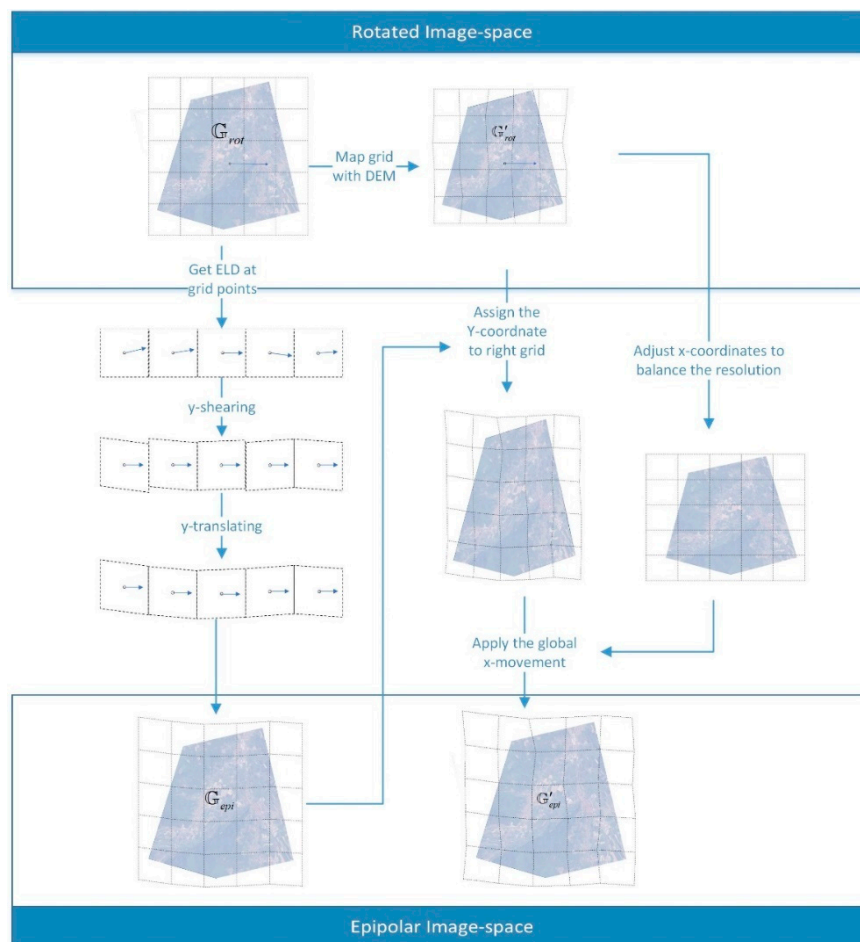


Figure 4. The process of transforming the grid from the rotated image-space to the epipolar image-space in the proposed approach.

4.2. Block-Wise Transformation of the Left Image

As explained in Section 3, the ELD may be influenced by different heights under the cross-track non-coplanar stereo condition and the CMDT stereo condition, which means the ELDs at different image locations may show small differences because of the uneven land. As a result, the block-wise homography transformation is chosen to ensure the ELDs at all the left-image locations point to the positive direction of the x -axis.

Denote the grids of the left image in the rotated image-space as:

$$\mathbb{G}_{rot} = \left\{ \mathbf{g}_{i,j}^{(rot)} \mid i = 0, 1, 2, \dots, m \text{ and } j = 0, 1, 2, \dots, n \right\} = \left\{ \left(x_{i,j}^{(rot)}, y_{i,j}^{(rot)} \right) \right\} \quad (7)$$

where \mathbb{G}_{rot} is a regular grid formed by squares, and covers the whole overlapping area of the left image in the rotated image-space. Denote the square surrounded by grid $\mathbf{g}_{i,j}^{(rot)}$, $\mathbf{g}_{i+1,j}^{(rot)}$, $\mathbf{g}_{i,j+1}^{(rot)}$, and $\mathbf{g}_{i+1,j+1}^{(rot)}$ as $\mathbf{sq}_{i,j}$. The side length of the squares is denoted as h .

Denote the corresponding grids of \mathbb{G}_{rot} in the epipolar image-space as:

$$\mathbb{G}_{epi} = \{ \mathbf{g}_{i,j}^{(epi)} \} = \{ (x_{i,j}^{(epi)}, y_{i,j}^{(epi)}) \} \tag{8}$$

The key is how to determine the location of $\mathbf{g}_{i,j}^{(epi)}$ s. In the proposed method, the x -coordinate of $\mathbf{g}_{i,j}^{(epi)}$ is inherited from $\mathbf{g}_{i,j}^{(rot)}$. Only the y -coordinates are adjusted to modify the ELDs. Thus, the transformation of grids from the rotated image-space to the epipolar image-space is:

$$\begin{aligned} x_{i,j}^{(epi)} &= x_{i,j}^{(rot)} \\ y_{i,j}^{(epi)} &= y_{i,j}^{(rot)} + \Delta y_{i,j} \end{aligned} \tag{9}$$

where the y -translation $\Delta y_{i,j}$ is determined in the following steps.

Denote the ELD at $\mathbf{g}_{i,j}^{(rot)}$ in the rotated image-space as $\mathbf{ed}_{i,j}^{(rot)}$ and denote the angle between $\mathbf{ed}_{i,j}^{(rot)}$ and the x -axis as $\theta_{i,j}^{(rot)}$. Assume we can transform the neighboring area of $\mathbf{g}_{i,j}^{(rot)}$ from the rotated image-space to the epipolar image-space by a y -shearing and a y -translation:

$$y^{(epi)} = y^{(rot)} - (x^{(rot)} - x_{i,j}^{(rot)}) \cdot \tan \theta_{i,j}^{(rot)} + \Delta y_{i,j} \tag{10}$$

where (x, y) is the coordinate of a point in the neighboring area of $\mathbf{g}_{i,j}^{(rot)}$. The y -shearing turns the ELD at the location of $\mathbf{g}_{i,j}^{(rot)}$ to the x -axis. The purpose of y -translation is to deal with the dislocation at the square center caused by the different y -shearings of its four grid points. For example, at the center of square $\mathbf{sq}_{i,j}$, of which the x -coordinate is $(x_{i,j}^{(rot)} + h/2)$, four different combinations of a y -shearing and a y -translation may be applied according to Equation (9):

$$\begin{aligned} T_{i,j,0,0} &= -(h/2) \cdot \tan \theta_{i,j}^{(rot)} + \Delta y_{i,j}; \text{ with grid } \mathbf{g}_{i,j}^{(rot)} \\ T_{i,j,1,0} &= (h/2) \cdot \tan \theta_{i+1,j}^{(rot)} + \Delta y_{i+1,j}; \text{ with grid } \mathbf{g}_{i+1,j}^{(rot)} \\ T_{i,j,0,1} &= -(h/2) \cdot \tan \theta_{i,j+1}^{(rot)} + \Delta y_{i,j+1}; \text{ with grid } \mathbf{g}_{i,j+1}^{(rot)} \\ T_{i,j,1,1} &= (h/2) \cdot \tan \theta_{i+1,j+1}^{(rot)} + \Delta y_{i+1,j+1}; \text{ with grid } \mathbf{g}_{i+1,j+1}^{(rot)} \end{aligned} \tag{11}$$

The difference between each pair of y -transformations at the center of square $\mathbf{sq}_{i,j}$ should be minimized. There are six difference values, denoted as:

$$\begin{aligned} dT_{i,j,1} &= T_{i,j,0,0} - T_{i,j,0,1} \\ dT_{i,j,2} &= T_{i,j,0,0} - T_{i,j,1,0} \\ dT_{i,j,3} &= T_{i,j,0,0} - T_{i,j,1,1} \\ dT_{i,j,4} &= T_{i,j,0,1} - T_{i,j,1,0} \\ dT_{i,j,5} &= T_{i,j,0,1} - T_{i,j,1,1} \\ dT_{i,j,6} &= T_{i,j,1,0} - T_{i,j,1,1} \end{aligned} \tag{12}$$

As a result, the y -translations $\Delta y_{i,j}$ are computed by solving the following constrained optimization problem:

$$\begin{aligned} \min_{\Delta y_{i,j}} & \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \sum_{k=1}^6 (dT_{i,j,k})^2 \\ \text{s.t.} & \Delta y_{\lfloor m/2 \rfloor, \lfloor n/2 \rfloor} = 0 \end{aligned} \tag{13}$$

where $\Delta y_{\lfloor m/2 \rfloor, \lfloor n/2 \rfloor}$ is the y -translation at the center grid, and the purpose of this constraining condition is to ensure a unique solution.

With the solution of $\Delta y_{i,j}$ s, the grids \mathbb{G}_{rot} in the epipolar image-space are obtained. Finally, the homography transformation $\mathbf{H}_{i,j}$ within the square $\mathbf{sq}_{i,j}$ is computed with the four pairs of grid points $(\mathbf{g}_{i,j}^{(rot)}, \mathbf{g}_{i,j}^{(epi)}), (\mathbf{g}_{i+1,j}^{(rot)}, \mathbf{g}_{i+1,j}^{(epi)}), (\mathbf{g}_{i,j+1}^{(rot)}, \mathbf{g}_{i,j+1}^{(epi)}),$ and $(\mathbf{g}_{i+1,j+1}^{(rot)}, \mathbf{g}_{i+1,j+1}^{(epi)})$.

4.3. Block-Wise Transformation of the Right Image

The block-wise homography transformation of the right image has two objectives: one is to eliminate the y -disparity between the corresponding points $(\mathbf{p}, \mathbf{p}')$, and the other is to balance the ground resolution in the x -direction and the y -direction.

Denote the grid on the right image in the rotated image-space as:

$$\mathbb{G}'_{rot} = \left\{ \mathbf{g}'_{i,j}^{(rot)} \mid i = 0, 1, 2, \dots, m \text{ and } j = 0, 1, 2, \dots, n \right\} = \left\{ \left(x'_{i,j}^{(rot)}, y'_{i,j}^{(rot)} \right) \right\} \tag{14}$$

where $\mathbf{g}'_{i,j}^{(rot)}$ is the corresponding point of the left-image-grid $\mathbf{g}_{i,j}^{(rot)}$ computed in three steps: (1) reverse global rotation on the left image; (2) the backprojection and projection process in Equation (4) with the aid of SRTM; and (3) rotation on the right image. Apparently, \mathbb{G}'_{rot} is an irregular grid and the homography transformation $\mathbf{H}'_{i,j}$ is applied within the polygon $\mathbf{P}'_{i,j}$ formed by the grid point $\mathbf{g}'_{i,j}^{(rot)}, \mathbf{g}'_{i+1,j}^{(rot)}, \mathbf{g}'_{i,j+1}^{(rot)}$ and $\mathbf{g}'_{i+1,j+1}^{(rot)}$.

Denote the corresponding grids of the right image in the epipolar image-space as:

$$\mathbb{G}'_{epi} = \left\{ \mathbf{g}'_{i,j}^{(epi)} \right\} = \left\{ \left(x'^{(epi)}_{i,j}, y'^{(epi)}_{i,j} \right) \right\} \tag{15}$$

The locations of $\mathbf{g}'_{i,j}^{(epi)}$ s are determined as follows:

$$\begin{aligned} x'^{(epi)}_{i,j} &= A \cdot x'^{(rot)}_{i,j} + B \cdot y'^{(rot)}_{i,j} + C \\ y'^{(epi)}_{i,j} &= y'_{i,j} \end{aligned} \tag{16}$$

The y -coordinate of $\mathbf{g}'_{i,j}^{(epi)}$ is inherited from the y -coordinate of $\mathbf{g}_{i,j}^{(epi)}$ (of the left image) to eliminate the y -disparity. The x -coordinate of $\mathbf{g}'_{i,j}^{(epi)}$ is adjusted with a linear transformation to balance the ground resolution. The parameters of the linear transformation $A, B,$ and C are obtained by solving the following optimization problem:

$$\min_{A,B,C} \sum_{i=0}^m \sum_{j=0}^n \left(A \cdot x'^{(rot)}_{i,j} + B \cdot y'^{(rot)}_{i,j} + C - x'^{(epi)}_{i,j} \right)^2$$

Finally, the homography transformation $\mathbf{H}'_{i,j}$ is computed with the four pairs of grid points on the right image: $(\mathbf{g}'_{i,j}^{(rot)}, \mathbf{g}'_{i,j}^{(epi)}), (\mathbf{g}'_{i+1,j}^{(rot)}, \mathbf{g}'_{i+1,j}^{(epi)}), (\mathbf{g}'_{i,j+1}^{(rot)}, \mathbf{g}'_{i,j+1}^{(epi)}),$ and $(\mathbf{g}'_{i+1,j+1}^{(rot)}, \mathbf{g}'_{i+1,j+1}^{(epi)})$.

5. Experiments

Seventeen scenes of satellite images with vendor-provided RFMs were used to demonstrate the proposed epipolar resampling approach. Two methods that can be easily implemented with RFM were used for comparison. For the convenience of listing the results, the proposed approach in this paper is called RBH (rotation and block-wise homography transformation); the approach based on the object-space project-reference-plane [21] is called PRP; and the approach using tie-points (TPs) to solve the 2D-polynomial transformations [24] is called TPP.

5.1. Data and Process

Table 1 lists the meta-information of the satellite images used in the experiments. There were three test areas. The first one was in Guangdong Province, China, covering the city of Guangzhou,

the city of Foshan, and some other areas on the west side of the Pearl River Estuary. The land is mostly plains, but some high hills exist in the north and in the west. The elevation in this area ranges from 0 m to 500 m. The plains are mostly dense urban areas, while heavy vegetation also exists on the hills. Many rivers exist in this area. The second test area included the plains near the city of Nanking, Jiangsu Province, China. The land is mostly plains, while some small hills exist. The elevation in this area ranges from 0 m to 230 m. The area is most rural. Many rivers and lakes exist in this area. The third test area was near the city of Seoul, South Korea. Half of the land is plains, and the other half is mountainous. The elevation in this area ranges from 10 m to 400 m. The plains are dense urban areas, while the mountains are covered by heavy vegetation.

The tested images were acquired by eight different satellites and nine sensors. In Table 1, ZY-3-NAD means a ZY-3 panchromatic nadir image; ZY-3-FWD means a ZY-3 panchromatic forward-view image; and ZY-3-BWD means a ZY-3 panchromatic backward-view image. ZY-3 is the first civil stereo mapping satellite of China. There are about 24 degree viewing-angles fixed between the ZY-3-NAD and the ZY-3-FWD/BWD. The ground sampling distance (GSD) on the ZY-3-NAD images is about 2.1 m, and the GSDs on the ZY-3-FWD/BWD images are about 3.2 m. GF-1-PMS and GF-2-PMS are the panchromatic images of the GF-1 satellite and GF-2 satellite, which were also launched in China. The GSD is about 2.0 m on the GF-1-PMS images and about 0.8 m on the GF-2-PMS images. The product level of the tested images captured by ZY-3, GF-1, and GF-2 is Level-1A, which means only radiative correction is made to the block-wise images, with an RFM generated with the initial rigorous georeferencing model.

Images captured by the panchromatic sensors on the IKONOS-2 and the WorldView-2/3, launched by Digital Globe Inc. from Westminster, Colorado, USA, were also tested. The GSD is about 1.0 m on the IKONOS-2 images, about 0.5 m on the WorldView-2 images, and about 0.4 m on the WorldView-3 images. The product level of the IKONOS-2 images is standard geometrically corrected. The product level of the WorldView-2/3 images is Level-1B. Both levels mean that radiative corrections and geometric corrections were made, and the images were not ortho-rectified with a DEM.

Two images captured by the Pleiades-HR 1A and the Pleiades-HR 1B were used in the South Korea test area. The product level of the Pleiades images is Sensor, which means the images were not ortho-rectified. The GSD is about 0.5 m on the Pleiades images.

Table 1. The meta-information of the satellite images.

Image Title	Satellite	GSD (m)	Product Level	Acquisition Date	Approximate Attitude (Degree)		
					Roll	Pitch	Yaw
Test area near the city of Guangzhou, Guangdong Province, China							
GD-Z3N-1	ZY-3-NAD	2.1	Level-1A	14 April 2015	0.00	0.22	0.19
GD-Z3F	ZY-3-FWD	3.2	Level-1A	14 April 2015	−0.00	−23.59	0.13
GD-Z3B	ZY-3-BWD	3.2	Level-1A	14 April 2015	−0.18	24.09	0.25
GD-Z3N-2	ZY-3-NAD	2.1	Level-1A	14 April 2015	0.00	0.22	0.19
GD-GF2	GF-2-PMS	0.8	Level-1A	23 January 2015	1.45	−0.36	0.01
GD-IK-1	IKONOS-2	1.0	SGC ¹	18 September 2001	12.17	22.21	0.00
GD-IK-2	IKONOS-2	1.0	SGC	23 November 2001	5.98	4.93	0.00
GD-W2-1	WorldView-2	0.5	Level-1B	28 December 2013	−1.14	0.38	0.01
GD-W2-2	WorldView-2	0.5	Level-1B	28 December 2013	−8.45	29.76	4.87
Test area in Southern Jiangsu Province, China							
JS-Z3N	ZY-3-NAD	2.1	Level-1A	17 October 2012	2.43	0.10	0.18
JS-GF1	GF-1-PMS	2.0	Level-1A	22 January 2014	1.71	−0.47	0.03
JS-GF2-1	GF-2-PMS	0.8	Level-1A	15 October 2015	15.88	−1.18	0.28
JS-GF2-2	GF-2-PMS	0.8	Level-1A	15 October 2015	15.88	−1.18	0.27
Test area near the city of Seoul, South Korea							
SK-W3-1	WorldView-3	0.4	Level-1B	7 August 2015	26.82	0.01	0.01
SK-W3-2	WorldView-3	0.4	Level-1B	7 August 2015	21.18	33.01	12.26
SK-P1A	Pleiades-HR 1A	0.5	Sensor	18 November 2017	−11.58	4.52	0.89
SK-P1B	Pleiades-HR 1B	0.5	Sensor	28 September 2017	1.56	13.78	0.28

¹ SGC, Standard Geometrically Corrected.

The relative positions of the images in the three test areas are shown in Figure 5. Combining the attitude angles in Table 1, we found that the pitch angles of the ZY-3-NAD images and GF-1/2-PMS images are very close to zero. Thus, the stereo image-pairs formed by ZY-3-NAD images and GF-1/2-PMS images are under the cross-track coplanar stereo condition (Figure 1a), and the image-pairs formed by a GF-1/2 PMS image and a ZY-3-FWD or ZY-3-BWD image follow the cross-track non-coplanar stereo condition (Figure 1c). The stereo image-pairs collected by ZY3 three-line-array and WorldView-2 in Guangdong and collected by WorldView-3 in South Korea are under the along-track-stereo condition (Figure 1b). For the images collected by IKONOS-2 in Guangdong, both the roll angle and the pitch angle were changed for side observation and stereo observation, thus the pair of IKONO-2 images is under the cross-track non-coplanar condition. As shown in Figure 5, the directions of the orbits of IKONOS-2, WorldView-2/3, and Pleiades are just south, while those of the ZY-3 and GF-1/2 have about 20-degree angles to the south direction. As a result, the stereo image pairs formed by a ZY-3/GF-1/GF-2 image and an IKONOS-2/WorldView-2/image are under the CMDT stereo condition (Figure 1d).

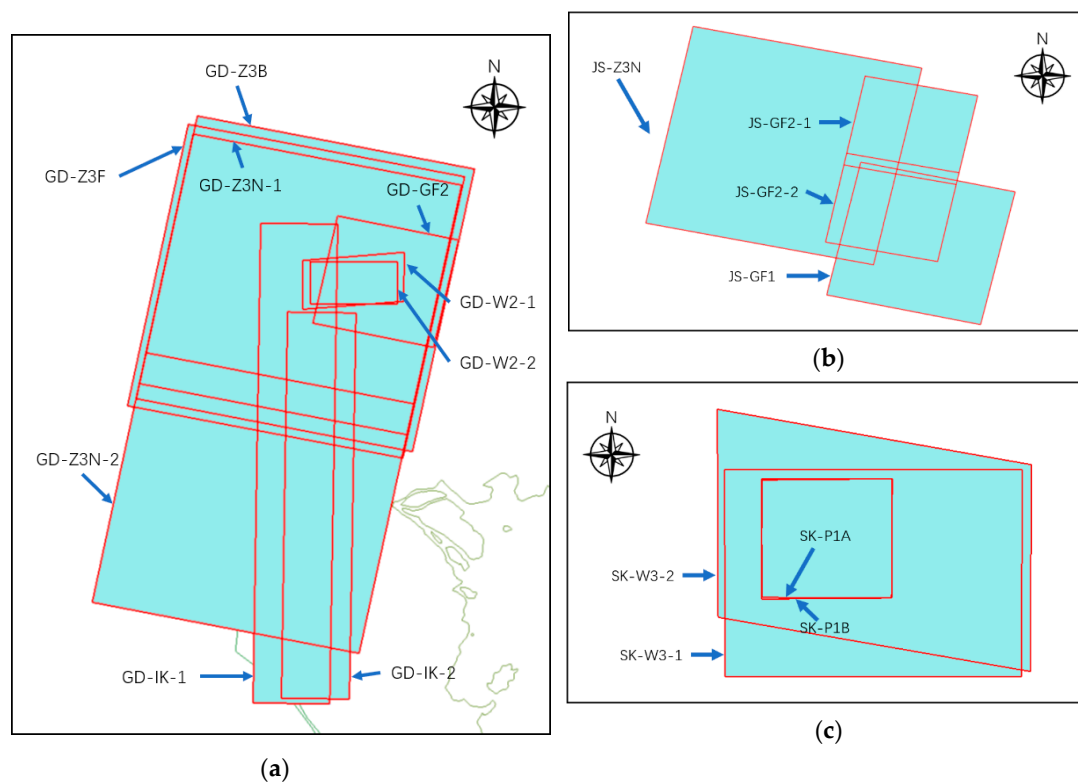


Figure 5. The relative positions of the tested satellite images: (a) the Guangdong test area; (b) the Jiangsu test area; and (c) the South Korea test area.

Tie-point matching [35], outlier filtering [36], and block adjustment [37] were processed before testing the epipolar resampling approaches, because the ZY-3/GF-1/GF-2 images are Level-1A images that were not geometrically corrected by the vendors. Meanwhile, although the images captured by IKONOS-2/WorldView-2/WorldView-3 were geometrically corrected by the vendors, their standard product accuracy (6 m for IKONOS-2 and 3 m for WorldView-2/3) cannot ensure the disparity perpendicular to the ELD is within one pixel of their GSDs.

The settings of the epipolar resampling process are as follows. When processing RBH (the proposed approach), the number of grid-points is at least 20 in the x -direction or y -direction; ∇H , the elevation interval for computing the ELD through the process in Equation (5), was set as 50 m. When processing PRP [21], the elevation of the projection reference plane was set as the average elevation of the overlapped area; the epipolar images were resampled in the same ground resolution

of the original left image. When processing TPP [24], two three-order polynomials were used as transformation models; each epipolar curve was expressed by 10 pieces of line-segments.

After preparing the transformation models, the accuracy was evaluated by the y -disparities of the virtual corresponding points (VCPs) and the tie-points (TPs). The VCPs were random left-image-points with their correspondences collected through the process in Equation (4); that is, ray-tracing with SRTM followed by projection. The TPs used for evaluation were selected with the following steps. First, map the left-image-point \mathbf{p} from to the right image \mathbf{I}' through the process in Equation (4) and denote it as \mathbf{p}'' . Second, obtain the ELD through the process in Equation (5) at the location of \mathbf{p}' , which is the matched point of \mathbf{p} on the right image \mathbf{I}' . Third, compute the point-to-line distance between the point \mathbf{p}'' and the straight line that crosses the point \mathbf{p}' and follows the direction of the ELD [36]. The TPs with less-than-0.5 pixel point-to-line distances were used for evaluation. It should be noted that the images with larger GSDs were set as the left images in the epipolar resampling, to avoid up-sampling the images. The y -disparities were computed with the unit of the left-image-pixel.

5.2. Experimental Results

The experimental results are listed in different groups in Tables 2–5. Each group corresponds to a stereo condition illustrated in Figure 1. The optimal values (closest to zero) are in bold font.

The results of two cross-track coplanar stereo image-pairs are listed in Table 2. All four images in the Jiangsu test area have close-to-zero pitch angles. With both the image pairs, the accuracy of the proposed RBH method and that of the PRP method evaluated by either the TPs or VCPs were very close; the y -disparities of the TPs ranged from about -0.5 pixels to 0.5 pixels, and the y -disparities of the VCPs ranged from about -0.1 pixels to 0.1 pixels. The accuracy of the TPP method was lower; the y -disparities of the TPs ranged from about -0.8 pixels to 0.9 pixels, and the y -disparities of the VCPs ranged from about -0.3 pixels to 0.5 pixels.

Table 2. The results of the image pairs in the cross-track coplanar stereo condition.

Left Image	Right Image	Approach	y -Disparity of TPs			y -Disparity of VCPs		
			RMS	Min	Max	RMS	Min	Max
JS-Z3N	JS-GF2-1	RBH	0.26	-0.53	0.53	0.01	-0.02	0.02
		PRP	0.24	-0.50	0.50	0.03	-0.04	0.04
		TPP	0.31	-0.77	0.67	0.11	-0.31	0.19
JS-GF1	JS-GF2-2	RBH	0.28	-0.50	0.52	0.02	-0.01	0.04
		PRP	0.25	-0.51	0.51	0.04	-0.07	0.06
		TPP	0.33	-0.63	0.90	0.17	-0.22	0.44

The results of four along-track stereo image pairs are listed in Table 3. The proposed RBH method still had the smallest y -disparity range for both TPs and VCPs. The y -disparities ranges of the PRP method and the TPP method, although a little wider than those of the RBH method, remained within the range of ± 1 pixel, except on the WorldView-3 image pairs in South Korea.

Table 3. The results of the image pairs in along-track stereo condition.

Left Image	Right Image	Approach	y -Disparity of TPs			y -Disparity of VCPs		
			RMS	Min	Max	RMS	Min	Max
GD-W2-1	GD-W2-2	RBH	0.23	-0.53	0.50	0.02	-0.05	0.04
		PRP	0.26	-0.72	0.76	0.14	-0.25	0.25
		TPP	0.24	-0.64	0.61	0.08	-0.18	0.15
GD-Z3B	GD-Z3N-1	RBH	0.19	-0.50	0.50	0.00	-0.01	0.01
		PRP	0.23	-0.70	0.71	0.12	-0.23	0.22
		TPP	0.26	-0.85	0.70	0.12	-0.38	0.21
GD-Z3F	GD-Z3N-1	RBH	0.19	-0.50	0.50	0.00	-0.01	0.01
		PRP	0.19	-0.58	0.56	0.04	-0.10	0.10

Table 3. Cont.

Left Image	Right Image	Approach	<i>y</i> -Disparity of TPs			<i>y</i> -Disparity of VCPs		
			RMS	Min	Max	RMS	Min	Max
SK-W3-1	SK-W3-2	TPP	0.26	−0.73	0.72	0.12	−0.27	0.24
		RBH	0.29	−0.59	0.51	0.05	−0.08	0.03
		PRP	0.53	−1.42	1.34	0.42	−0.82	0.76
		TPP	0.41	−0.66	1.46	0.32	−0.29	1.00

The results of four cross-track non-coplanar image pairs are listed in Table 4. In this group, the ranges of the *y*-disparities of VCPs with the RBH methods were obviously narrower than those with the PRP method or with the TPP method. For the pairs of GD-1K-1/GD-W2-1 and SK-P1A/SK-W3-2 with the TPP method, the ranges of the *y*-disparities of both TPs and VCPs exceeded the range of ± 1 pixel, which may have had a bad influence on one-dimensional matched point searching.

Table 4. The results of the image pairs in the cross-track non-coplanar stereo condition.

Left Image	Right Image	Approach	<i>y</i> -Disparity of TPs			<i>y</i> -Disparity of VCPs		
			RMS	Min	Max	RMS	Min	Max
GD-Z3B	GD-GF2	RBH	0.25	−0.50	0.50	0.01	−0.05	0.07
		PRP	0.26	−0.81	0.80	0.26	−0.46	0.41
		TPP	0.36	−1.06	1.04	0.37	−0.74	0.70
GD-Z3F	GF-GF2	RBH	0.24	−0.50	0.50	0.01	−0.03	0.07
		PRP	0.22	−0.71	0.78	0.20	−0.35	0.37
		TPP	0.31	−0.91	1.05	0.32	−0.58	0.74
GD-1K-1	GD-W2-1	RBH	0.26	−0.50	0.48	0.00	−0.00	0.01
		PRP	0.39	−0.79	0.86	0.22	−0.36	0.46
		TPP	0.63	−1.20	1.59	0.50	−0.94	1.18
SK-P1A	SK-W3-2	RBH	0.27	−0.54	0.53	0.01	−0.02	0.01
		PRP	0.35	−0.90	1.01	0.25	−0.49	0.68
		TPP	1.02	−2.58	2.81	1.00	−2.06	2.45

The results of three CMDT stereo image pairs are listed in Table 5. In this group, the accuracy of the proposed RBH method is much better than that of either the PRP method or the TPP method, whether evaluated with the *y*-disparities of the TPs or evaluated with the *y*-disparities the VCPs. In all three image pairs, only the RBH method kept the *y*-disparities of TPs within the range of -1 to 1 pixels. The *y*-disparity ranges with both the PRP method and the TPP method exceeded the range of ± 1 pixel. In the pair of GD-Z3N-2/GD-1K-2, the *y*-disparity of the TPs exceeded 3 pixels with the PRP method and exceeded 5 pixels with the TPP method, which would result in failures in dense image matching.

Table 5. The results of the image pairs in the cross-multi-directional-track stereo condition.

Left Image	Right Image	Approach	<i>y</i> -Disparity of TPs			<i>y</i> -Disparity of VCPs		
			RMS	Min	Max	RMS	Min	Max
GD-GF2	GD-W2-2	RBH	0.28	−0.67	0.51	0.06	−0.11	0.04
		PRP	0.40	−1.21	1.14	0.35	−0.69	0.68
		TPP	0.73	−1.90	1.99	0.73	−1.40	1.63
GD-Z3N-2	GD-1K-2	RBH	0.27	−0.51	0.51	0.01	−0.04	0.04
		PRP	1.60	−3.15	3.24	1.57	−2.85	2.84
		TPP	3.18	−5.74	5.99	3.16	−5.58	5.59
GD-Z3N-1	GD-W2-2	RBH	0.26	−0.65	0.50	0.11	−0.32	0.11
		PRP	0.49	−1.14	1.62	0.38	−0.74	0.71
		TPP	0.92	−2.02	2.00	0.84	−1.66	1.56

5.3. Analysis and Discussion

In Section 3.1, we point out that different elevations will not influence the computation of the ELD in the cross-track coplanar stereo condition and along-track stereo condition, but will influence it in the cross-track non-coplanar stereo condition and CMDT stereo condition. The experimental results presented in Section 5.2 demonstrate the above conclusion. With the cross-track coplanar stereo image pairs and the along-track stereo image pairs, the proposed RBH method had almost the same accuracy as the PRP method and the TPP method, although they computed the ELDs with different strategies. However, with the cross-track non-coplanar stereo image pairs and the CMDT stereo image pairs, the proposed RBH method had obviously better accuracy than the PRP method and the TPP method, which demonstrates that medium-accuracy elevations can greatly improve the computing accuracy of the ELDs. Moreover, in some image pairs (e.g., GD-Z3N-2/GD-IK-2), only the proposed RBH method could obtain accurate enough results, since the PRP method and TPP method exceeded the y -disparity-range of ± 1 -pixel with some TPs.

For a better understanding, Figure 6 illustrates the $\theta_c^{(org)} \sim H$ curves of four image-pairs belonging to four different stereo conditions, where $\theta_c^{(org)}$ is the angle between the ELD at the center of the overlapping area on the left image and the positive direction of the x -axis. The angle $\theta_c^{(org)}$ was computed with the elevation H ranging from 0 m to 1000 m. As shown in Figure 6a, the $\theta_c^{(org)}$ s computed with 0 m elevation and with 1000 m elevation had less than 0.04-degree differences with the cross-track coplanar stereo image pair. As shown in Figure 6b, the $\theta_c^{(org)}$ s were almost the same, with different elevations with the along-track stereo image pair. As shown in Figure 6c, the $\theta_c^{(org)}$ s had about a 0.12-degree difference with 0 m elevation and with 1000 m elevation with the cross-track non-coplanar stereo image pair. As shown in Figure 6d, the difference of the $\theta_c^{(org)}$ s could be as large as 0.22 degrees with 0 m elevation and with 1000 m elevation, which could cause about 2.6 pixels difference in a distance of 1000 pixels. The variation of the $\theta_c^{(org)}$ with different elevations demonstrates the importance of computing accurate ELDs with the aid of SRTM when using epipolar resampling with cross-track non-coplanar stereo image pairs or CMDT stereo image pairs.

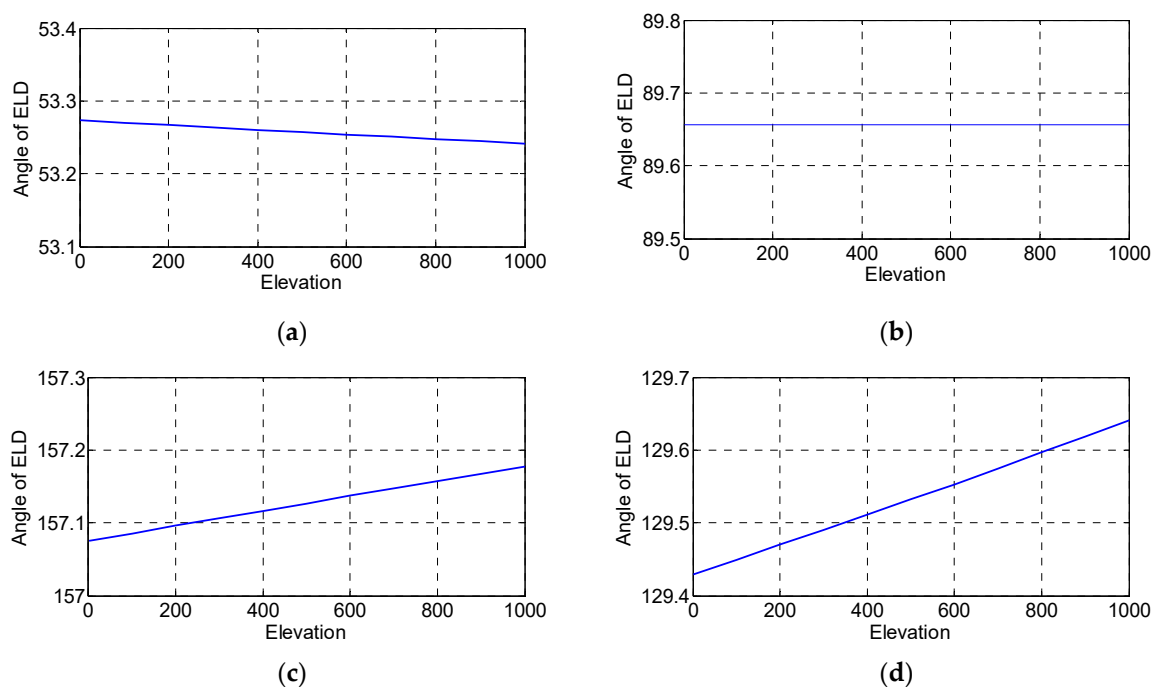


Figure 6. The $\theta_c^{(org)} \sim H$ curves of: (a) a cross-track coplanar stereo image pair (JS-Z3N/JS-GF2-1); (b) an along-track stereo image pair (GD-Z3B/GD-Z3N-1); (c) a cross-track non-coplanar stereo image pair (SK-P1A/SK-W3-2); and (d) a CMDT stereo image pair (GD-Z3N-2/GD-IK-2).

In almost all tested image pairs, the PRP method achieved better accuracies than the TPP method even when evaluated by the TPs, demonstrating that a corresponding-point-free method is a better choice since both the ELD and the corresponding epipolar curve can be accurately determined with pre-corrected RFMs.

6. Conclusions

This paper classifies the geometric conditions of satellite-image-based binocular stereo observation into four classes: cross-track coplanar stereo, along-track stereo, cross-track non-coplanar stereo, and CMDT stereo. Under the cross-track coplanar stereo condition and along-track stereo condition, the ELDs are not influenced by the choice of elevation in computing. However, under the cross-track non-coplanar stereo condition and the CMDT stereo condition, the ELDs are influenced by elevation. To deal with this problem, this paper proposes a novel SRTM (or other public medium-accuracy elevation data source)-aided method of epipolar resampling for satellite stereo images. The proposed method transforms a point from the original image-space to the epipolar image-space through two transformations: a global rotation, followed by a block-wise homography transformation. The SRTM supplies medium-accuracy elevations to compute an accurate ELD when preparing the block-wise homography transformation. The experiments demonstrated that, compared with two other methods that compute the ELDs with a unified elevation plane, the proposed method is much more accurate. On some image pairs, only the proposed method could ensure the y -disparities of the TPs remained within the range of ± 1 pixels. Since most multi-source stereo image pairs are cross-track non-coplanar stereo and CMDT stereo, the contributions of this paper are very important for research regarding the problem of multi-source satellite stereo observation.

There is still work to do to improve the proposed method in the future. First, the block-wise transformation from the epipolar image-space to the rotated image-space on both the left image and the right image uses irregular grids, which is time-consuming since the transformation is very frequently required in the resampling process. Thus, optimizations in efficiency are required. Second, the proportion between the x -disparity and the height difference is uneven since the proposed method works on the image-space, which makes the post-processing of the dense image matching inefficient because we need to transform the matching points from the epipolar image space to the original image-space, and then obtain accurate elevations through forward-intersecting. Similar approaches may be developed in the object-space to deal with this problem.

In this study, SRTM was used to constrain the ELD of an image location because the corresponding object-space point was close to the elevation plane of SRTM. In the research on multi-source satellite stereo, SRTM constrained dense image-matching is also a potential technical route that may lead to important breakthroughs.

Author Contributions: J.H. had the original idea for the study. G.-S.X. supervised the research and contributed to the article's organization. J.H. drafted the manuscript, which was revised by all authors. All authors read and approved the submitted manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant numbers 61771014, 61771350, and 41820104006.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Scharstein, D.; Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **2002**, *47*, 7–42. [[CrossRef](#)]
2. Zhang, Y.; Zhang, Y.; Mo, D.; Zhang, Y.; Li, X. Direct digital surface model generation by semi-global vertical line locus matching. *Remote Sens.* **2017**, *9*, 214. [[CrossRef](#)]
3. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multiview stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1362–1376. [[CrossRef](#)] [[PubMed](#)]

4. Garg, R.; Vijay Kumar, B.G.; Carneiro, G.; Reid, I. Unsupervised cnn for single view depth estimation: Geometry to the rescue. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin, Germany, 2016; pp. 740–756.
5. Godard, C.; Mac Aodha, O.; Brostow, G.J. Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 270–279.
6. Cao, Y.; Wu, Z.; Shen, C. Estimating depth from monocular images as classification using deep fully convolutional residual networks. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *28*, 3174–3182. [[CrossRef](#)]
7. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [[CrossRef](#)]
8. Toutin, T.; Schmitt, C.; Wang, H. Impact of no GCP on elevation extraction from WorldView stereo data. *ISPRS J. Photogramm. Remote Sens.* **2012**, *72*, 73–79. [[CrossRef](#)]
9. Wang, T.; Zhang, G.; Li, D.; Tang, X.; Jiang, Y.; Pan, H.; Zhu, X.; Fang, C. Geometric accuracy validation for ZY-3 satellite imagery. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1168–1171. [[CrossRef](#)]
10. Zbontar, J.; LeCun, Y. Stereo matching by training a convolutional neural network to compare image patches. *J. Mach. Learn. Res.* **2016**, *17*, 2.
11. Zbontar, J.; Le Cun, Y. Computing the stereo matching cost with a convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1592–1599.
12. Zhang, Z. Determining the epipolar geometry and its uncertainty: A review. *Int. J. Comput. Vis.* **1998**, *27*, 161–195. [[CrossRef](#)]
13. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
14. Gupta, R.; Hartley, R.I. Linear pushbroom cameras. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 963–975. [[CrossRef](#)]
15. Kim, T. A study on the epipolarity of linear pushbroom images. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 961–966.
16. Al-Rousan, N.; Cheng, P.; Petrie, G.; Toutin, T.; Zoj, M.V. Automated DEM extraction and orthoimage generation from SPOT level 1B imagery. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 965–974.
17. Bang, K.I.; Jeong, S.; Kim, K.-O.; Cho, W. Automatic DEM generation using IKONOS stereo imagery. In Proceedings of the 2003 IEEE International Geoscience and Remote Sensing Symposium (IGARSS'03), Toulouse, France, 21–25 July 2003; pp. 4289–4291.
18. Morgan, M.; Kim, K.-O.; Jeong, S.; Habib, A. Epipolar resampling of space-borne linear array scanner scenes using parallel projection. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 1255–1263. [[CrossRef](#)]
19. Morgan, M.F. *Epipolar Resampling of Linear Array Scanner Scenes*; University of Calgary: Calgary, AB, USA, 2004.
20. Hu, F.; Wang, M.; Li, D. A novel epipolarity model of satellite stereo-imagery based on virtual horizontal plane of object-space. In Proceedings of the International Conference on Earth Observation Data Processing and Analysis (ICEODPA), Wuhan, China, 28–30 December 2008; International Society for Optics and Photonics: Bellingham, WA, USA, 2008; p. 72851D.
21. Wang, M.; Hu, F.; Li, J. Epipolar resampling of linear pushbroom satellite imagery by a new epipolarity model. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 347–355. [[CrossRef](#)]
22. Dowman, I.J.; Michalis, P. In Generic rigorous model for along track stereo satellite sensors. In Proceedings of the ISPRS Workshop on High Resolution Mapping from Space, Hanover, Germany, 9–13 September 2003.
23. Fraser, C.S.; Dial, G.; Grodecki, J. Sensor orientation via RPCs. *ISPRS J. Photogramm. Remote Sens.* **2006**, *60*, 182–194. [[CrossRef](#)]
24. Idrissa, M.; Beumier, C. Generic epipolar resampling method for perspective frame camera and linear push-broom sensor. *Int. J. Remote Sens.* **2016**, *37*, 3494–3504. [[CrossRef](#)]
25. Jannati, M.; Valadan Zoj, M.J.; Mokhtarzade, M. Epipolar resampling of cross-track pushbroom satellite imagery using the rigorous sensor model. *Sensors* **2017**, *17*, 129. [[CrossRef](#)]
26. Koh, J.-W.; Yang, H.-S. Unified piecewise epipolar resampling method for pushbroom satellite images. *EURASIP J. Image Video Process.* **2016**, *2016*, 11. [[CrossRef](#)]

27. Toutin, T. Spatiotriangulation with multisensor VIR/SAR images. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 2096–2103. [[CrossRef](#)]
28. Toutin, T. Spatiotriangulation with multisensor HR stereo-images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 456–462. [[CrossRef](#)]
29. Jeong, J.; Kim, T. Analysis of dual-sensor stereo geometry and its positioning accuracy. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 653–661. [[CrossRef](#)]
30. Jeong, J.; Kim, T. Comparison of positioning accuracy of a rigorous sensor model and two rational function models for weak stereo geometry. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 172–182. [[CrossRef](#)]
31. Jeong, J.; Yang, C.; Kim, T. Geo-positioning accuracy using multiple-satellite images: IKONOS, QuickBird, and KOMPSAT-2 stereo images. *Remote Sens.* **2015**, *7*, 4549–4564. [[CrossRef](#)]
32. Jeong, J.; Kim, T. Quantitative estimation and validation of the effects of the convergence, bisector elevation, and asymmetry angles on the positioning accuracies of satellite stereo pairs. *Photogramm. Eng. Remote Sens.* **2016**, *82*, 625–633. [[CrossRef](#)]
33. Sun, G.; Ranson, K.J.; Kharuk, V.I.; Kovacs, K. Validation of surface height from shuttle radar topography mission using shuttle laser altimeter. *Remote Sens. Environ.* **2003**, *88*, 401–411. [[CrossRef](#)]
34. Sheng, Y. Theoretical analysis of the iterative photogrammetric method to determining ground coordinates from photo coordinates and a DEM. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 863–871. [[CrossRef](#)]
35. Ling, X.; Zhang, Y.; Xiong, J.; Huang, X.; Chen, Z. An image matching algorithm integrating global STRM and image segmentation for multi-source satellite imagery. *Remote Sens.* **2016**, *8*, 672. [[CrossRef](#)]
36. Wan, Y.; Zhang, Y. The P2L method of mismatch detection for push broom high-resolution satellite images. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 317–328. [[CrossRef](#)]
37. Zhang, Y.; Wan, Y.; Huang, X.; Ling, X. Dem-assisted RFM block adjustment of pushbroom nadir viewing HRS imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1025–1034. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).