Optimization method for DSM reconstruction of equivalent pinhole model based on iterative resampling and chunking process

Yongjian Li, Song Ji, Danchao Gong, Dazhao Fan, Jiaxuan Song, Ding Gao

Abstract—The development of satellite remote sensing technology has made it easier to obtain satellite imagery. Compared to imagery obtained from aerial photography, satellite imagery has the advantages of wide coverage, high acquisition efficiency, and periodic revisits. In photogrammetry, the 3D reconstruction technology of satellite imagery often requires optimization and adjustment of numerous RPC parameters, which to some extent limits the speed and accuracy of 3D reconstruction. At the same time, the progress in 3D reconstruction technology in the field of computer vision has shown certain advantages in terms of accuracy and speed. However, these methods are specifically designed for pinhole imaging models and cannot be directly applied to the 3D reconstruction of satellite imagery with row-sampled central projection. The introduction of the equivalent pinhole imaging model enables computer vision methods to perform 3D reconstruction on satellite imagery. This local approximation introduces re-projection errors when the RPC model is equivalent to the pinhole imaging model, thereby affecting the accuracy of 3D reconstruction.This paper investigates the causes and patterns of re-projection errors in the equivalent pinhole imaging model and proposes a method for generating pseudo-images through iterative resampling, as well as a method for partitioning satellite images to equivalently approximate the pinhole imaging model. Test results on the MVS3D dataset show that both methods can reduce re-projection errors, thereby improving the accuracy of 3D reconstruction of satellite images using the equivalent pinhole imaging model.

Index Terms—Equivalent Pinhole Camera Models, RPC Model, 3D Reconstruction.

I. INTRODUCTION

REMOTE sensing satellite imagery plays a crucial role in target detection and tracking, terrain mapping and analysis, map production and updating, 3D urban modeling, and navigation. With the increase in the number of remote sensing satellites and the advancement of sensors, satellites can image the same geographic area within seconds, days, months, or even years, generating a sufficient number of images for 3D reconstruction. In this context, there is a growing interest among researchers in using satellite imagery for large-scale 3D reconstruction of the Earth's surface [1].

Photogrammetry and Remote Sensing, as a discipline, are intersecting and integrating with Computer Vision, with their boundaries becoming increasingly blurred [2]. There is a convergence and mutual borrowing of techniques and methods between them [3]. However, there are still distinct differences in the aspect of 3D reconstruction:

- In different fields, there are significant differences in imaging models. Satellite cameras typically use rational function models to characterize complex imaging relationships, while computer vision use simple pinhole camera models.
- In the field of remote sensing, integrating multiple stereo reconstructions can produce better results, as seen in the representative S2P satellite reconstruction process [4]. Due to the inherent revisit cycles of satellite imaging, combined with factors such as cloudy weather, multi-view satellite images of specific locations may span several years. Changes in image features caused by variations in lighting, seasons, and weather conditions make remote sensing techniques more suitable for processing satellite images with minor temporal changes and insignificant terrain alterations. In the domain of computer vision, multi-view stereo methods have long been explored by researchers because of their advantages in dealing with occlusions, repetitions, and temporal changes.

3D reconstruction methods in the field of computer vision have developed rapidly. Techniques such as Structure from Motion (SfM) and Multi-view Stereo (MVS) have been widely applied to the 3D reconstruction of drone imagery and ground photographs, achieving good results. However, due to differences in imaging modalities, SfM and MVS cannot be directly applied to satellite imagery. Although researchers have proposed equivalent pinhole camera models to establish a connection between remote sensing and computer vision, these models do not accurately approximate the Rational Polynomial Coefficients (RPC) model of satellite images, especially for large-scale images. This is because the equivalent pinhole imaging model for satellite images introduces significant reprojection errors, which affect the accuracy of 3D reconstruction from the images.

The challenges faced in using computer vision methods for 3D reconstruction of satellite imagery include: (1) For Structure from Motion (SfM): The main difficulty stems from the RPC parameters widely used in satellite imagery, which cannot be directly employed as camera pose parameters in vision-based reconstruction [5]. Methods that locally approximate the RPC model to a pinhole camera model enable us to utilize computer vision SfM methods for bundle adjustment of camera parameters [6]. (2) For Multi-View Stereo (MVS): Due to the vast distance between the satellite and the Earth, depth
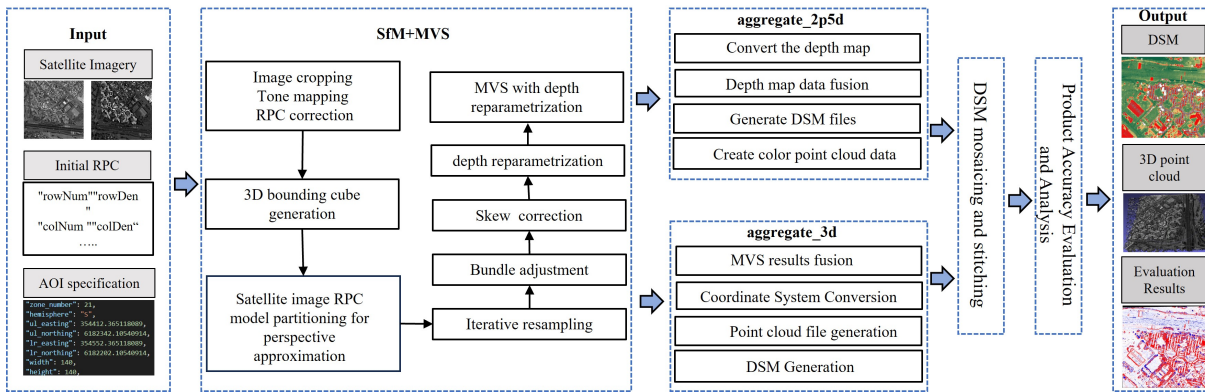
Fig. 1: Technical Flowchart

will be concentrated far from the sensor plane, leading to a large mean depth distribution with low variance. Consequently, the computation of depth values suffers from a lack of numerical precision. Depth can be reparameterized using a reference plane, ensuring the numerical precision of operations.

In this paper, we analyze the equivalent transformation methods between different imaging models. To address the projection errors introduced by approximating the RPC model locally to a pinhole imaging model, we propose an iterative resampling method to generate pseudo-images and a block-based reconstruction scheme. We have demonstrated that these two methods can reduce the reprojection errors caused by the equivalent model, thereby improving the accuracy of 3D reconstruction. We apply the 3D reconstruction module of COLMAP, a computer vision 3D reconstruction pipeline. COLMAP performs Structure from Motion (SfM) and Multi-view Stereo (MVS) for 3D reconstruction.

To enhance the precision of 3D reconstruction of satellite images using equivalent pinhole imaging models, we have designed the technical workflow as depicted in Fig. 1. The input data consists of three parts: satellite image files, Rational Polynomial Coefficients (RPC) parameters, and the area of interest extent. During the construction of the equivalent pinhole model, we reduce the reprojection error of the equivalent model by locally approximating the RPC model and generating pseudo-images through iterative resampling. Subsequently, skew correction is applied to eliminate the skew parameters in the projection matrix P of the equivalent pinhole model. This allows us to handle the pseudo-images with standard procedures for pinhole camera models, including bundle adjustment and depth estimation, among other steps. We used two methods to generate point clouds and Digital Surface Model(hereinafter referred to as DSM). The first method involves creating DSM based on the depth maps of each camera, then fusing these DSMs to produce the final DSM product, and subsequently extracting pseudo-3D point clouds from the merged DSM. The second method utilizes the COLMAP Stereo_fusion tool to convert depth maps into 3D point clouds, which are then projected onto a geographic grid to derive the DSM. Finally, the DSMs from each segment are stitched together according to the partitioning scheme, and the accuracy is assessed based on the reference true value DSM obtained from LiDAR data.

In the MVS3D dataset, an experimental area of 500 meters by 500 meters was selected for the experiment. The median error of the DSM reconstructed using the equivalent pinhole camera model proposed by Zhang et al [7]. is 0.51 meters. After two rounds of image resampling and reconstruction, the median error of the DSM is 0.38 meters. The median error of the DSM obtained by reconstructing the entire experimental area is 0.23 meters. The median error of the DSM obtained by dividing the experimental area into a 4 by 4 grid and reconstructing in blocks is 0.09 meters.These experimental results indicate that the iterative resampling strategy and block-based reconstruction strategy proposed in this study can reduce reprojection errors, thereby improving the accuracy of 3D reconstruction from satellite images using the equivalent model.

II. RELATED WORKS

A. Research on RPC Model

In the realms of photogrammetry and remote sensing, the Rational Polynomial Coefficients (RPC) model, introduced by B. J. Nelson in the 1980s, assumes a pivotal role. The standard RPC model consists of 78 coefficients and 10 normalization constants [8]. This model precisely describes the spatial relationship between ground points and their corresponding points in the image, and can be represented as follows:

$$u = \mu_u + \sigma_u \cdot g\left(\frac{x - \mu_x}{\sigma_x}, \frac{y - \mu_y}{\sigma_y}, \frac{z - \mu_z}{\sigma_z}\right) \quad (1)$$

$$v = \mu_v + \sigma_v \cdot h\left(\frac{x - \mu_x}{\sigma_x}, \frac{y - \mu_y}{\sigma_y}, \frac{z - \mu_z}{\sigma_z}\right), \quad (2)$$

where x, y, z represent latitude, longitude, and elevation, respectively, and g, h denote the ratio of two cubic polynomials represented by 39 coefficient parameters, $\mu_{\{x,y,z,u,v\}}$ and $\sigma_{\{x,y,z,u,v\}}$ are normalization constants. Although RPC parameters do not have a direct geometric interpretation corresponding to camera attitudes, the RPC model has been successfully applied to downstream photogrammetric tasks such as orthoimage generation and DEM production.RPC model has become the primary method for reflecting the relationship between satellite image point coordinates and ground point coordinates [8].

Research on the Rational Polynomial Coefficients (RPC) model is primarily divided into two domains: on one hand, there is the study of RPC parameter estimation, bundle adjustment, and residual minimization algorithms [9]. However, parameter estimation relies on specific information about the physical sensor, which is typically conducted by satellite image providers and is often not readily accessible to users. On the other hand, numerous studies have been dedicated to analyzing the errors inherent in the RPC model and have introduced the concept of calibrating RPC model parameters [10]. Some research has found that the bias in the RPC model can be equated to translations of pixels in 2D image space, and this bias can be corrected using artificially measured ground control points [11]. Compensating for and correcting pixel coordinates in 2D image space to enhance the accuracy of the RPC model is essentially a process of re-sampling images and re-arranging pixel positions. This concept of correction in image space can also be applied to correct the bias between the equivalent pinhole model and the RPC model, which is the method proposed in this paper of generating pseudo-images through iterative resampling.

B. Research on 3D Reconstruction of Satellite Imagery in Remote Sensing

In the field of remote sensing, research and application of 3D reconstruction algorithms based on the RPC model are continuously making new progress. The RPC model, introduced by B. J. Nelson in the 1980s, has become a key tool in this field because it can accurately describe the spatial relationship between ground points and corresponding points in the image [12] [13].

Currently, reconstruction algorithms based on the RPC model, represented by S2P, follow this approach: first, complete stereo reconstruction, and then achieve multi-view reconstruction through DSM fusion algorithms. The reconstruction process mainly includes three key steps: obtaining dense correspondence between images using parallax(D'Angelo and Reinartz, 2012 [14]; Kuschk, 2013 [15]; Rupnik et al. 2017 [16]) or optical flow method; optimizing the RPC model estimation residuals of optical satellite images through bundle adjustment; and performing triangulation based on the RPC model to recover the 3D structure of the scene corresponding to the images.For example, Bosch et al. (2017) optimized the RPC model through bundle adjustment to achieve maximum consistency between models, and used the RPC model for triangulation to recover the 3D structure of the ground scene. De Franchis et al. (2014) [17] used an RPC model elevation iteration method, correcting the pointing error between two views, avoiding complex nonlinear bundle adjustment, and achieving the 3D reconstruction process of optical remote sensing images.

In addition, Facciolo et al. (2017) [18] analyzed the main factors affecting the accuracy of stereo reconstruction, sorted the image pairs in the dataset based on these factors, and ultimately selectively fused the independent stereo reconstruction results. These studies demonstrate the potential of RPC-based satellite image 3D reconstruction methods in improving reconstruction accuracy. It is worth noting that although the accuracy and speed of RPC-based satellite image 3D reconstruction methods are continuously improving, these methods cannot avoid solving the complex cubic polynomials of the RPC model. The parameter estimation of the RPC model requires at least 39 sets of ground control points and corresponding image point relationships, which to some extent limits the speed and accuracy of 3D reconstruction.

C. Research on 3D Reconstruction of Satellite Imagery in Computer Vision.

In the field of computer vision, significant progress has been made in 3D scene reconstruction methods based on optical images(Govindu, 2004 [19]; Snavely et al. 2008 [20]; Wilson and Snavely, 2014 [21]; Schonberger and Frahm, 2016 [22]; Yu, Y. et al. 2019 [23]; Xue, J. S. et al. 2020 [24]). These methods are typically suitable for processing images captured by area-array cameras, but they are not applicable to images obtained from push-broom linear array cameras mounted on satellites. Consequently, researchers have been exploring how to apply these reconstruction algorithms to remote sensing images. Fraser et al. (2014) [25] directly fitted an affine camera model using the true correspondences between ground control points and image points, achieving affine reconstruction of optical remote sensing images through the fitted affine camera model. However, this method operates on the entire optical remote sensing image, resulting in lower reconstruction accuracy.

Wang et al. proposed an incremental optical satellite image reconstruction method based on the affine imaging model, which no longer relies on the RPC model and only requires at least four ground control points for scene reconstruction [26]. Additionally, Chen et al. introduced a method for fast 3D reconstruction of satellite images using a global affine model [27]. This method first crops multi-view satellite images into a set of local image patches and calculates the corresponding 3D affine point clouds for each local scene. It then employs a global affine motion matrix estimation method based on local point clouds to compute the affine motion matrix for each viewpoint, utilizing a few ground control points to recover the Geometric structure of the scene.

Zhang et al.(2019) [7] fitted a weak perspective projection model using the known RPC model and introduced a depth reparameterization method based on COLMAP to achieve 3D reconstruction of optical remote sensing images within a computer vision reconstruction pipeline. They also designed a conversion method between the RPC model and the pinhole imaging model, constructing an equivalent pinhole imaging model for satellite images, and successfully performed 3D reconstruction using Structure from Motion (SfM) and Multi-view Stereo (MVS), achieving satisfactory results in their experiments.

The processes developed in the field of computer vision utilize the easily interpretable pinhole camera model instead of the RPC model. The parameters of the pinhole camera model correspond to the camera's position and orientation, having

clear physical significance. The pinhole camera model can be expressed in the following form:

$$[u; v; 1] = \mathbf{K}[\mathbf{R}|\mathbf{t}][x; y; z; 1] \tag{3}$$

The implication matrix $\mathbf{K} \in \mathcal{R}^{3\times3}$ takes the form $[f_x, s, c_x; 0, f_y, c_y; 0, 0, 1]$, while the outer matrix consists of camera rotations $\mathbf{R} \in \mathrm{SO}(3)$ and translations $\mathbf{t} \in \mathcal{R}^3$. In many cases, zero skew ($s = 0$), unit aspect ratio ($f_x = f_y$) and centered principal points ($c_x, c_y = 0.5$ in normalized image coordinates) can be assumed.Most computer vision reconstruction processes include Structure from Motion (SfM) and Multi-View Stereo (MVS) stages. The goal of SfM is to obtain accurate camera parameters by iteratively optimizing the intrinsic and extrinsic parameters of the projection matrices, as well as triangulating sparse point clouds. Given a set of images and their corresponding camera parameters, MVS reconstructs a dense point cloud or mesh. Key issues of MVS include view selection based on camera baselines, determination of overlapping areas, and robustly recovering dense depth maps [28].

D. Research on Equivalent Pinhole Imaging Model

The main difference between photogrammetric satellite image 3D reconstruction and computer vision reconstruction methods lies in the camera model (RPC model and pinhole imaging model). Zhang et al. tried to bridge this gap by locally approximating the RPC model to a pinhole model [7]. After this approximation, SfM processes such as VisualSFM [29] and COLMAP SfM [6] can be used to adjust camera parameters.

To tailor satellite imagery for Structure from Motion (SfM), Zhang et al. have conducted systematic work, which can be encapsulated in six main aspects [7]:

- They defined SfM coordinate systems adapted for satellite imagery, encompassing the geodetic coordinate system, the tangent plane coordinate system, and the local Universal Transverse Mercator (UTM) projection coordinate system.
- They devised a scheme for cropping satellite images within the UTM projection framework.
- They employed Gamma correction to rectify the long-tail distribution of the original satellite images.
- They mathematically demonstrated the local approximation of the RPC camera to a perspective camera and provided numerical methods for estimating camera parameters.
- While keeping other intrinsic and extrinsic parameters constant, they performed bundle adjustment by adjusting the camera's principal point and added regularization terms to the coordinates of sparse 3D points to ensure they do not deviate significantly from their original coordinates during the bundle adjustment process.
- They designed a simple and effective skew correction step, which involves decomposing the intrinsic parameter matrix into the product of a transformation matrix and an intrinsic parameter matrix with zero skew, using the transformation matrix to resample the original image to

obtain a new image. After the aforementioned transformation, the intrinsic parameter matrix with zero skew can be considered as the intrinsic parameter matrix of the new image.

To adapt Multi-View Stereo (MVS) for satellite imagery, Zhang et al. have implemented the following strategies [7]:

- They introduced a depth reinitialization step during the MVS phase, effectively addressing the issues of numerical precision and GPU consumption caused by the significant difference in height between the satellite camera and ground elevatio.
- They Derived a method for directly computing the homography matrix from the projection matrix, resolving the issue of numerical instability in the homography matrix calculated in single precision.

Hong et al. have analyzed the conditions for establishing an equivalent pinhole model, theoretically deducing that the reprojection error of the equivalent pinhole model may be related to the undulations of the ground elevation [30]. They have also derived and validated that the reprojection error at the edges of satellite images is greater than that at the image center. Additionally, they conducted 3D reconstruction experiments using ZY-3 and GF-7 satellite data, comparing the accuracy and efficiency of the equivalent pinhole model reconstruction method with existing satellite image 3D reconstruction methods across various satellite datasets. The quantitative evaluation results on the WHU-TLC dataset are shown in TABLE I (the best results are indicated in bold. ↓ indicates that the lower the value, the better; ↑ indicates that the higher the value, the better). The DSM generated from the improved equivalent pinhole camera model (PREM) has the highest RMSE accuracy, validating the feasibility of this idea.

III. MATERIALS AND METHODS

Satellite image 3D reconstruction based on the equivalent pinhole model has been proven feasible in both theory and practice, offering certain advantages in terms of accuracy and speed [7] [30]. However, the equivalent pinhole imaging model inevitably introduces reprojection errors for the following reasons: Satellite images are captured using linear array push-broom imaging, with a projection method based on line-center projection. While individual lines in the image conform to rigorous imaging models, the entire satellite image involves multiple projection centers and complex imaging relationships, which cannot be accurately modeled using strict models. As a result, researchers have proposed using complex RPC (Rational Polynomial Coefficient) models to describe the intricate imaging relationships of satellite imagery. The rigorous imaging model typically uses 12 parameters (such as those in a projection matrix), while the RPC model consists of 78 parameters. Therefore, a 12-parameter model cannot adequately capture the complexity of a 78-parameter model.

On this basis, this paper studies the factors affecting the reprojection error of the equivalent pinhole model and designs two methods to reduce the reprojection error to improve the accuracy of 3D reconstruction: First, during the construction of the equivalent pinhole camera model, this paper generates

TABLE I: Quantitative evaluation of the DSM calculation on the WHU-TLC test set.

Methods	MAE(m)↓	RMSE(m)↓	Comp2.5(%)↑	Comp5.0(%)↑	Time(min)↓
CATALYSTa(Catalyst,2023)	3.454	7.939	52.310	82.520	**3.800**
Metashapea(Agisoft,2023)	2.693	13.047	56.590	75.460	24.510
SDRDISa(Sdrdis,2016)	4.496	15.012	47.580	73.570	9.410
Sat-MVSFa(Gao et al.,2023)	1.895	3.654	64.820	80.050	5.870
S2P(Gao et al.,2023)	1.692	8.710	66.738	94.652	7.487
LPS(Leica, 2023)	3.581	13.934	62.242	91.729	8.062
PCAM(Zhang et al., 2019)	1.766	3.334	73.926	**97.428**	7.171
RPEM(Hong et al.,2024)	**1.666**	**3.150**	**75.826**	97.308	7.317

pseudo-images that better conform to the imaging characteristics of the pinhole camera through iterative resampling; Second, this paper segments the original image into blocks for separate reconstruction, and then mosaics and stitches the reconstructed DSM.

A. Equivalent pinhole camera model

In order to further enhance the accuracy of the equivalent pinhole model, this paper proposes an iterative resampling scheme. By performing multiple resamplings of the satellite imagery, the images are made to more closely align with the imaging relationship of a pinhole camera, thereby enabling us to obtain pseudo-satellite images that are nearly identical in imaging characteristics to those of a pinhole camera.

For satellite images, the depth value of each ground point is much larger than the differences in depth values between different ground points. Under such circumstances, we have demonstrated that both perspective cameras and linear push-broom cameras can be accurately approximated by what is known as a weak perspective camera, thus allowing us to approximately convert a linear push-broom camera into a perspective camera. The first step is the model approximation of the perspective camera. Due to the minimal changes in scene depth and the very large average scene depth, the perspective camera is simplified to a weak perspective camera using the following form: $u \approx \frac{f_x}{Z}X + \frac{s}{Z}Y + c_x, v \approx \frac{f_y}{Z}Y + c_y$,Where (X, Y, Z) is a point in the camera coordinate system. Under the same conditions, a linear push-broom camera can also be approximated as a weak perspective camera. The imaging model of the linear push-broom camera can be modeled as: $u = a_1 x + b_1 y + c_1 z + d_1, v = \frac{a_2 x + b_2 y + c_2 z + d_2}{a_3 x + b_3 y + c_3 z + d_3}$,where (x, y, z) is a point in the scene coordinate system. The satellite image exhibits linearity along the row axis u and a weak perspective effect along the column axis v [7]. Assuming $f, \tilde{c}_x, \tilde{c}_y, \mathbf{R}, \mathbf{t}$ are the intrinsic and extrinsic parameters of the perspective camera along the column axis (with zero skew and a unit aspect ratio), we then use $\mathbf{R}.\mathbf{t}$ to transform the scene coordinates (x, y, z) into camera coordinates (X, Y, Z), and rewrite the linear push-broom camera model. Through approximation, the result is still a weak perspective camera [7]:

$$u = \hat{a}_1 X + \hat{b}_1 Y + \hat{c}_1 Z + \hat{d}_1$$
$$= \hat{a}_1 X + \hat{b}_1 Y + \hat{c}_1 \bar{Z} \cdot \left(1 + \frac{Z - \bar{Z}}{\bar{Z}}\right) + \hat{d}_1$$
$$\approx \hat{a}_1 X + \hat{b}_1 Y + \hat{c}_1 \bar{Z} + \hat{d}_1 \qquad (4)$$
$$\mathbf{v} = f\frac{Y}{Z} + \tilde{c}_y \qquad (5)$$

Comparing the two weak perspective camera models obtained from approximation yields the following equation:$\hat{a}_1 = \frac{f_x}{Z}, \hat{b}_1 = \frac{s}{Z}, \hat{c}_1 \bar{Z} + \hat{d}_1 = c_x, f = f_y, \tilde{c}_y = c_y$. By solving for f_x, f_y, s, c_x, c_y as well as $\mathbf{R}, \mathbf{t}$, we obtain approximate values for the parameters of the perspective camera model, which is referred to as the equivalent pinhole imaging model in other sections of this paper.

The above derivation only establishes the feasibility of a perspective approximation for the satellite camera within a local area. To actually solve for the approximate perspective camera model, three steps need to be carried out: placing virtual control points using the RPC model, fitting the projection matrix, and decomposing the projection matrix.

First, We generate a set of virtual control points from the RPC model (correspondences between 3D points in space and 2D points in the image plane), a process that can be referred to as constructing a virtual control grid in the field of remote sensing (also known as terrain-independent control points).Specifically, by evenly dividing the x, y, and z axes into M, M, and N sampling positions respectively, we discretize the ENU (East, North, Up) 3D bounding cube into a finite grid. Within this grid, $M \times M \times N$ grid sample points are generated with coordinates (x, y, z). These grid sample points are then converted from ENU coordinates to (latitude, longitude, elevation) coordinates and projected onto pixel coordinates (u, v) through the RPC model. This constructs a total of $M \times M \times N$ pairs of virtual control points $(x_i, y_i, z_i, u_i, v_i)$, where $i = 1, \dots, M \times M \times N$. Finally, we eliminate the virtual control points whose pixel coordinates fall outside the image boundaries.

Solve a 3×4 projection matrix $\mathbf{P} = [\mathbf{p}_1^T; \mathbf{p}_2^T; \mathbf{p}_3^T]$ based on these $M \times M \times N$ virtual control points. Assuming that the coordinates of the scene points are $\mathbf{x} = [x; y; z; 1]$, then the projection equations for our perspective camera are $u = \mathbf{p}_1^T \mathbf{x}/\mathbf{p}_3^T \mathbf{x}, v = \mathbf{p}_2^T \mathbf{x}/\mathbf{p}_3^T \mathbf{x}$.

Given L pairs of virtual control points $(\mathbf{x}_i, u_i, v_i), i = 1, \dots, L$, this paper employs the standard Direct Linear Transformation (DLT) method to solve for P. Subsequently, we decompose P into its canonical form K[R t]. It should be noted that the numerical solution of this projection matrix

TABLE II: Statistical values of object space and image space reprojection errors for virtual control points.

Image number	mean proj_err (pixels)	median proj_err (pixels)	max_proj_err (pixels)	mean inv_proj_err (meters)	median inv_proj_err (meters)	max inv_proj_err (meters)
00_WV03	0.016005	0.012989	0.08596	0.008019	0.006509	0.043071
01_WV03	0.024054	0.020372	0.116291	0.007606	0.006442	0.036777
02_WV03	0.024261	0.02036	0.113656	0.007645	0.006415	0.035801
03_WV03	0.028095	0.023443	0.154878	0.009569	0.007983	0.05277
04_WV03	0.027743	0.023766	0.139919	0.009661	0.008276	0.048704
05_WV03	0.024351	0.019732	0.122837	0.007989	0.006474	0.040283
06_WV03	0.026159	0.021457	0.140299	0.008849	0.007257	0.047496
07_WV03	0.0269	0.022059	0.141588	0.009112	0.007471	0.047978
08_WV03	0.022638	0.019161	0.1085	0.007429	0.006288	0.035592
09_WV03	0.022202	0.018554	0.097264	0.007232	0.006045	0.031669

will have a non-zero skew parameter, which is uncommon in typical images but can be explained by the equation $s = b_1 \bar{Z}$ in the derivation above. Satellite images are constructed by stitching together rows of images captured at slightly different moments; hence, these rows may not be perfectly aligned.

Using the equivalent pinhole imaging model to locally approximate the RPC model offers several advantages beyond its physical interpretability, simplicity, and efficiency. Firstly, it enables the transformation of the RPC model from satellite images into an equivalent pinhole imaging model for use in existing Structure from Motion (SfM) and Multi-view Stereo (MVS) workflows. Secondly, this approximation indicates that satellite images approximately possess linear epipolar geometry. The equivalent pinhole imaging model avoids dealing with the complex epipolar described by the RPC model, which is crucial for standard vision-based reconstruction processes. Thirdly, the bundle adjustment of the equivalent pinhole imaging model is simplified, whereas the bundle adjustment for the RPC model is computationally demanding and requires prior knowledge of the sensor.

B. Iterative resampling of images

The equivalent pinhole imaging model uses a 3×4 projection matrix in mathematics to replace the 90 RPC parameters for describing the mapping relationship from object space to image space in satellite imagery. This significantly reduces the number of parameters but also decreases the precision of the mapping relationship. We can reflect the accuracy of the model through the reprojection error of virtual control points, as shown in TABLE II, which presents the statistical results of the reprojection errors for the object and image sides of $100 \times 100 \times 20$ virtual control points. The local approximation range of the equivalent pinhole imaging model is $500m \times 500m$.

TABLE II lists the average, median, and maximum values of the image space projection errors and object space projection errors for a $500m \times 500m$ image. The results show that the image space projection errors are within 0.2 pixels, and the object space projection errors are less than 0.05 meters, indicating that the projection errors generated by using the projection matrix $\mathbf{P}$ to describe the local imaging relationship of satellite images are limited. In other words, within a certain level of accuracy, the pinhole imaging model can replace the RPC model.

In order to better describe the projection relationship between the object space and image space of satellite imagery using the projection matrix, this paper adopts a resampling method to correct the satellite images. Since there is a significant difference between the line array push-broom imaging and pinhole imaging, it is difficult to correct this difference using a simple affine transformation. This paper designs a high-order polynomial iterative resampling scheme to generate pseudo-images, as shown in Fig. 2.

Firstly, based on the UTM coordinates of the experimental area, the range of the object space region is roughly determined, and a virtual control point grid of $100 \times 100 \times 20$(corresponding to the ENU three dimensions) is constructed within the object space, as shown in the Fig. 1 by the blue point set P (hereinafter referred to as Blue P). According to the RPC model, the corresponding image point coordinates of Blue P are calculated to obtain the red point set P' (hereinafter referred to as Red P'), and the PCM (Pinhole Camera Model) parameters are calculated based on Blue P and Red P'. Then, using the PCM model, the corresponding image point coordinates of Blue P are calculated to obtain the green point set P (hereinafter referred to as Green P), and the distance between the homologous points of Green P and Red P' is referred to as the reprojection error. To reduce the reprojection error, Green P is used as the image point coordinates of the target image, and Red P' is used as the coordinates of the original image I (the image to be corrected), to calculate the polynomial coefficients of the transformation from Red P' to Green P (i.e., the polynomial coefficients from the original image to the target image); then, the original image is resampled using the polynomial model to generate a pseudo-image I_1, and the resampled coordinates of Red P' are Red P'', which completes the first resampling.Iterative resampling refers to the process of using P'' as the image point coordinates of the pseudo-image I_1, and using Green P as the image point coordinates of the target image, to calculate the polynomial coefficients for the transformation from Red P'' to Green P (i.e., the polynomial coefficients from the pseudo-image I_1 to the target image). Then, the pseudo-image I_1 is resampled using the polynomial model to generate the pseudo-image I_2, and the resampled coordinates of Red P'' become Red P'''. This completes the second iteration, and so on.
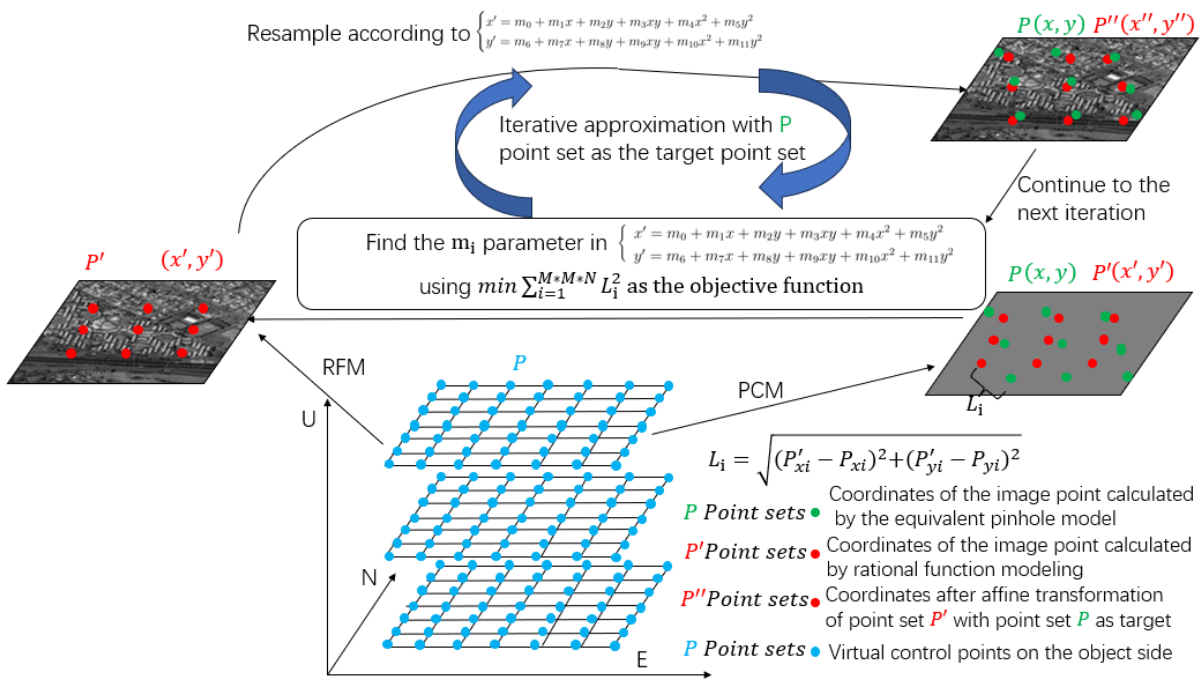
Fig. 2: Iterative resampling of satellite imagery

C. Skew parameter correction of the projection matrix

The pinhole model is used to locally approximate the RPC model, and the projection matrix is used to replace the RPC parameters to describe the imaging relationship between the object space and image space of satellite imagery. This reduces the number of parameters but also decreases the precision of the description. We can reflect the accuracy of the equivalent pinhole model by calculating the reprojection error of virtual control points. The skew parameter s included in the intrinsic matrix $\mathbf{K}$ of the equivalent pinhole model is not zero.In computer vision methods, the skew parameter in the intrinsic matrix is usually assumed to be zero, and existing computer vision methods cannot directly handle images with non-zero skew parameters.Therefore, it is necessary to correct the skew parameter of the intrinsic matrix. The method for skew correction is as follows: The original image is $\mathbf{I}$, and the corresponding intrinsic parameter matrix is $\mathbf{K}$. Decompose the intrinsic matrix $\mathbf{K}$ into $\mathbf{MK_1}$, where $\mathbf{M}$ is the transformation matrix, and $\mathbf{K_1}$ is the intrinsic parameter matrix with zero skew parameters.Then, apply the inverse of $\mathbf{M}$ to transform the original image to obtain the new image $\mathbf{I_1}$. The matrix $\mathbf{K_1}$ is the intrinsic parameter matrix of the new image $\mathbf{I_1}$, and this process is shown in Fig.3.

$\mathbf{X}$ represents the 3D point coordinates in the world coordinate system, and x represents the 2D point coordinates in the image plane coordinate system. P is a 3x4 homogeneous camera projection matrix. Under homogeneous coordinates, the relationship from a 3D point to a 2D point is given by x=PX, where P can be decomposed into the intrinsic matrix K, the rotation matrix $\mathbf{R}$, and the translation vector t, i.e., $\mathbf{P} = \mathbf{K[R|t]}$. In the equivalent pinhole camera model of satellite imagery, $\mathbf{K}$ includes a non-zero skew parameter s. To ensure that the skew parameter ss of the image is zero, the following

method is designed: $\mathbf{K}$ is decomposed into $\mathbf{MK_1}$, where $\mathbf{K_1}$ does not contain the skew parameter, and the skew parameter is factored into $\mathbf{M}$, thus obtaining $\mathbf{x} = \mathbf{MK_1[R|t]X}$. By rearranging, we get $\mathbf{M^{-1}}x = \mathbf{K_1[R|t]X}$. Multiplying the image point coordinates x on the left by M^{-1} is essentially resampling the original satellite image. Assuming the new image after resampling is $\mathbf{I_1}$, then $\mathbf{K_1}$ is the intrinsic matrix of $\mathbf{I_1}$. After correction, the new image $\mathbf{I_1}$ and the intrinsic matrix $\mathbf{K_1}$ with a skew parameter of 0 are obtained.

D. Chunking of satellite images and DSM stitching

The width of satellite images can reach tens to hundreds of kilometers. This paper has found that an increase in the image size corresponding to the equivalent pinhole camera model results in a larger reprojection error. To improve the accuracy of 3D reconstruction, a block processing method has been designed for large-format satellite images. First, the satellite image is uniformly blocked and cropped according to UTM coordinates (cropping RPC parameters at the same time); then, the cropped image blocks are equivalent to the pinhole camera model, and 3D reconstruction is performed using SfM and MVS; finally, the DSM generated from the image blocks are stitched together to obtain a 3D reconstruction product corresponding to the entire image. To ensure a certain degree of overlap when stitching DSMs later, each image block should be cropped with a certain degree of overlap.

In the experiment, the resolution of the generated DSM) was uniformly set to 0.3 meters, and UTM coordinates were used consistently, ensuring that all DSMs could be spatially aligned with a consistent resolution. For pixels in the overlapping areas, this experiment took the DSM value from any one of the DSMs as the value for that pixel, rather than taking the average or median value of the overlapping areas. This is because the
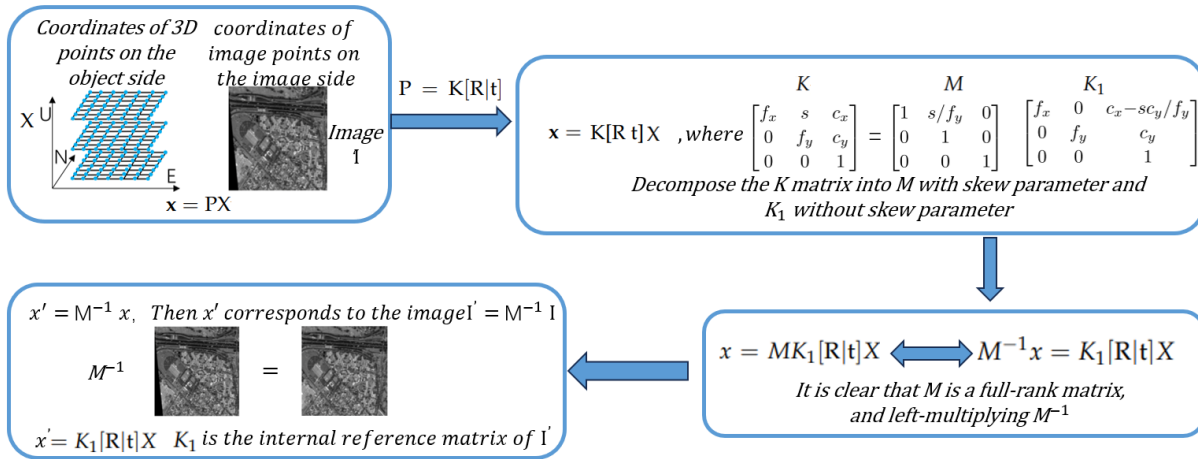
Fig. 3: Correction of the skew parameters of the projection matrix

more partitions there are, the larger the overlapping areas will be, and the stitching method that takes the average value will provide a more significant accuracy improvement for schemes with more partitions. The purpose of using this stitching method in this experiment was to ensure that any differences in the accuracy of the resulting DSM were entirely due to the different partitioning methods.

E. Accuracy metrics

The accuracy assessment metrics used in this paper are generally divided into two categories: one is for the assessment of reprojection error, and the other is for evaluating the reconstruction accuracy of the DSM.

Reprojection errors can be categorized into two types. The first type is image space reprojection error, which is the difference between the projected coordinates of the object space points in the image plane and their true image space coordinates. The second type is object space reprojection error, which is the difference between the projected coordinates of the image space points in object space and their true object space coordinates. Assuming the image space coordinates of the control point are x, the object space coordinates are X, and the projection matrix is P.

- The image space reprojection error $proj\_err$(pixels) can be expressed as: $proj\_err = |\mathbf{PX} - \mathbf{x}|$.
- Similarly, the object space reprojection error $inv\_proj\_err(pixels)$ (meters) can be expressed as: $inv\_proj\_err = |\mathbf{P}^{-1}\mathbf{x} - \mathbf{X}|$.
- Since there are many control points, we statistically analyze the reprojection errors and take the maximum, median, and mean values as assessment indicators of the reprojection error, resulting in six accuracy assessment metrics. The three evaluation indicators for the image space are denoted as $Mean\ P\_err$ (mean image space reprojection error), $Median\ P\_err$ (median image space reprojection error), and $Max\ P\_err$ (maximum image space reprojection error).
- The three evaluation indicators for the object space are denoted as $Mean\ I\_P\_err$ (Mean Inverse Reprojection

Error), $Median\ I\_P\_err$ (Median Inverse Reprojection Error), and $Max\ I\_P\_err$(Maximum Inverse Reprojection Error).

The accuracy assessment indicators for DSM are as follows:

- Error Map, under the condition of geographical coordinate alignment, the difference in elevation values for each grid is taken to obtain the error map. The method for calculating the error map is as follow: $err\_map = |DSM - Ground\_Truth|$, where DSM is the matrix representing the DSM, and Ground_Truth is the matrix representing the actual ground surface model,which is obtained from LiDAR point clouds;
- The Mean Absolute Error (MAE) in elevation is the average of the absolute values of elevation errors:$MAE = \frac{1}{N}\left(\sum_{i=1}^{N}\left|\hat{h}_i - h_i\right|\right)$, where N is the number of elements in the err_map matrix, $h_i$is the value of the i-th element in the err_map, and $\hat{h}_i$ is the mean value of the elements in the err_map;
- The $err\_num\_percentile$ refers to the error value corresponding to a specific percentile point when all elements of the error map are sorted in ascending order. The percentile error can reflect the distribution characteristics of the errors, similar to the role of an error histogram. For example, $err\_10\_percentile = 0.3m$ means that 10% of the elements in the err_map are less than $0.3m$.

In addition to accuracy assessment, we also evaluate the efficiency in terms of reconstruction time, which refers to the time span from the input of images to the generation of DSM products.

F. Experimental programme

We tested the effectiveness of the iterative resampling and block-based reconstruction methods proposed in this paper on the IARPA MVS 3DM dataset, which was used in the IARPA Multi-View Stereo 3D Mapping Challenge [31]. The dataset comprises 50 DigitalGlobe WorldView-3 panchromatic images covering a 100 square kilometer area near San Fernando, Argentina. These images have a minimum ground sampling distance of 30 cm and were acquired over a 14 month period,

TABLE III: Partitioning method for satellite image block reconstruction.

Partitioning Method	Reconstruction range	GSD	Number of partitions	Size of partition	Number of pixels	Overlap range
Method 1	500m×500m	0.3m	1	500m×500m	1650×1650	20m
Method 2	500m×500m	0.3m	4	260m×260m	850×850	20m
Method 3	500m×500m	0.3m	9	180m×180m	600×600	20m
Method 4	500m×500m	0.3m	16	140m×140m	450×450	20m

with most images taken on different dates. The dataset also includes airborne lidar-derived terrain truth at a 30 cm resolution for a $20km^2$ subset of the area. During the evaluation process, the ground truth and the reconstructed point cloud are projected onto a pre-designated geographic grid with a specification of 0.3 meters by 0.3 meters, resulting in two elevation maps with a resolution of 0.3 meters. Then, these maps are aligned and their differences are calculated to generate an error map (err_map). Based on the err_map, evaluation metrics are computed, including the percentile elevation error $err\_num\_percentile$ and the mean absolute elevation error MAE.

The experimental data used in this paper are World-View3 multiview stereoscopic images (50 views).the equipment used for the experiment is a laptop computer, running on 32G of RAM, and the Central Processing Unit is AMD_Ryzen_9_6900hx with radeon graphics 16, and the graphics card is NVIDIA _Corporation/AMD_Yellow_carp.

The experimental plan for this paper is as follows: Experiment 1: Calculate the reprojection error of the equivalent pinhole model to explore the feasibility of approximating a pinhole camera. If the reprojection error is within a certain tolerance, it indicates that satellite images described by the RPC model can be locally approximated using the projection matrix P. Analyze the changes in reprojection error with iterative resampling correction to verify the effectiveness of the iterative image resampling method proposed in this paper.

Experiment 2: Compare the reprojection errors of equivalent pinhole imaging models for satellite images of different sizes to explore the relationship between the size of satellite images and the reprojection error of the equivalent pinhole imaging model.

Experiment 3: Partition-based Reconstruction of Satellite Images and DSM Mosaicking. This study involves the 3D reconstruction of partitioned satellite images and the subsequent mosaicking of the generated DSM. The experiment evaluates the accuracy of DSMs produced by various reconstruction schemes and analyzes the optimal partitioning strategy. The paper presents four partitioning schemes as illustrated in TABLE III.

IV. RESULTS

A. Reprojection error after iterative resampling of satellite images

In Section III A, we theoretically derived the possibility of locally approximating the PRC model with a pinhole model and experimentally validated that the reprojection error of virtual control points calculated using the equivalent pinhole model is within a certain tolerance range. On the premise

that the equivalent pinhole imaging model is feasible, this paper also designed an iterative resampling scheme to further reduce the reprojection error. We tested the reprojection errors of satellite images after different numbers of iterative resamplings, with the test results of three images shown in Fig.4. The first row of the Fig. 4 set displays the reprojection errors of virtual control points for the uncorrected images. The second row of the Fig. 4 set shows the reprojection errors of virtual control points for images after one iteration of resampling, with the third and fourth rows following accordingly. The figure in Fig. 4 named Resample_1_00 indicates that the number of iterations is 1, and the image number is 00.

For 50 WorldView-3 images, an iterative resampling test was conducted on the reprojection error of the approximate pinhole model. The experiment showed that although the initial reprojection error varied from image to image, the reprojection error gradually decreased during multiple resampling processes, presenting the pattern shown in the three example images of Fig. 4. This iterative resampling scheme effectively resolved the projection error issues stemming from the insufficient local approximation accuracy of the pinhole imaging model.

To further test the impact of the iterative resampling scheme on the accuracy of 3D reconstruction from satellite imagery, this paper uses lidar-scanned terrain data as the ground truth to assess the accuracy of the reconstructed DSM. As shown in Fig. 5, the reconstruction area in this experiment is a square region with a side length of 500 meters. The yellow and dark green in the figure represent the magnitude of the error of the reconstructed DSM compared to the ground truth, with the specific mapping relationship shown in the legend's color band. It can be seen from the error map that the errors of the DSM are mainly concentrated in areas with significant elevation changes, such as the edges of residential houses, vegetation, and roads. The experimental results show that the 3D reconstruction error range of the resampled images is smaller, and the boundary of the corresponding error map is clearer. The essence of image resampling is to rearrange the local pixels on the satellite image, generating pseudo-images by reconfiguring the pixel arrangement on the satellite image. This makes the mapping relationship between the object space and the image space in the strict imaging model of the pseudo-image more accurate, thereby improving the accuracy of 3D reconstruction.

In order to quantitatively study the changes in the accuracy of the DSM before and after resampling, this paper calculates the percentage error values in the DSM error map, as shown in TABLE IV. Curve graphs were created based on the percentage errors, as shown in Fig. 6 , Fig. 7, corresponding to the 2p5d reconstruction results and 3D reconstruction results in Fig. 1,
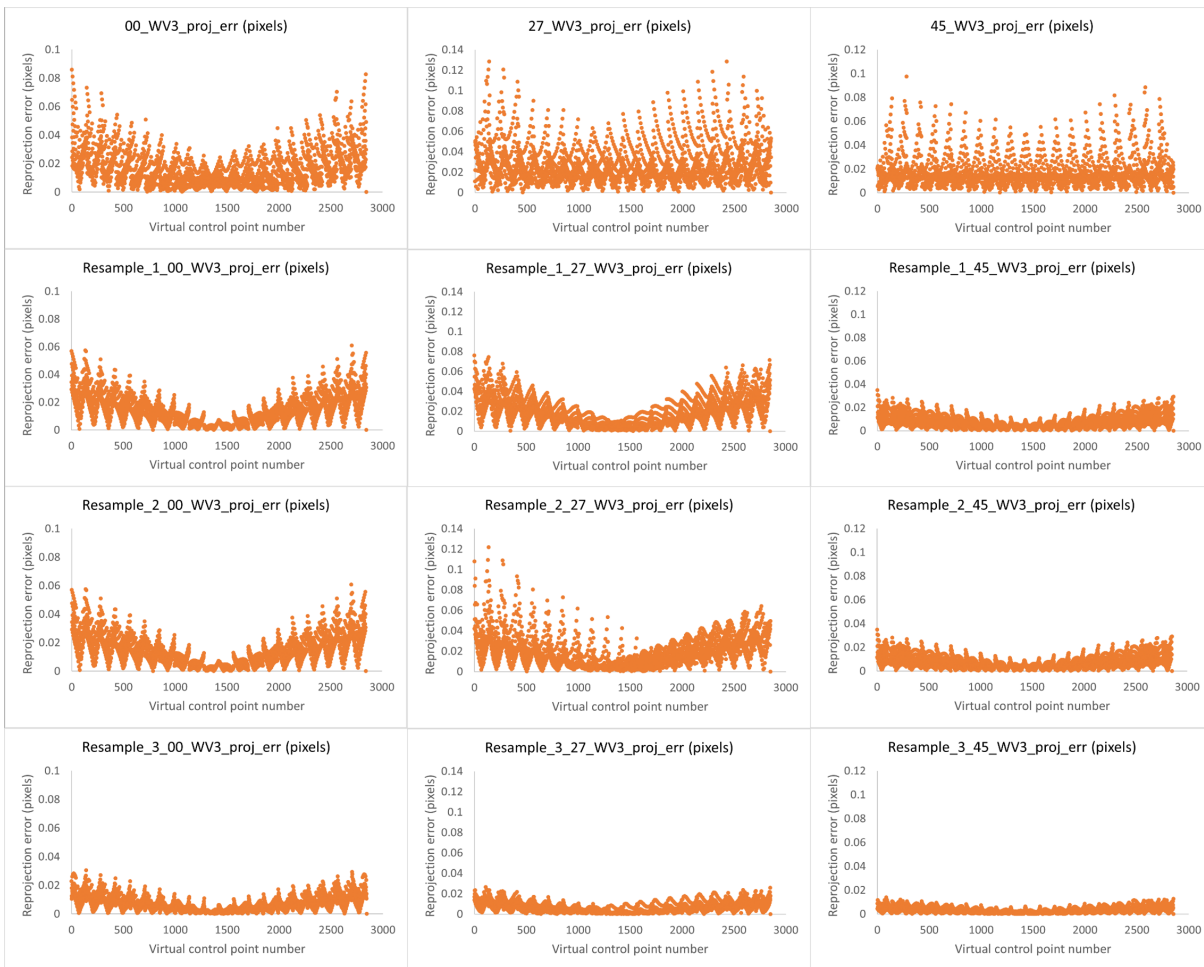
Fig. 4: Reprojection error of the equivalent pinhole camera for different image sizes

respectively. The curves in Fig. 6 and Fig. 7 indicate that in the percentile error statistics curve, the DSM error reconstructed using the resampled image is less than the error of the DSM reconstructed using the original image. This suggests that using the iterative resampling method can improve the reconstruction accuracy of the DSM.

TABLE IV: Statistical values of object space and image space reprojection errors for virtual control points.

%	2p5d 0_iters (pixels)	2p5d 1_iters (pixels)	2p5d 2_iters (pixels)	3d 0_iters (pixels)	3d 1_iters (pixels)	3d 2_iters (pixels)
0	0	0	0	0	0	0
10	0.180592	0.06275	0.058563	0.135599	0.073566	0.071903
20	0.27096	0.127726	0.120243	0.245644	0.150173	0.145491
30	0.343691	0.201033	0.187553	0.345008	0.234657	0.224815
40	0.417559	0.293962	0.268705	0.447403	0.336418	0.316416
50	0.518047	0.425749	0.382908	0.569099	0.471537	0.436619
60	0.701374	0.636036	0.585171	0.740398	0.671076	0.615639
70	1.097548	1.061097	1.038439	1.074598	1.009964	0.965623
80	2.164554	2.104804	2.128172	2.009983	1.85307	1.892384
90	5.632227	5.452448	5.482693	5.394635	5.030376	5.093053

B. Reprojection errors for regions of different sizes

In Section IV A, experiments have shown that pseudo-images generated by iterative resampling can reduce the reprojection error of the equivalent pinhole model. To further enhance the accuracy of 3D reconstruction, we conducted tests using images of different extents to explore the relationship between image size and the reprojection error of the equivalent pinhole camera, with the specific results as shown in Fig. 8.

Fig. 8 displays the test results from 50 WorldView-3 images, encompassing six assessment indicators: image space projection errors (mean, maximum, median) and object space projection errors (mean, maximum, median). The findings indicate that the smaller the image area corresponding to the equivalent pinhole model, the lower the reprojection error of the control points, and the better the approximation of the RPC model by the projection matrix. These experimental results also provide a strategy for 3D reconstruction of satellite imagery based on the equivalent pinhole model. This strategy involves dividing large satellite images into several smaller sections, treating each section as an equivalent pinhole camera model, conducting 3D reconstruction based on these equivalent pinhole models, and then mosaicking the reconstructed DSMs to obtain a DSM for the entire area.
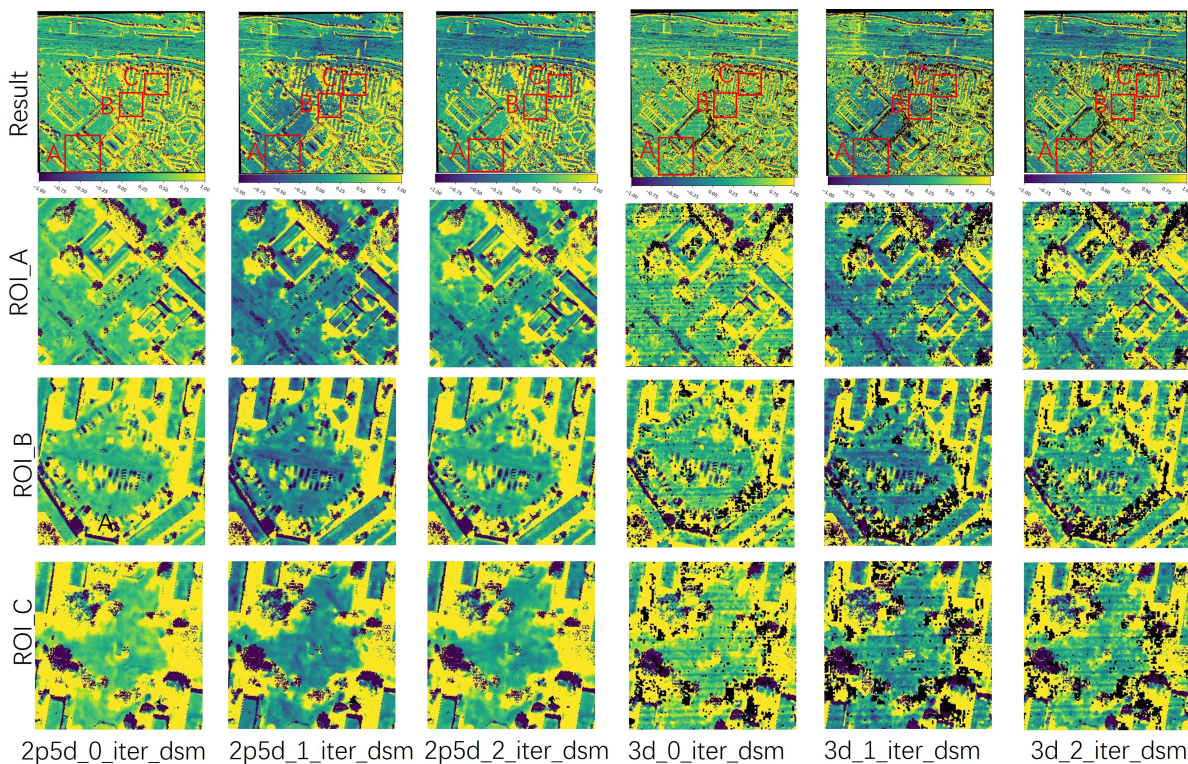
Fig. 5: Error map of 3D reconstruction of resampled images with different iteration times
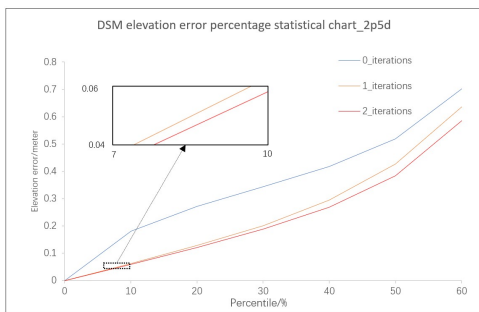

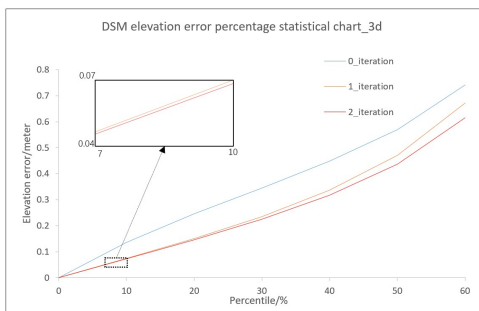Fig. 6: Relationship between DSM error and resampling times_2p5d

Fig. 7: Relationship between DSM error and resampling times_3d

C. Satellite Image Partition Reconstruction and DSM Mosaicing

In Section IV B, experiments revealed that the reprojection error of the RPC model approximated by the pinhole imaging model is related to the extent of the local approximation; the smaller the approximation range, the smaller the reprojection error. Against this backdrop, in order to further enhance the accuracy of 3D reconstruction, a block-based reconstruction scheme for DSM was designed in the Section III D. The experimental results are depicted in Fig. 9. This paper tabulates the reprojection errors of the approximated pinhole imaging model and the errors in the reconstructed DSM during the block-based reconstruction process, as shown in TABLE V. A stacked curve of DSM errors was plotted to reflect the distribution of DSM errors, as illustrated in Fig. 10.

The results from Fig. 9 demonstrate that the brightness values of the error maps for the block-reconstructed DSM are lower, indicating higher accuracy of the DSMs. Due to the block processing of the target area, there are differences in accuracy between different regions, leading to noticeable stitching seams on the error maps. Combining the data in TABLE IV, it can be observed that increasing the number of blocks reduces the area of each individual region, decreases the reprojection error, and enhances the accuracy of the reconstructed DSM, but at the same time, it consumes a longer duration.

As shown in Fig. 9, there are terrain differences across the various experimental regions. Although the reconstruction accuracy varies, with higher accuracy in the plains and lower accuracy in areas with significant elevation variation, both methods proposed in this study effectively improve the accuracy of the equivalent pinhole imaging model, regardless of the terrain type.

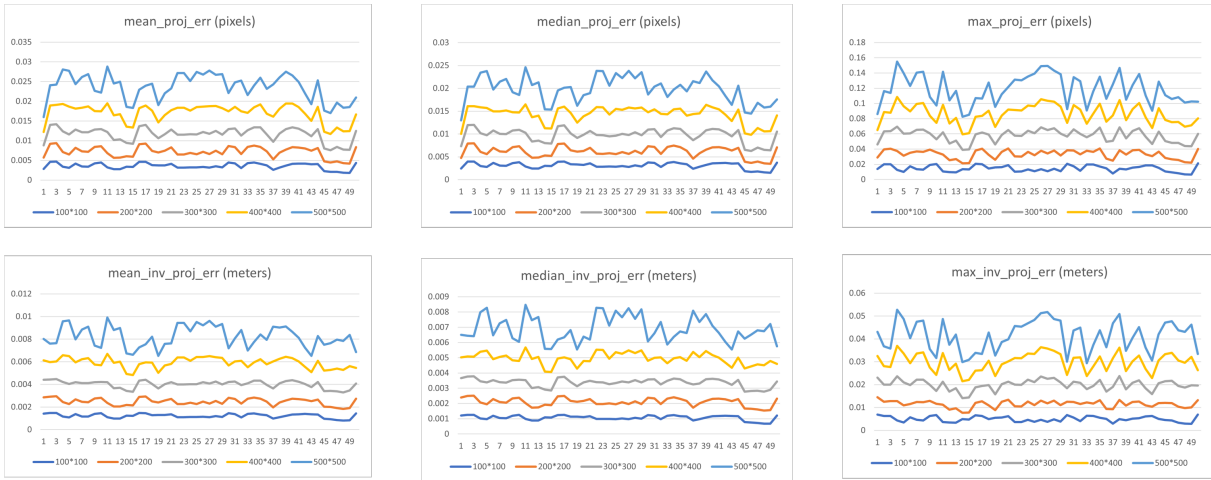Fig. 10 represents the stacked error curves of the DSMs

Fig. 8: Reprojection error of the equivalent pinhole camera for different image sizes
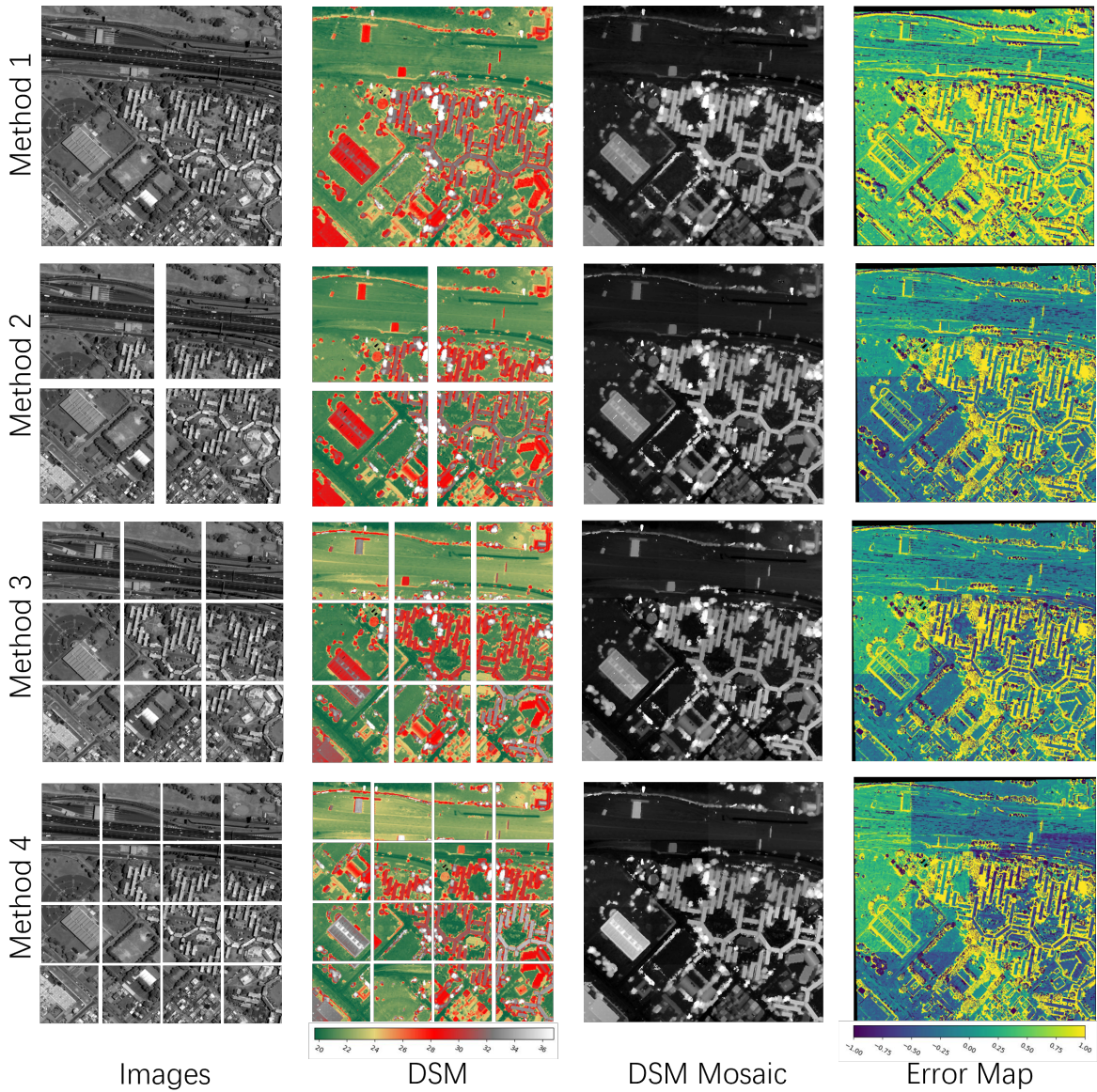

Fig. 9: Reprojection error and elevation error of DSM reconstruction by different zoning schemes

TABLE V: Reprojection Errors and DSM Elevation Errors of Equivalent Pinhole Models for Different Subdivision Schemes.

| Method No.	Mean P_err /pixel	Median P_err /pixel	Max P_err /pixel	Mean I_P_err /meter	Median I_P_err /meter	Max I_P_err /meter	2p5d MAE /meter	3d MAE /meter	Time /min
Method1	0.02366	0.01987	0.11879	0.00821	0.00690	0.04140	0.23783	0.31313	32.71238
Method2	0.00974	0.00821	0.04734	0.00336	0.00283	0.01641	0.13391	0.23077	57.59312
Method3	0.00638	0.00546	0.02906	0.00220	0.00188	0.01004	0.09542	0.19195	81.32990
Method4	0.00491	0.00424	0.02122	0.00169	0.00146	0.00732	0.09074	0.19809	111.78648
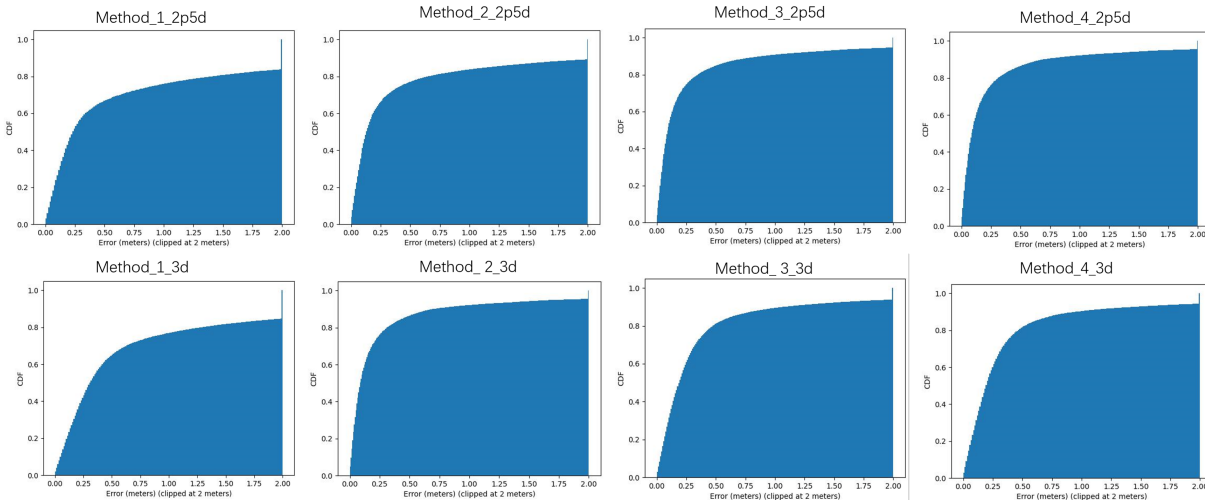
Fig. 10: Reprojection error of the equivalent pinhole camera for different image sizes

generated by different partitioning schemes. From the error stacking curves, it can be observed that the smaller the partitioning range, the fuller the convex curve of the error stacking graph, meaning that the error values of the DSM grids are more distributed on the side closer to zero. Therefore, regardless of whether a 2.5D DSM reconstruction scheme or a 3D DSM reconstruction scheme is used, block-based reconstruction can improve the accuracy of DSM reconstruction.

D. Analysis and Discussion

Zhang et al. selected three areas on the MVS3D dataset to test the feasibility of the equivalent pinhole model. The experimental results indicated that the precision and speed of the equivalent pinhole model reconstruction method were comparable to S2P, and even with the addition of MRF (Markov Random Fields) optimization, the generated DSM achieved higher accuracy than S2P [7].Hong et al. employed an improved equivalent pinhole camera model, denoted as PREM in the text, to generate DSMs with the smallest Root Mean Square Error (RMSE) compared to other methods [30]. Building upon the foundational work of previous researchers, we have developed an iterative resampling method and a block-based reconstruction strategy, thereby enhancing the accuracy of DSM reconstruction based on the equivalent pinhole imaging model.

The iterative resampling method also has its limitations. In order to preserve the features of the image, cubic spline interpolation algorithms are required during the resampling process, coupled with multiple iterations of sampling, the process of generating pseudo-images involves significant com-

putational consumption. Iterative resampling, in its pursuit of "high precision" equivalence, inevitably leads to the rearrangement of local pixels, which may disrupt the information contained in the original local image. Excessive resampling may destroy the original neighborhood relationships between pixels, leading to distortion in the resampled pseudo-images. Considering that the enhancement of image precision with the number of iterative resamplings will reach saturation, the number of iterative resamplings in the experiment should be appropriate.

In the partition reconstruction experiment, we observed that smaller areas had lower reprojection errors in the equivalent pinhole model and higher accuracy in the reconstructed DSM. This is likely due to the fact that larger ground areas exhibit more significant height differences, making the imaging relationship between the object space and image space of the corresponding image more complex. The complex imaging relationship leads to increased reprojection errors in the local approximation of the RPC model by the equivalent pinhole model, thereby introducing greater errors in 3D reconstruction. However, if the reconstruction area is too small, it may result in inaccurate depth estimation and imprecise positioning. To accurately align the reconstructed DSMs, a greater number of control points or actual DSM data are required. Without proper alignment and calibration, the partitioned DSMs may exhibit noticeable overall shifts. Furthermore, the partitioned DSMs need to be stitched together, which means maintaining overlap between partitions is essential. It should also be considered that excessive partitioning can lead to excessive overlap, thereby reducing the efficiency of reconstruction.

V. CONCLUSIONS

To better adapt satellite imagery to 3D reconstruction methods in the field of computer vision, this paper investigates the differences between the RPC model and the projection matrix, and analyzes the causes and patterns of reprojection errors in the equivalent pinhole imaging model. The projection matrix with fewer parameters cannot accurately fit the complex object-to-image mapping relationship of satellite imagery. To reduce the reprojection errors when fitting satellite imagery with the projection matrix, this paper designs the following two methods: First, generating pseudo-images through multiple resamplings to better satisfy the object-to-image mapping relationship of the pinhole imaging model; second, the imagery is divided into blocks and approximated locally as a pinhole imaging model, which reduces the size of the satellite imagery fitted by the equivalent pinhole model and decreases the complexity of the object-to-image mapping relationship. The test results on the MVS3D dataset demonstrate that this paper has improved the 3D reconstruction accuracy of satellite images by optimizing the reprojection error, enabling the equivalent pinhole imaging model to be better applied to the 3D reconstruction of remote sensing satellite images.

It should be noted that although both methods can improve accuracy of 3D reconstruction, they also have the following limitations: Iterative resampling to reduce reprojection errors has a certain degree of limitation;excessive resampling can disrupt the adjacency of pixels in the original image and the epipolar geometry, consuming a huge amount of computational power without significantly enhancing the accuracy of reconstructed DSM. Image clipping and block processing can improve the internal accuracy of local area DSM reconstruction, but if the block area is too small, it can cause an overall shift in the generated DSM (reduced external accuracy), requiring control point data or true value DSM to correct the shift. Moreover, an increase in the number of blocks requires the reconstruction of more overlapping areas for stitching and mosaicking, which consumes more computational power. Therefore, we need to set reasonable numbers of iterations and partitioning schemes based on the specific scene's requirements for accuracy and efficiency.

VI. BIOGRAPHY SECTION

Yongjian Li obtained a Bachelor's degree in Remote Sensing Science and Technology from the Information Engineering University in 2022. He is currently pursuing a Doctoral degree in Photogrammetry and Remote Sensing at the College of Geospatial Information, Information Engineering University. His main research directions include photogrammetry and remote sensing, 3D reconstruction and computer vision.

Song Ji received the Ph.D. degree in photogrammetry and remote sensing from Information Engineering University, in 2012. He is currently an Associate Professor with Information Engineering University. His current research interests include image processing and 3D reconstruction.

Danchao Gong received his PhD in Photogrammetry and Remote Sensing from the College of Geospatial Information of Information Engineering University in 2003. He is a researcher and doctoral supervisor at Xi'an Institute of Surveying and Mapping.

Dazhao Fan received the Ph.D. degree in photogrammetry and remote sensing from Information Engineering University, in 2007. He is currently a professor and doctoral supervisor at the Information Engineering University. His current research interests include digital photogrammetry and its applications.

Jiaxuan Song graduated from the Information Engineering University with a degree in Remote Sensing Science and Technology in 2020.She received her Master's degree in Resources and Environment from Information Engineering University in 2023. He is currently pursuing the PhD degree in Photogrammetry and Remote Sensing in School of Geospatial Information, Information Engineering University. His main research interests include photogrammetry and remote sensing, 3D reconstruction.

Ding Gao obtained a Bachelor's degree in Surveying and Mapping Engineering from North China Institute of Science and Technology in Hebei, China, in 2021. He is currently pursuing a Master's degree in Resources and Environment at the College of Geospatial Information, Information Engineering University. His main research directions include photogrammetry and remote sensing, 3D reconstruction and computer vision.

References

[1] M. Bosch, Z. Kurtz, S. Hagstrom, and M. Brown, "1," in *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2017.

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski, "Building rome in a day," in *2009 IEEE 12th International Conference on Computer Vision (ICCV)*, 2009.

[3] S. Fuhrmann, F. Langguth, and M. Goesele, "Mve - a multi-view reconstruction environment," *Eurographics Association*, 2014.

[4] G. Facciolo, C. D. Franchis, and E. Meinhardt-Llopis, "Automatic 3d reconstruction from multi-date satellite images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.

[5] Y. Hu, V. Tao, and A. Croitoru, "Understanding the rational function model: Methods and applications," in *International Archives of Photogrammetry and Remote Sensing*, 2004.

[6] J. L. Schonberger and J. M. Frahm, "Structure-from-motion revisited," in *IEEE Conference on Computer Vision Pattern Recognition*, 2016, pp. 4104–4113.

[7] K. Zhang, N. Snavely, and J. Sun, "Leveraging vision reconstruction pipelines for satellite imagery," *IEEE*, 2019.

[8] Y. Hu, V. Tao, and A. Croitoru, "Understanding the rational function model: Methods and applications," in *International Archives of Photogrammetry and Remote Sensing*, 2004.

[9] T. Saxena, Richard I. Hartley, "The cubic rational polynomial camera model," in *Image Understanding Workshop*, 1997.

[10] C. S. Fraser and H. B. Hanley, "Bias-compensated rpcs for sensor orientation of high-resolution satellite imagery," *Photogrammetric Engineering Remote Sensing*, vol. 71, no. 8, pp. 909–915, 2005.

[11] J. Grodecki and G. Dial, "Block adjustment of high-resolution satellite images described by rational polynomials," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 1, pp. 59–68, 2003.

[12] R. A. Beyer, O. Alexandrov, and S. McMichael, "The ames stereo pipeline: Nasa's open source software for deriving and processing terrain data," *Earth and Space Science*, vol. 5, no. 9, pp. 537–548, 2018.

[13] T. Saxena, Richard I. Hartley, "The cubic rational polynomial camera model," in *Image Understanding Workshop*, 1997.

[14] P. D'Angelo and P. Reinartz, "Semiglobal matching results on the isprs stereo matching benchmark," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3819, pp. 79–84, 2012.

[15] G. Kuschk and P. D'Angelo, "Fusion of multi-resolution digital surface models," 2013.

[16] O. Martinez-Rubi and F. Nex..., "Improving foss photogrammetric workflows for processing large image datasets," *Open Geospatial Data Software Standards*, vol. 2, no. 1, 2017.

[17] C. D. Franchis, E. Meinhardtllopis, J. Michel, J. M. Morel, and G. Facciolo, "An automatic and modular stereo pipeline for pushbroom images," 2014.

[18] G. Facciolo, C. D. Franchis, and E. Meinhardt-Llopis, "Automatic 3d reconstruction from multi-date satellite images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017.

[19] Govindu and M. V., "[ieee proceedings of the 2004 ieee computer society conference on computer vision and pattern recognition, 2004. cvpr 2004. - washington, dc, usa (27 june - 2 july 2004)] proceedings of the 2004 ieee computer society conference on computer vision and patte," vol. 1, pp. 684–691, 2004.

[20] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, 2008.

[21] K. Wilson and N. Snavely, "Robust global translations with 1dsfm," *Springer, Cham*, 2014.

[22] J. L. Schonberger and J. M. Frahm, "Structure-from-motion revisited," in *IEEE Conference on Computer Vision Pattern Recognition*, 2016, pp. 4104–4113.

[23] Y. Yu and Y. Zhang, "A incremental structure from motion method of robustness enhancement and accuracy improvement," *Journal of Geodesy and Geoinformation Science*, vol. 48, no. 2, p. 9, 2019.

[24] J. Xue and H. Yi, "A hybrid multi-view 3d reconstruction method based on scene graph partition," *ACTA AUTOMATICA SINICA*, vol. 46, no. 4, p. 14, 2020.

[25] R. Qin, A. Gruen, and C. Fraser, "Quality assessment of image matchers for dsm generation - a comparative study based on uav images," in *35th Asian Conference on Remote Sensing 2014, ACRS 2014: Sensing for Reintegration of Societies*, 2014.

[26] P. Wang, L. Shi, B. Chen, Z. Hu, Q. Dong, and J. Qiao, "Pursuing 3d scene structures with optical satellite images from affine reconstruction to euclidean reconstruction," 2022.

[27] C. Bao, W. Pinhe, and D. Qiulei, "Fast 3d reconstruction of satellite images via the global affine model," *Journal of Remote Sensing*, vol. 28, no. 6.

[28] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo," in *Computer Vision– ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14.* Springer, 2016, pp. 501–518.

[29] C. Wu, "Towards linear-time incremental structure from motion," in *2013 International Conference on 3D Vision-3DV 2013.* IEEE, 2013, pp. 127–134.

[30] D. Hong, A. Yu, S. Ji, X. Lu, W. Guo, X. Cao, and C. Qiu, "Refined equivalent pinhole model for large-scale 3d reconstruction from spaceborne ccd imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 134, p. 104164, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S156984322400520X

[31] M. Bosch, Z. Kurtz, S. Hagstrom, and M. Brown, "A multiple view stereo benchmark for satellite imagery," in *2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR).* IEEE, 2016, pp. 1–9.