

Article

An Improved 3D Reconstruction Method for Satellite Images Based on Generative Adversarial Network Image Enhancement

Henan Li ^{1,2}, Junping Yin ^{2,3,*} and Liguo Jiao ^{1,2,*}

¹ Academy for Advanced Interdisciplinary Studies, Northeast Normal University, Changchun 130024, China; lihn117@nenu.edu.cn

² Shanghai Zhangjiang Institute of Mathematics, Shanghai 201203, China

³ Institute of Applied Physics and Computational Mathematics, Beijing 100094, China

* Correspondence: yinjp829829@126.com (J.Y.); jiaolg356@nenu.edu.cn (L.J.)

Abstract: Three-dimensional reconstruction based on optical satellite images has always been a research hotspot in the field of photogrammetry. In particular, the 3D reconstruction of building areas has provided great help for urban planning, change detection and emergency response. The results of 3D reconstruction of satellite images are greatly affected by the input images, and this paper proposes an improvement method for 3D reconstruction of satellite images based on the generative adversarial network (GAN) image enhancement. In this method, the perceptual loss function is used to optimize the network, so that it can output high-definition satellite images for 3D reconstruction, so as to improve the completeness and accuracy of the reconstructed 3D model. We use the public benchmark dataset of satellite images to test the feasibility and effectiveness of the proposed method. The experiments show that compared with the satellite stereo pipeline (S2P) method and the bundle adjustment (BA) method, the proposed method can automatically reconstruct high-quality 3D point clouds.

Keywords: optical satellite imagery; 3D reconstruction; deep learning; generative adversarial network (GAN); RPC model



Citation: Li, H.; Yin, J.; Jiao, L. An Improved 3D Reconstruction Method for Satellite Images Based on Generative Adversarial Network Image Enhancement. *Appl. Sci.* **2024**, *14*, 7177. <https://doi.org/10.3390/app14167177>

Academic Editor: Mauro Lo Brutto

Received: 16 July 2024

Revised: 11 August 2024

Accepted: 13 August 2024

Published: 15 August 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, optical satellite remote sensing technology has developed rapidly, which has further reduced the cost of obtaining satellite imagery. Compared with other photogrammetry techniques, 3D reconstruction from satellite imagery has the advantages of wide coverage and no geographical limitations [1–3]. As we all know, the input of 3D reconstruction from satellite imagery can be roughly divided into the method for stereo pair and the method for multi-view, among which the stereo pair 3D reconstruction method needs only two satellite images as input; still, the multi-view 3D reconstruction method requires more than two satellite images with different view angles as input. Moreover, multi-view 3D reconstruction consists of two methods: true multi-view and multi-view stereo [4]. The true multi-view method processes all the satellite images simultaneously, while the multi-view stereo method selects stereo pairs from all the satellite images and then processes them by the 3D reconstruction method for stereo pairs. It has been shown that the reconstruction accuracy of the true multi-view method is not as good as that of the multi-view stereo method [1,4]. Therefore, most of the commonly used 3D reconstruction methods of satellite imagery are based on the reconstruction method for stereo pairs.

The satellite stereo pair-based 3D reconstruction mainly includes three steps: stereo rectification, stereo matching, and altitude estimation. One of the purposes of stereo rectification is to make sure that there is only horizontal disparity in the satellite stereo pair. The stereo rectification process can reduce the matching complexity and save the matching time. The other purpose of stereo rectification is to eliminate the epipolar constraint error of satellite stereo pair as much as possible. Different from the traditional image

pinhole model, satellite imagery is imaged by a pushbroom camera, and the general Rational Polynomial Camera (RPC) model is often used to describe the imaging process [5]. However, pushbroom camera imaging will lead the epipolar line of the satellite stereo pair to the epipolar curve, resulting in an epipolar constraint error [6]. At present, a popular method of stereo rectification of satellite image pairs is affine model approximation [7], that is, the affine matrix is generated according to the matching point pairs of the stereo pairs, and then, the affine transformation is performed on the satellite image stereo pair. The affine model approximation method is simple and possesses high rectification accuracy.

Stereo matching is an important step in the 3D reconstruction of satellite images. The most commonly used stereo-matching method is using the semi-global matching (SGM) algorithm [8–11]. The SGM algorithm approximates the two-dimensional matching cost computation by means of one-dimensional path cost aggregation. However, the SGM algorithm is prone to generating error streaks. Therefore, some researchers propose the more global matching (MGM) algorithm [12], which extends the one-dimensional path cost aggregation to the two-dimensional quadrant aggregation. Even though this method improves the matching accuracy, it is time consuming. Besides the SGM and MGM algorithm, some researchers use the patch match method in the field of computer vision to implement stereo matching [13,14]. The patch match method has the problem of fixed disparity in the window, which will lead to the distortion of the edges in images during matching. With the popularity of deep learning, some researchers propose an intelligent stereo-matching method, such as using a convolutional neural network (CNN) to calculate the matching cost [15] and using smart aggregation strategy in the SGM algorithm [16]. Although these methods can obtain good matching results, they require a large number of satellite images for training.

The final step of the satellite images 3D reconstruction is altitude estimation, which uses triangulation to calculate 3D coordinates according to the disparity map generated by stereo matching. The triangulation of satellite imagery requires the RPC model [17]. Therefore, the accuracy of the RPC model is particularly important. Mari et al. proposed an RPC model refinement method based on bundle adjustment [18], which can improve the accuracy of satellite image 3D reconstruction.

From previous studies, it can be seen that the quality of satellite images has a great impact on 3D reconstruction [4,19]. Therefore, a GAN image enhancement-based 3D reconstruction method is proposed in this paper. The GAN is an unsupervised deep learning model [20], which has good applications in the fields of image generation, image completion, and image quality improvement [21–23]. The GAN consists of a generator and a discriminator. Some researchers add a local discriminator to improve the brightness of the images [24], but this will make the images lose some colors. Jiang et al. extracted and fused multi-scale features under the framework of the GAN [25]. This method has good results on image enhancement, but it lacks image details. There are also some researchers who integrate attention mechanisms and global features into the GAN structure to enhance image quality [26].

In view of the facts that the accuracy of the satellite images 3D reconstruction results depends heavily on the quality of the input images, we use the GAN model to enhance the input images to obtain high-quality satellite images for 3D reconstruction. In this paper, we propose an improved 3D reconstruction method of satellite images based on GAN image enhancement. The contributions of this paper are as follows:

1. A new improvement method for automatic 3D reconstruction of satellite images is proposed, which can generate high-quality reconstruction results without any ground control points.
2. The perceptual loss function is applied to GAN image enhancement to improve the clarity of satellite images, further improving the quality of 3D reconstruction results of satellite images.

This article is organized as follows: The key imaging model and the network used in the proposed method are introduced in Section 2. The framework and details of the

proposed method are described in Section 3. We compare our method with other satellite images 3D reconstruction methods and analyze the reconstruction results in Section 4. There are some useful conclusions in Section 5.

2. RPC Model and Generative Adversarial Network

As an approximation of the satellite imaging model, the RPC model is an important input for the 3D reconstruction of satellite images. In addition, the method proposed in this paper is based on GAN image enhancement. Therefore, this section will briefly introduce the principles of the RPC model and the GAN model.

2.1. RPC Model

For reasons of technical secrecy and extended utility, commercial satellite companies replaced physical sensor models with generic RPC models. The generic RPC model is a mathematical model that represents the mapping relationship between image coordinates and spatial 3D coordinates. Equation (1) shows the projection of the generic RPC model.

$$(x, y) = P(X, Y, Z) = \left(\frac{NumL(X, Y, Z)}{DenL(X, Y, Z)}, \frac{NumS(X, Y, Z)}{DenS(X, Y, Z)} \right), \tag{1}$$

where (x, y) and (X, Y, Z) represent the normalized coordinates of the image point and the 3D point, respectively, $P(\cdot)$ denotes satellite camera projection. As shown in Equation (2), the coordinates (x, y) and (X, Y, Z) are all normalized to the range of $[-1, 1]$. The original coordinates of the satellite image point and the 3D point are presented as (\tilde{x}, \tilde{y}) and $(\tilde{X}, \tilde{Y}, \tilde{Z})$, respectively. There are usually 10 normalization parameters in the RPC model, which contains 5 offset parameters and 5 scale parameters, as shown in Equation (2).

$$\begin{cases} X = \frac{\tilde{X} - LAT_OFF}{LAT_SCALE} \\ Y = \frac{\tilde{Y} - LONG_OFF}{LONG_SCALE} \\ Z = \frac{\tilde{Z} - HEIGHT_OFF}{HEIGHT_SCALE} \\ x = \frac{\tilde{x} - LINE_OFF}{LINE_SCALE} \\ y = \frac{\tilde{y} - SAMP_OFF}{SAMP_SCALE} \end{cases} \tag{2}$$

In Equation (1), $NumL(\cdot)$, $DenL(\cdot)$, $NumS(\cdot)$, and $DenS(\cdot)$ stand for the rational polynomials in RPC models, which have the following form:

$$\begin{aligned} NumL(X, Y, Z) &= a_0 + a_1Y + a_2X + a_3Z + a_4YX + a_5YZ + a_6XZ + a_7Y^2 + a_8X^2 \\ &+ a_9Z^2 + a_{10}XYZ + a_{11}Y^3 + a_{12}YX^2 + a_{13}YZ^2 + a_{14}Y^2X \\ &+ a_{15}X^3 + a_{16}XZ^2 + a_{17}Y^2Z + a_{18}X^2Z + a_{19}Z^3. \end{aligned} \tag{3}$$

The forms of $DenL(\cdot)$, $NumS(\cdot)$, and $DenS(\cdot)$ are the same as $NumL(\cdot)$, so that the RPC model has 80 rational polynomial coefficients. The RPC model, which has the advantages of no oscillation on the coefficients and independence from sensors, approximates the satellite imaging model in the form of rational functions.

2.2. Generative Adversarial Network

In this paper, the GAN model is used to improve the quality of the input satellite images, where I^{HQ} is used to represent high-quality satellite images and I^{LQ} is used to represent low-quality satellite images. The GAN model consists of a generator $G_\theta(\cdot)$ and a discriminator $D_\rho(\cdot)$, whose goal is to train a generator network that the generated satellite images are judged to be true by the discriminator. Here θ and ρ are the tuning parameters

to be calculated for the GAN model. These two sets of parameters contain the weights and biases of the generator network and the discriminator network. The θ and ρ are obtained by optimizing the loss function in Equation (4). Assuming that there are N low-quality satellite images and N high-quality satellite images to be trained, Equation (4) needs to be solved for the GAN image enhancement.

$$\min_{\theta} \frac{1}{N} \sum_{n=1}^N l^{IE} \left(G_{\theta} \left(I_n^{LQ} \right), I_n^{HQ} \right), \quad (4)$$

where l^{IE} indicates the loss function to be optimized. In addition to optimizing the generator network, the GAN model also needs a discriminator network, which is obtained by optimizing and solving Equation (5).

$$\min_{\theta} \max_{\rho} \left\{ \mathbb{E} \left[\log D_{\rho} \left(I^{HQ} \right) \right] + \mathbb{E} \left[\log \left(1 - D_{\rho} \left(G_{\theta} \left(I^{LQ} \right) \right) \right) \right] \right\}, \quad (5)$$

here we input the low-quality satellite images I^{LQ} into the generator and hope that the generator will output the high-quality satellite images I^{HQ} that can fool the discriminator. The discriminator is trained to distinguish between the real high-quality satellite images and the generated high-quality satellite images. When we solve Equation (5), we need to alternatively optimize the generating function and the discriminating function.

3. The Improved 3D Reconstruction Method of Satellite Images

In this paper, we propose an improved satellite image reconstruction method based on the GAN image enhancement model. The processing flow chart of the proposed reconstruction method is shown in Figure 1.

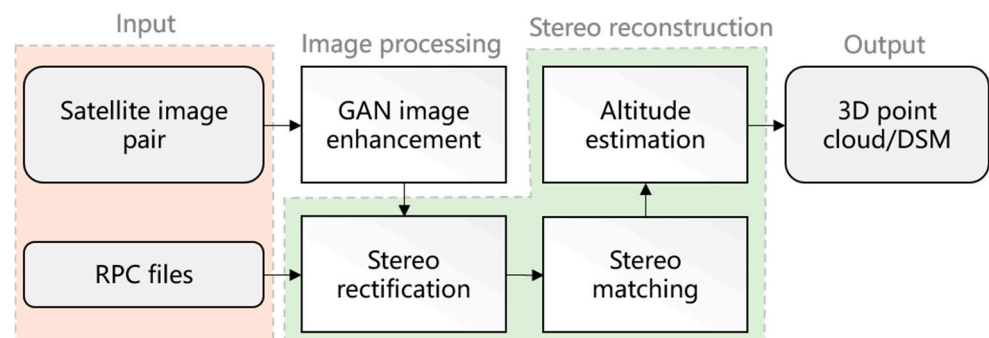


Figure 1. The processing flow chart of the proposed reconstruction method.

There are two types of input data, which are satellite image pair and the corresponding RPC files. Firstly, the proposed method implements enhancement processing on the input satellite images, which is based on the GAN model. Secondly, the enhanced satellite image pair and its corresponding RPC files are processed by stereo rectification and stereo matching. After these two processes, a dense disparity map will be generated. Triangulation-based altitude estimation is then performed on the disparity map, and finally, a high-quality 3D point cloud and its digital surface model (DSM) are generated.

3.1. GAN-Based Image Enhancement

According to the above introduction, the proposed method implements image enhancement on the satellite images by the GAN model. Here, we propose a modified generator network as shown in Figure 2, which is learned from the network designed in Ref. [21]. The generator network is mainly composed of the convolution layers, the batch-normalization layers, the residual blocks, the up-sampling block, and the down-sampling block. This generator network is trained to solve the minimization problem in Equation (5). We adopt Parametric-ReLU as the activation function and add a down-sampling block to

the generator network of the super-resolution GAN model to improve the quality of the output satellite images.

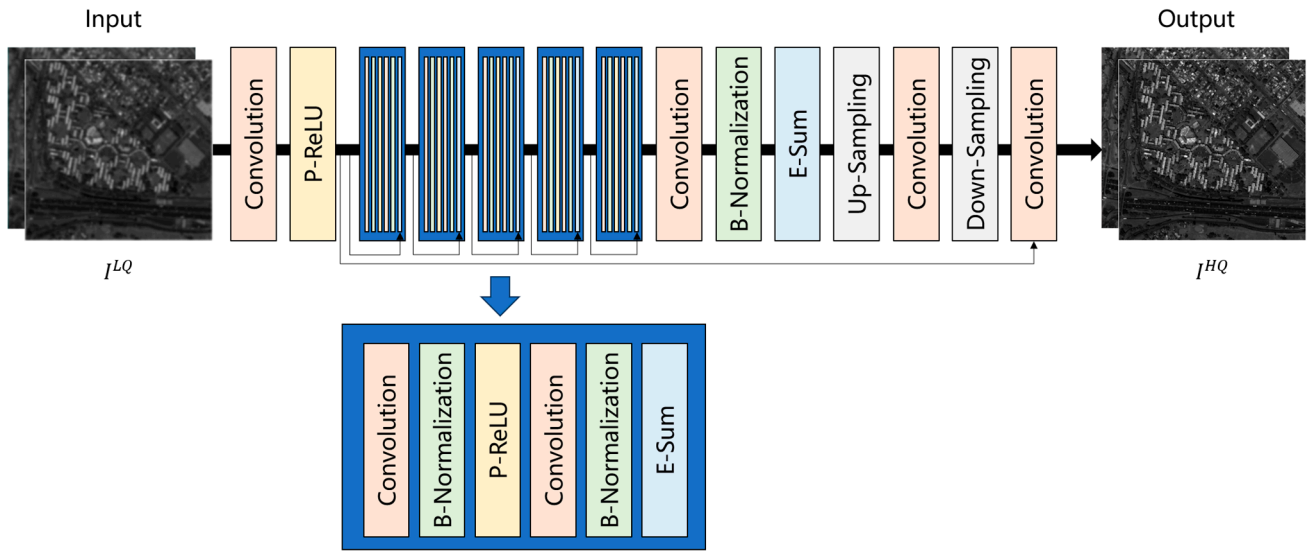


Figure 2. The architecture of the generator network in the proposed method.

Figure 3 shows the structure of the discriminator network, which is used to discriminate whether the samples generated by the generator network are high-quality satellite images. There are eight convolutional layers, seven batch-normalization layers, and two dense layers in the discriminator network. Here, we adopt Leaky-ReLU as the activation function, which is used to avoid maximum pooling [21]. The final activation function takes the form of the sigmoid function. We train this discriminator network to solve the maximization problem in Equation (5).

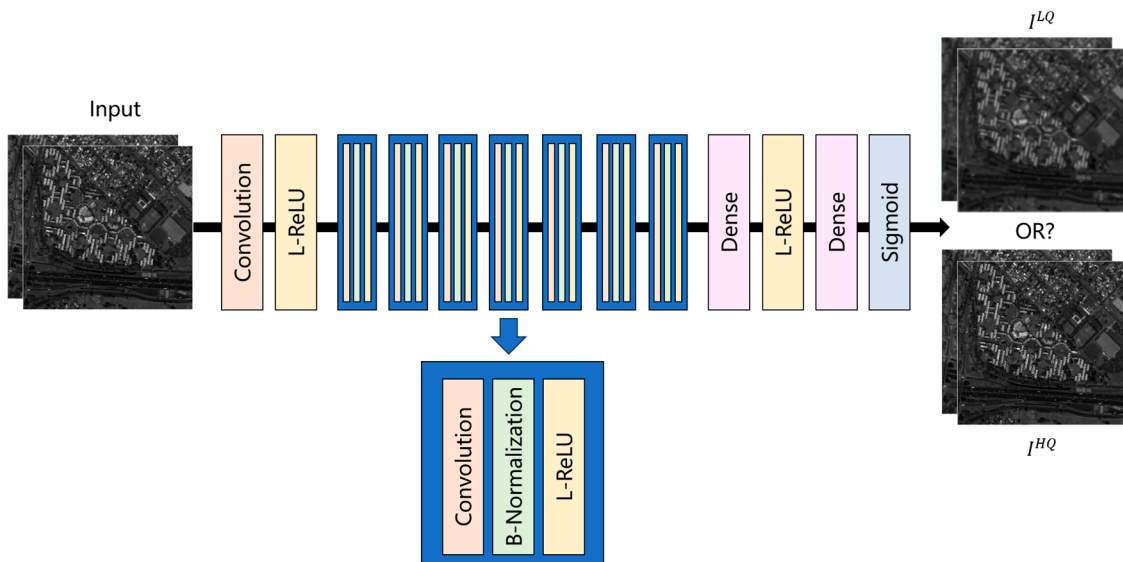


Figure 3. The architecture of the discriminator network in the proposed method.

The definition of the loss function directly affects the performance of the generator network. MSE is usually used to construct the loss function in image quality improvement research, but this method often lacks high-frequency details, which is inconsistent with our

desire to make the image sharper. Therefore, the proposed method adopts a combination of MSE loss, VGG loss, adversarial loss, and regularization loss, as shown in Equation (6).

$$l^{IE} = l_{MSE}^{IE} + 6 \times 10^{-3} l_{VGG}^{IE} + 10^{-3} l_{Adv}^{IE} + 2 \times 10^{-8} l_{Reg}^{IE}, \quad (6)$$

where l_{MSE}^{IE} represents the MSE loss, whose structure is as below:

$$l_{MSE}^{IE} = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H \left(I_{w,h}^{HQ} - G_{\theta} \left(I_{w,h}^{LQ} \right) \right)^2, \quad (7)$$

In Equation (7), W and H denote the width and the height of the high-quality satellite image. We define the MSE loss based on the Euclidean distance between the pixels of the reconstructed high-quality satellite image and the reference high-quality satellite image. The meaning of I^{HQ} , I^{LQ} , and $G_{\theta}(\cdot)$ are the same as in Equation (4). In Equation (6), l_{VGG}^{IE} represents the VGG loss in Equation (6), whose structure is shown in Equation (8).

$$l_{VGG}^{IE} = \frac{1}{W'H'} \sum_{w'=1}^{W'} \sum_{h'=1}^{H'} \left(\phi \left(I_{w,h}^{HQ} \right) - \phi \left(G_{\theta} \left(I_{w,h}^{LQ} \right) \right) \right)^2, \quad (8)$$

where W' and H' describe the dimensions of the feature maps in the VGG network, and $\phi(\cdot)$ stands for the obtained feature map. In fact, the VGG loss is defined based on the Euclidean distance between the reconstructed satellite image features and the reference satellite image features. As mentioned in Section 2.2, the generator network is important for the entire training process. In order to generate high-quality satellite images, we add adversarial loss to the loss function, as shown in Equation (9).

$$l_{Adv}^{IE} = - \sum_{n=1}^N \log D_{\rho} \left(G_{\theta} \left(I^{LQ} \right) \right), \quad (9)$$

Here, we define the adversarial loss based on the sum of the probabilities that the high-quality satellite image generated by the generator network $G_{\theta}(\cdot)$ will be judged true by the discriminator network $D_{\rho}(\cdot)$. We adopt the logarithmic form in Equation (9) to facilitate the solution later. In order to prevent the model from overfitting, the proposed method adds regularization loss to l^{IE} ,

$$l_{Reg}^{IE} = \frac{1}{WH} \sum_{w=1}^W \sum_{h=1}^H \left\| \nabla_{G_{\theta}} \left(I_{w,h}^{LQ} \right) \right\|_2^2. \quad (10)$$

We use the two-norm of the generator network gradient to construct the regularization loss.

3.2. Stereo Rectification

After the satellite images are enhanced by the GAN model, the proposed method performs stereo rectification on the enhanced satellite image pair. The purpose of this process is to reduce the epipolar constraint error of the image pair and make it suitable for the subsequent stereo matching. The principle of the epipolar constraint error is shown in Figure 4. As the matching point of x_m , ideally, the x'_m should coincide with the projection point of the X_m on the epipolar curve $epi_{uv}^{x'_m}(\cdot)$. However, the existence of the epipolar constraint error leads to the distance between the x'_m and the $epi_{uv}^{x'_m}(X_m)$, and the Euclidean distance between the two points is defined as the epipolar constraint error.

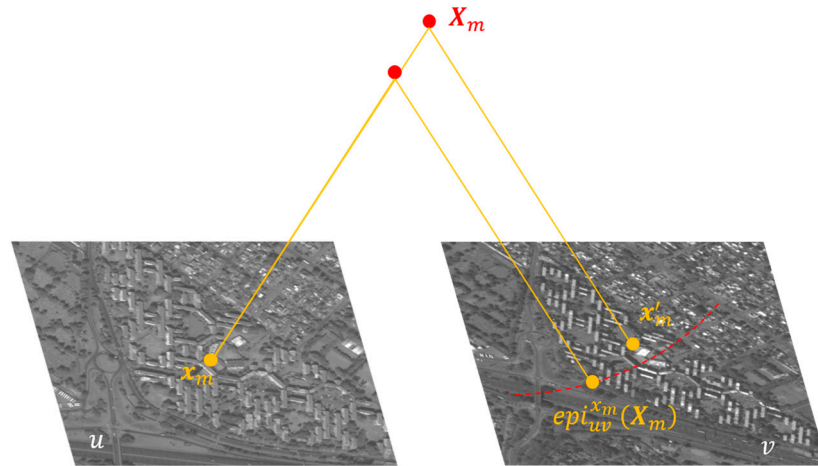


Figure 4. The principle of the epipolar constraint error.

Suppose that there is a satellite stereo pair consisting of an image u and an image v , and there are M pairs of matching points $(x, x')_{m=1,2,\dots,M}$ on the images. The epipolar constraint error can be expressed as Equation (11).

$$E_{uv} = \frac{1}{M} \sum_{m=1}^M (x'_m, epi_{uv}^{x_m}(X_m))^2, \tag{11}$$

where $x'_m = (x'_m, y'_m)$ on satellite image v is the m -th matching point of $x_m = (x_m, y_m)$ on satellite image u , and the corresponding 3D point is expressed as $X_m = (X_m, Y_m, Z_m)$. The meaning of $epi_{uv}^{x_m}(\cdot)$ is the epipolar curve of x_m , whose form is as below.

$$epi_{uv}^{x_m}(X_m) = P_v(P_u^{-1}(X_m)), \tag{12}$$

the $P_v(\cdot)$ represents the satellite camera projection of the image v and $P_u^{-1}(\cdot)$ denotes the localization model of the image u . The purpose of the stereo rectification is to minimize the epipolar constraint error by image transformation (R, t) . Therefore, Equation (11) can be reorganized as

$$E_{uv} = \frac{1}{M} \sum_{m=1}^M (Rx'_m + t, epi_{uv}^{x_m}(X_m))^2, \tag{13}$$

where R and t denote the rotation and the translation of the point x'_m , respectively. Actually, the proposed method adopts a rigid 2D transformation on the satellite image, which is calculated according to the matching point pairs.

3.3. Stereo Matching

The satellite stereo pair will be matched after rectification for obtaining dense disparity map. The traditional SGM algorithm adopts mutual information to calculate the matching cost, but the mutual information needs to obtain the information entropy of the satellite image, which leads to the low computational efficiency of the algorithm. The information entropy of image I can be calculated by Equation (14).

$$H_I = - \int_0^1 p_I(i) \log p_I(i) di, \tag{14}$$

where $p_I(i)$ denotes the probability distribution of the luminance of pixel i on image I . The proposed method uses Census cost [27] to calculate the matching cost, and Census cost utilizes the luminance difference in the local pixel neighborhood to convert the luminance into bits, which can significantly improve the efficiency of the matching algorithm.

After matching cost calculation, the SGM algorithm optimizes the initial cost through the matching cost aggregation, so that the optimized matching cost will be robust to the local noise. This process is essentially a dynamic programming problem. Different from the traditional SGM algorithm, when aggregating the matching costs, we replace the disparity information in one direction with the mean of it in two vertical directions. Actually, we add a direction that rotates the original direction by 90° counterclockwise for aggregation. The final step in stereo matching is the disparity map generation. The proposed method selects the disparity value corresponding to the minimum aggregating matching costs for each pixel as the final disparity; then, we use a left–right check [28] to refine the final disparity.

3.4. Altitude Estimation

Altitude estimation is the final step of the proposed method, which requires triangulation in combination with the dense disparity map generated in Section 3.3 and the corresponding RPC file of the satellite image. We bring the matching point pairs into the epipolar curve equation and iterate over it to obtain the final altitude value.

4. Experimental Results and Discussion

The proposed method is tested on a benchmark dataset about satellite images, the corresponding experimental results are analyzed thoroughly, and the proposed method is verified to be feasible and effective.

4.1. Dataset and Metrics

The public benchmark dataset we used in this section is IARPA Multi-View Stereo 3D Mapping Challenge [29], which contains 50 panchromatic satellite images captured by WorldView-3 satellite. The acquisition time of these satellite images is from November 2014 to January 2016. The nadir resolution of the satellite images is 30 cm, and these images cover an area of 100 km² near San Fernando, Argentina. The IARPA MVS3DM dataset also provides airborne lidar data with 20 cm nadir resolution that can be used as ground truth DSM.

The metrics of the experimental results evaluation in this paper is followed the previous quantitative evaluation [1,2,7,18], which contains completeness (CP), root-mean-square error (RMSE) and median error (ME). All these three metrics require aligning the reconstructed DSM with the ground truth DSM and comparing them with pixel-wise. Among the three metrics, CP stands for the percentage of altitude errors less than 1m in a reconstructed 3D point cloud, RMSE represents the root-mean-square error of the altitude, and ME is defined as the median altitude error.

4.2. Performance Analysis of the Proposed Method

We select eight sites from IARPA MVS3DM dataset. There are 400 images from the eight sites, and when training the generator network, we selected 289 images as the training set, 29 images as the validation set, and 82 images as the test set. Since the dataset contains multi-date satellite images, we use an experience-based satellite stereo pair selection strategy [4]. The input satellite images are shown in Figure 5. Site 1 is a mixed area of low-height buildings, medium-height buildings and roads, whose altitude range is from 18 m to 35 m. Site 2, Site 5, and Site 6 are low-height building areas with an altitude range of 18 m–35 m. Site 3 is a mixed area of medium-height buildings and low-height buildings with an altitude range of 18 m–42 m, whose altitude of medium-height buildings is a little bit higher than that of Site 1. Site 4 and Site 7 are areas that have high buildings with an altitude more than 80 m. Site 8 is a park area with an altitude range of 15 m–55 m, and there are some water surfaces in this area. These 8 sites contain a variety of heights and different types of buildings, which can fully verify the performance of the proposed method.

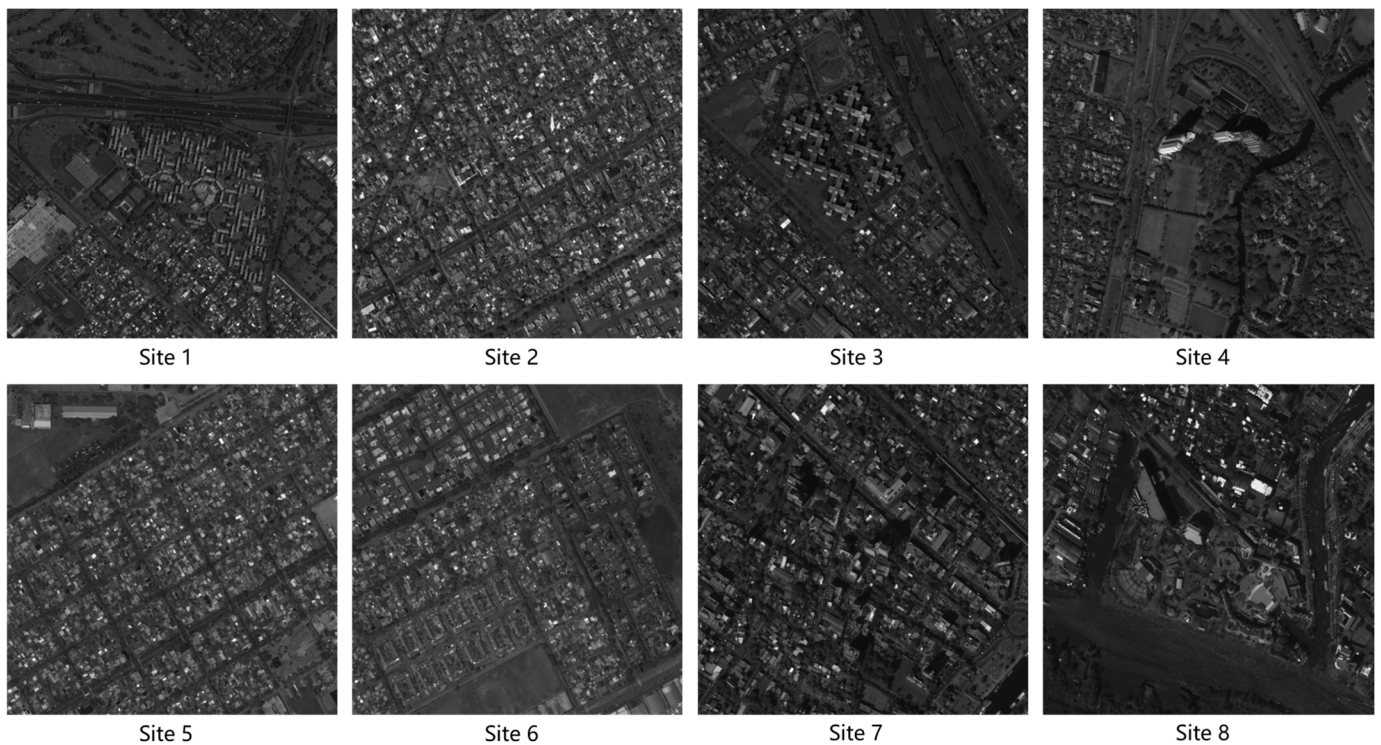


Figure 5. The input satellite images of the 8 selected sites from the IARPA MVS3DM dataset.

We compare our method with the S2P method [7] and the BA method [18]. The S2P method is one of the most accurate 3D reconstruction methods at present, which achieves high-precision and rapid reconstruction through ingenious geometric rectification of satellite images. The BA method is a 3D reconstruction method of satellite images that optimizes the RPC parameters by the bundle adjustment algorithm, and this method can generate good reconstruction results when there are multi-view satellite images. The ground truth DSMs and the reconstructed DSMs are shown in Figure 6. For comparison and visualization, we represent DSMs as RGB images, using red for high-altitude areas and blue for low-altitude areas. The elevation range for each site is indicated on the rightmost part of Figure 6. The details of the reconstructed results of Site 1 are shown in Figure 7.

For Site 1, it can be seen from Figures 6 and 7 that the S2P method generates poor reconstruction results in the building area of the upper-left corner, and the BA method has a low reconstruction elevation in the 35 m building area (this area is orange instead of red), and the method proposed in this paper has a better reconstruction result in this area. As for Site 2 and Site 3, the reconstructed results of the S2P method in the lower-left and middle areas are still obviously distorted, while the reconstruction results of the BA method and the proposed method are blurred at the edge of the high-rise buildings and the left edge of the DSM. As can be seen from the fourth row of Figure 6, the reconstructed DSMs of Site 4 are poor for all three methods, due to the fact that Site 4 contains too many vegetated areas. For Site 5 and Site 6, the visual quality of the reconstruction results of the S2P method and the proposed method is slightly better than that of the BA method, which may be due to the fact that the BA method does not obtain the optimal value when bundle adjustment optimizes the RPC parameters. The results of Site 7 and Site 8 reconstructed by the three methods has poor visual quality due to the fact that these two sites contain multiple forms of buildings and the height difference between buildings is pretty large. Due to the large architectural difference, the 3D reconstruction methods will reduce the number of matching points for stereo matching, resulting in inaccurate altitude estimation.

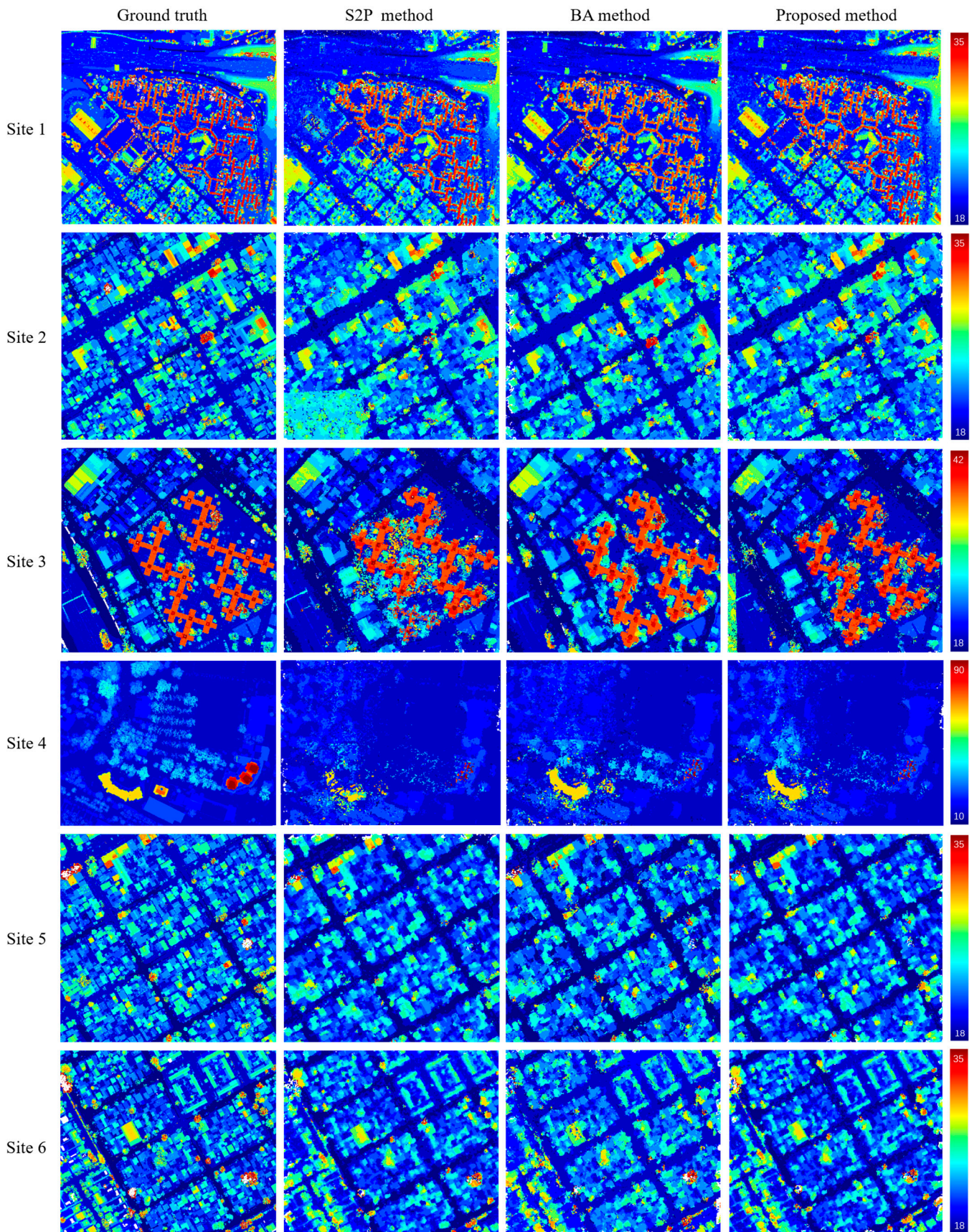


Figure 6. Cont.

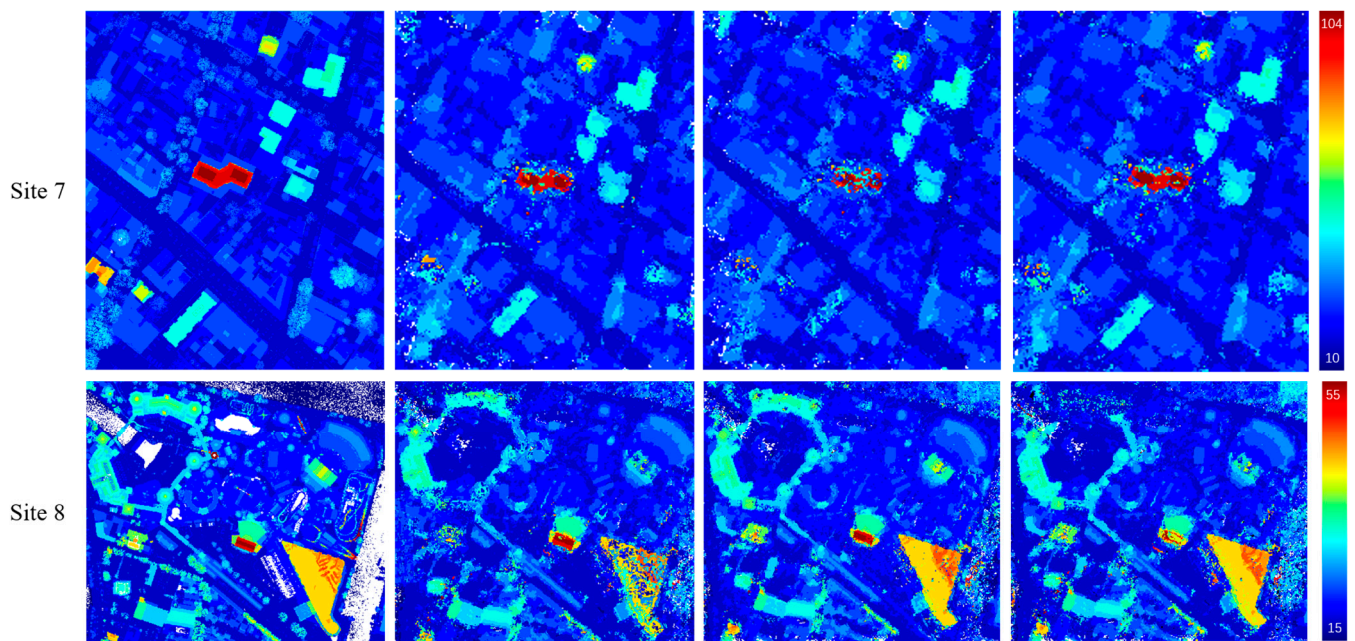


Figure 6. The ground truth DSMs and the reconstructed DSMs of the S2P method [7], the BA method [18], and the proposed method.

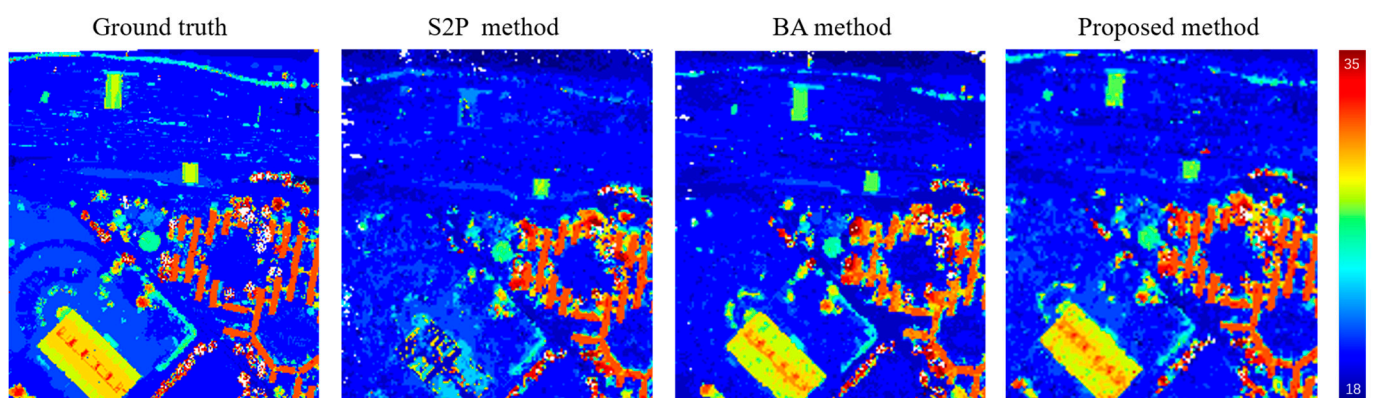


Figure 7. The details of the reconstructed results of Site 1.

We performed quality evaluation experiments on the reconstructed DSMs shown in Figure 6, the experimental results of which are listed in Table 1. The metrics marked in bold and green are those with better quality evaluation.

According to the introduction of Section 4.1, for the reconstruction results, a higher CP means more good points in the reconstructed 3D point cloud. Here we define the reconstructed 3D points with an altitude error of less than 1m as good points (since the nadir resolution of the satellite image is 0.3 m, it is reasonable to set the error threshold at the 1m). The RMSE and ME metrics both represent the reconstruction accuracy of the 3D point cloud, and the RMSE finds the root-mean-square error of the altitude of all the reconstruction 3D points, including some 3D points with very large errors, but this kind of 3D points may be the reconstruction points that we are not concerned about, such as the altitude of trees. Besides RMSE, this paper also uses ME to represent the median error of the reconstructed 3D point cloud altitude, and ME can be understood as the altitude error of the first 50% of the 3D points with the best reconstruction altitude accuracy. Obviously, we hope the RMSE and ME of the reconstructed 3D point cloud are as low as possible.

Table 1. The quality evaluation results of the reconstructed results.

Site	Method	CP (%)	RMSE (m)	ME (m)
1	S2P method [7]	73.98	2.60	0.39
	BA method [18]	72.60	2.68	0.65
	Proposed method	74.69	2.47	0.42
2	S2P method [7]	60.77	2.74	0.57
	BA method [18]	60.66	2.64	0.55
	Proposed method	65.08	2.22	0.50
3	S2P method [7]	67.21	8.87	0.35
	BA method [18]	66.91	3.99	0.47
	Proposed method	68.88	5.58	0.34
4	S2P method [7]	50.21	11.14	0.98
	BA method [18]	42.04	9.34	1.45
	Proposed method	51.50	10.74	0.89
5	S2P method [7]	71.19	1.92	0.52
	BA method [18]	68.61	2.31	0.46
	Proposed method	71.45	1.88	0.51
6	S2P method [7]	68.57	2.17	0.53
	BA method [18]	53.02	3.74	0.90
	Proposed method	68.74	2.17	0.54
7	S2P method [7]	59.20	7.56	0.74
	BA method [18]	54.71	6.37	0.82
	Proposed method	58.67	4.97	0.74
8	S2P method [7]	63.12	5.03	0.47
	BA method [18]	62.66	3.80	0.66
	Proposed method	63.34	4.85	0.48

As shown in Table 1, the proposed method has the highest CP in the reconstruction results of most sites, and the CP of the proposed method is also close to the best CP for Site 7. In addition, except for Site 4, Site 7, and Site 8, the CP of the proposed method is more than 65%, indicating that the 3D point cloud reconstructed by our method has good completeness. The slightly worse CP of the reconstructed results for Site 4, Site 7, and Site 8 indicates bad 3D reconstruction from satellite images, which is also verified in Figure 6. In terms of accuracy, the proposed method can achieve an RMSE of less than 5.0 m and an ME of less than 0.6 m for the reconstruction results of most sites (more than five sites), which means that the proposed method can generate 3D reconstructed point clouds with high accuracy.

5. Conclusions

In this paper, we propose an improved 3D reconstruction method for satellite images based on GAN image enhancement, which uses the perceptual loss function training network to enhance the satellite image. The loss function of the proposed method is a good combination of MSE loss, VGG loss, adversarial loss, and regularization loss. This method improves the quality of satellite images and then improves the accuracy of the reconstruction results of the satellite images. The proposed method in this paper not only obtains 3D point clouds with good visual quality but also has a good objective quality evaluation of the reconstruction results. Comparative experiments with other methods prove the effectiveness and feasibility of the proposed method.

Author Contributions: Conceptualization, H.L. and J.Y.; methodology, H.L. and L.J.; software, H.L.; validation, H.L.; formal analysis, H.L., L.J. and J.Y.; investigation, H.L. and L.J.; resources, H.L.; data curation, H.L.; writing—original draft preparation, H.L.; writing—review and editing, H.L., L.J. and J.Y.; visualization, H.L. and L.J.; supervision, J.Y.; project administration, J.Y.; funding acquisition, H.L. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by Major Program of National Natural Science Foundation of China NSFC (Nos.12292980, 12292984); by National Key R&D Program of China (Nos. 2023YFA1009000, 2023YFA1009004, 2020YFA0712203, 2020YFA0712201); by Key Projects of National Natural Science Foundation of China NSFC (no. 12031016); by Beijing Natural Science Foundation (no. BNSF-Z210003); by the Department of Science, Technology and Information of the Ministry of Education (no. 8091B042240); and by the Fundamental Research Funds for the Central Universities (2412022QD024).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Zhang, K.; Snavely, N.; Sun, J. Leveraging vision reconstruction pipelines for satellite imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.
2. Qin, R.; Song, S.; Ling, X.; Elhashash, M. 3D reconstruction through fusion of cross-view images. In Proceedings of the Recent Advances in Image Restoration with Applications to Real World Problems, London, UK, 4 November 2022.
3. Wang, P.; Shi, L.; Chen, B.; Hu, Z.; Qiao, J.; Dong, Q. Pursuing 3-D scene structures with optical satellite images from affine reconstruction to Euclidean reconstruction. *IEEE Trans. Geosci. Remote* **2022**, *60*, 1–14. [\[CrossRef\]](#)
4. Facciolo, G.; De Franchis, C.; Meinhardt-Llopis, E. Automatic 3D reconstruction from multi-date satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017.
5. Wang, M.; Hu, F.; Li, J. Epipolar arrangement of satellite imagery by projection trajectory simplification. *Photogramm. Rec.* **2010**, *25*, 422–436. [\[CrossRef\]](#)
6. Oh, J.; Lee, W.H.; Toth, C.K.; Grejner-Brzezinska, D.A.; Lee, C. A piecewise approach to epipolar resampling of pushbroom satellite images based on RPC. *Photogramm. Eng. Rem. Sens.* **2010**, *76*, 1353–1363. [\[CrossRef\]](#)
7. De Franchis, C.; Meinhardt-Llopis, E.; Michel, J.; Morel, J.-M.; Facciolo, G. Automatic sensor orientation refinement of Pléiades stereo images. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014.
8. Ghuffar, S. Satellite stereo based digital surface model generation using semi global matching in object and image space. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2016**, *3*, 63–68. [\[CrossRef\]](#)
9. Hirschmuller, H. Stereo processing by semiglobal matching and mutual information. *IEEE Trans. Pattern Anal.* **2007**, *30*, 328–341. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Tatar, N.; Saadatseresht, M.; Arefi, H.; Hadavand, A. Quasi-epipolar resampling of high resolution satellite stereo imagery for semi global matching. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2015**, *40*, 707–712. [\[CrossRef\]](#)
11. Ye, L.; Peng, M.; Di, K.; Liu, B.; Wang, Y. Lunar terrain reconstruction from multi-view Lroc Nac images based on semi-global matching in object space. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2020**, *43*, 1177–1183. [\[CrossRef\]](#)
12. Facciolo, G.; De Franchis, C.; Meinhardt, E. MGM: A significantly more global matching for stereovision. In Proceedings of the British Machine Vision Conference 2015, Swansea, UK, 7–10 September 2015.
13. Qayyum, A.; Malik, A.S.; Saad, M.N.B.M.; Abdullah, F.; Iqbal, M. Disparity map estimation based on optimization algorithms using satellite stereo imagery. In Proceedings of the IEEE International Conference on Signal and Image Processing Applications, Kuala Lumpur, Malaysia, 19–21 October 2015.
14. Bleyer, M.; Rhemann, C.; Rother, C. Patchmatch stereo-stereo matching with slanted support windows. In Proceedings of the British Machine Vision Conference 2011, Dundee, UK, 29 August–2 September 2011.
15. Zbontar, J.; LeCun, 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.
16. Poggi, M.; Mattoccia, S. Learning a general-purpose confidence measure based on O(1) features and a smarter aggregation strategy for semi global matching. In Proceedings of the Fourth International Conference on 3D Vision, Stanford, CA, USA, 25–28 October 2016.
17. Bosch, M.; Kurtz, Z.; Hagstrom, S.; Brown, M. A multiple view stereo benchmark for satellite imagery. In Proceedings of the IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, 18–20 October 2016.
18. Mari, R.; De Franchis, C.; Meinhardt-Llopis, E.; Anger, J.; Facciolo, G. A generic bundle adjustment methodology for indirect RPC model refinement of satellite imagery. *Image Process. Line* **2021**, *11*, 344–373. [\[CrossRef\]](#)
19. Qin, R. A critical analysis of satellite stereo pairs for digital surface model generation and a matching quality prediction model. *ISPRS J. Photogramm.* **2019**, *154*, 139–150. [\[CrossRef\]](#)
20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, PQ, Canada, 8–13 December 2014.

21. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
22. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. WESPE: Weakly supervised photo enhancer for digital cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
23. Sun, X.; Li, M.; He, T.; Fan, L. Enhance images as you like with unpaired learning. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Online, 19–26 August 2021.
24. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [[CrossRef](#)] [[PubMed](#)]
25. Jiang, L.; Zhang, C.; Huang, M.; Liu, C.; Shi, J.; Loy, C.C. TSIT: A simple and versatile framework for image-to-image translation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020.
26. Ni, Z.; Yang, W.; Wang, S.; Ma, L.; Kwong, S. Towards unsupervised deep image enhancement with generative adversarial network. *IEEE Trans. Image Process.* **2020**, *29*, 9140–9151. [[CrossRef](#)] [[PubMed](#)]
27. Fife, W.S.; Archibald, J.K. Improved census transforms for resource-optimized stereo vision. *IEEE Trans. Circ. Syst. Vid.* **2012**, *23*, 60–73. [[CrossRef](#)]
28. Jie, Z.; Wang, P.; Ling, Y.; Zhao, B.; Wei, Y.; Feng, J.; Liu, W. Left-right comparative recurrent model for stereo matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
29. Bosch, M.; Leichtman, A.; Chilcott, D.; Goldberg, H.; Brown, M. Metric evaluation pipeline for 3D modeling of urban scenes. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, *42*, 239–246. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.