TWO-VIEW STRUCTURE FROM MOTION FOR BUILDING MODEL RECONSTRUCTION FROM VHR SATELLITE IMAGERY

Elisabeth Johanna Dippold [1] and Fuan Tsai * [1, 2]
[1] Dept. of Civil Engineering, National Central University, 300, Zhongda Rd., Zhongli, Taoyuan 32001, Taiwan
Email: elisabeth.dippold@gmail.com

[2] Center for Space and Remote Sensing Research, National Central University, 300, Zhongda Rd., Zhongli, Taoyuan, 32001, Taiwan
Email: ftsai@csrsr.ncu.edu.tw

KEY WORDS: Structure from Motion, Satellite Imagery, Moving Window, Sparse Point Cloud

ABSTRACT: 3D information can be of a great use for different applications. The generation of point cloud (PCL) for building reconstruction from very high resolution (VHR) satellite imagery can be achieved in a variety of ways. Structure from Motion (SfM) generates 3D PCL from a set of 2D images based on the sensor information, the scene captured and the motion of the sensor. This study utilizes multiple space-borne images captured by Pelaides-1B VHR satellite as tri-stereo pairs. The key of this study is the generation of sparse PCL from VHR satellite imagery with two-view SfM by utilizing moving window operation. The generation of PCL from VHR satellite imagery can be accomplished in four major steps. Firstly, feature detection by utilizing multiple feature detector operators (MFDO) withing a moving window operation. Secondly, masking out features with Homography matrix and removing detected outliers. Thirdly, matching the remaining features. Finally, pose estimation and point triangulation to generate sparse PCL. This study emphasizes the importance of detecting robust features by applying MFDO as well as the number and quality of features. The applied two-view SfM algorithm is resource aware and an efficient solution for 3D reconstruction. Further, the incorporation of a moving window operation increases the number of effectives features whereas the masking enhances the quality of the SfM operation. The resulting sparse point clouds show great alignment and the target building is recognizable. The efficiency and resource awareness of our study lead to promising results for subsequent sophisticated applications.

1. INTRODUCTION AND RELATED WORK

1.1 Introduction

The 3D modelling from VHR satellite imagery is a hot topic in remote sensing and related research areas (Xiao et al., 2023). The coverage of a large area, relatable images like stereo and efficient processing chain are beneficial key elements. Structure from Motion (SfM) is an algorithm that generates a 3D model from a set of 2D images. One crucial step in SfM based reconstruction, among others, is the initial match. Therefore, this study focuses on utilizing two-view SfM algorithm that requires only two images (Dippold and Tsai, 2022; Wang et al., 2021; Zhou and Zhang, 2015) to generate 3D point cloud building model from VHR satellite images.

The generation of 3D point cloud (PCL) from a set of 2D images with SfM can include multiple feature detector operators (MFDO). This approach increases the number of detected features significantly. For example, Dippold and Tsai (2023) utilized different MFDO, including speeded-up robust features (SURF), features from accelerated segment test (FAST) with binary robust independent elementary features (BIREF) and oriented FAST and rotated BRIEF (ORB). The application of a moving widow operation further increases the number of detected features for matching reconstruction in the SfM operation. However, the number of outliers will increase as well. Masking to remove outliers can preserve the quality of matches. The Moving Window (MW) is a plotting technique with a consisted series of sub data as result. This can be beneficial in detecting vegetation and describing patterns as well as for boundary analysis (Erdős et al., 2014). In contrast, the moving window technique can be used for efficient calculation by utilizing multi-core architectures. Consequently, the combination of multicore application and geospatial index calculation like aspect or curvature, leads to a resource aware DEM generation (Bylina et al., 2023).

Accordingly, this study focuses on the five major steps to generate PCL from VHR satellite imagery for building reconstruction using two-view SfM algorithms. The steps include feature detection and matching with moving window and MFDO, masking outliers, pose estimation and triangulation to generate 3D point cloud.

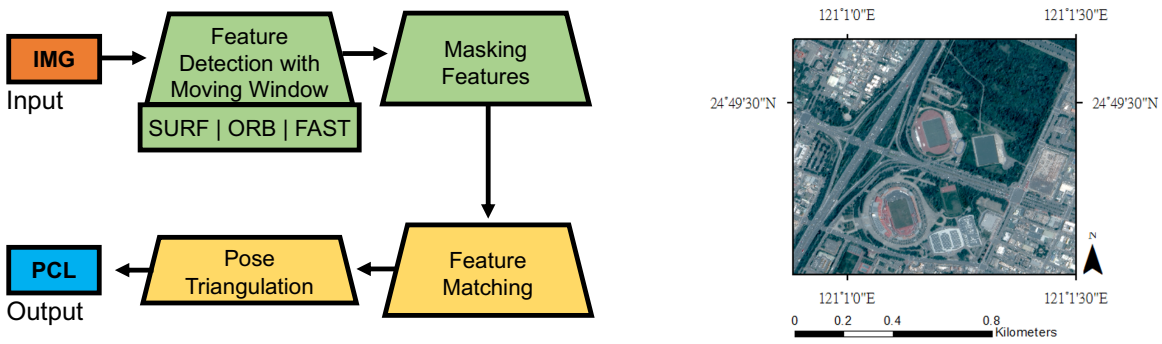
1.2 Related Work

The establishment of a relation based on pattern, classes and properties are often benefiting from MW operation. The utilization of MW with K-means was successfully carried out to relate an urban form (landscape) to land surface temperature (LST) (Zawadzka et al., 2021). In addition, examination with respect to canopy height moving window based statistics was also proposed to detect and describe trees (Blaschke et al., 2004). On the other hand, 3D reconstructions are the generation of a three-dimensional model from one or more 2D images with auxiliary data. However, 2D images are projections of the 3D world onto a 2D plane so that the third dimension, depth, is lost in the process. The reconstruction process faces different challenges in terms of calibration, matching, reconstruction and the density of the 3D model.

The 3D model can be obtained by disparity, parallax, blur, silhouette and, as already introduced, structure from motion. In addition, models can also be generated by artificial intelligence (AI) (Moncef Aharchi and Kbir, 2020). Furthermore, deep learning (DL) as well as other machine learning (ML) techniques which are often combined with traditional methods, in order to achieve 3D reconstruction. As a result, neural networks (NN), support vector machines (SVM) and generative adversarial networks (GAN) are wildly used for 3D building reconstruction (Buyukdemircioglu et al., 2022). In contrast, software and applications as listed in a previous review (Djuric et al., 2021) including the well-known open-source VisualSFM (Wu et al., 2011) are utilized for 3D reconstructions. However, most the existing algorithms were developed for close-range or airborne data. Other applications which can process satellite imagery within an open-source photogrammetry software are MicMac (Rupnik et al., 2017), S2P (de Franchis et al., 2014) and CARS (Youssefi et al., 2020). These applications need further information like accurate rational polynomial coefficients (RPC) which provide information about the geometry of image and ground. This study investigates on the issues in utilizing the two-view SfM algorithm to reconstruct 3D building models directly from VHR satellite images and without RPC or other orbit information.

2. MATERIALS AND AREA OF INTEREST

A tri-stereo pair of Pleiades-1B VHR satellite images, including multispectral and panchromatic bands are the primary materials used in his study. The tri-stereo images are acquired as nadir, forward and backward views and summarized in Table 1 and the area of interest (AOI) is a small urban area as displayed in Figure 1. The target area consists mainly of a stadium with park areas in the north, residential in the east and south and high-way in the west. The panchromatic band of the satellite images can be used to pansharped the multispectral (MS) in order to generated a higher resolution image with the spectral information preserved. This study used coupled nonnegative matrix factorization (CNMF) to pansharped MS with PAN (Yokoya et al., 2012).



(a) (b)

Figure 1. Workflow and Area of Interest (AOI)

Table 1. Dataset of VHR satellite imagery.

Name	Direction	Time	IA
Img 1 (70)	Nadir	2019-08-27T 09:18:44.807	12.57°
Img 2 (74)	Forward	2019-08-27T 08:43:17.817	17.71°
Img 3 (93)	Backward	2019-08-27T 08:47:06.900	16.62°

3. METHODOLOGY

The generation of PCL with a two-view SfM algorithm can be divided into five steps (Figure 1 (a)). Firstly, read in the images with auxiliary data of focal length and principial point. Secondly, perform feature detection utilizing the entire image with MFDO – SURF, FAST and ORB. Then use moving window operation to increase the number of features detected further. Thirdly, mask outliers out with Homography matrix to preserve quality of quantity. Then, match features and stack all points up. The final step is pose estimation and triangulation for PCL generation with the remaining points.

The two-view SfM algorithm starts with feature detection of two images, as introduced and the PAN and RGB are tested. In general, there are two ways to detect features within a moving window operation, using the whole image at once or partially. A moving window operation slide over a predefined area with or without defined overlap. This study utilizes moving window in two different ways, the complete image (Figure 2 (a) and (b)) and strides (Figure 2 (c) and (d)). In addition, the subdivision of the whole image (756 x 756) into five strides (128 x 756) for further feature detection. In both cases (Figure 2 (a-b) and (c-d)), the moving window size is 128x128 and moves two pixels down. One downward movement of the moving window is shown in Figure 2 (e), to visualize and emphasize the possible processing volume. The positive quantitative aspect is quality-wise enhanced by masking outliers out. After outlier removal, adding the detected points of both operations together for each feature detector respectively.

The pose estimation with the remaining points to estimate Fundamental matrix and Essential matrix to recover the pose is the next process. Finally, triangulation is performed based on previous processes to generate 3D point cloud. The postprocessing includes the removal of random noise, error analysis and a noise free result analysis.

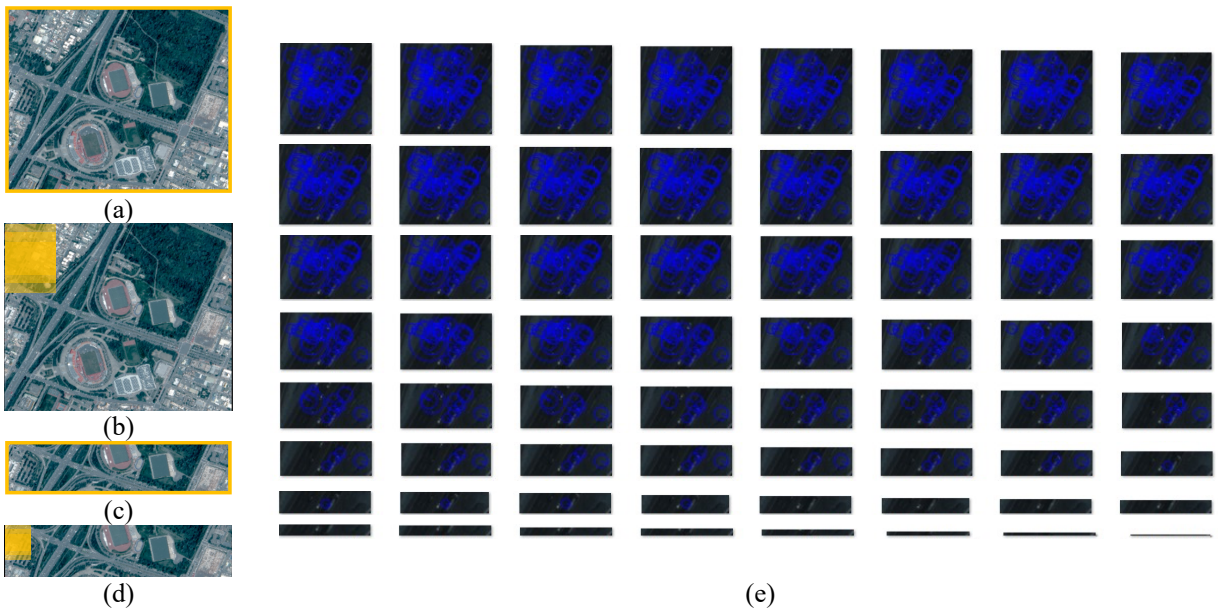


Figure 2. Moving Window Operation (a) to (d) and Example Window Movement (e)

4. RESULTS AND DISCUSSIONS

The result evaluation based on a density analysis and visible inspection was conducted on the generated 3D point cloud model. The noise removal is performed manually in CloudCompare. Whereas, the visible inspection to visualize the target and the achieved level of detail as grid-based plot was carried out in MATLAB. The ground truth is generated with MicMac from VHR tri-stereo pairs with RPC of the entire scene. The evaluation will focus on the main target – the stadium and its surrounding in the area of interest.

4.1 Noise and Matching Errors

In general, the process faces different challenges in terms of calibration and matching besides the density of the 3D model. Figure 3 displays different error sources summarized in which this study faced in order to generate sparse point cloud from VHR satellite imagery using a two-view SfM algorithm.

The curvy stream of points indicates warping in combination with distortion shown in the first case from PCL (a) to (c), which is often caused by three major issues. Firstly, the warping indicates that there are too few points especially in the border area of the image. The distortion, the curved shaping of the model implies that the used parameters for calibration, focal length and principial point, are inaccurate. In contrast, the mismatch of two images can lead to a point cloud as shown in Figure 3 (f) and (g). However, this curial error needs to be analyzed further. This PCL was generated with two tiles of 1000x1000 pansharpened images and a moving window of 500x500 (Figure 3 (h)) with a movement of 250 pixels (9 windows). In addition, to adjust the moving window, the finetune of the tiles size and the parameters of focal length and principial point need further investigation. Finally, the amount of noise as shown in the generated PCL (d) and (e) indicate the need of a better strategy to improve the quality. Therefore, masking the outliers out with Homography matrix seems to be a commonly used approach.

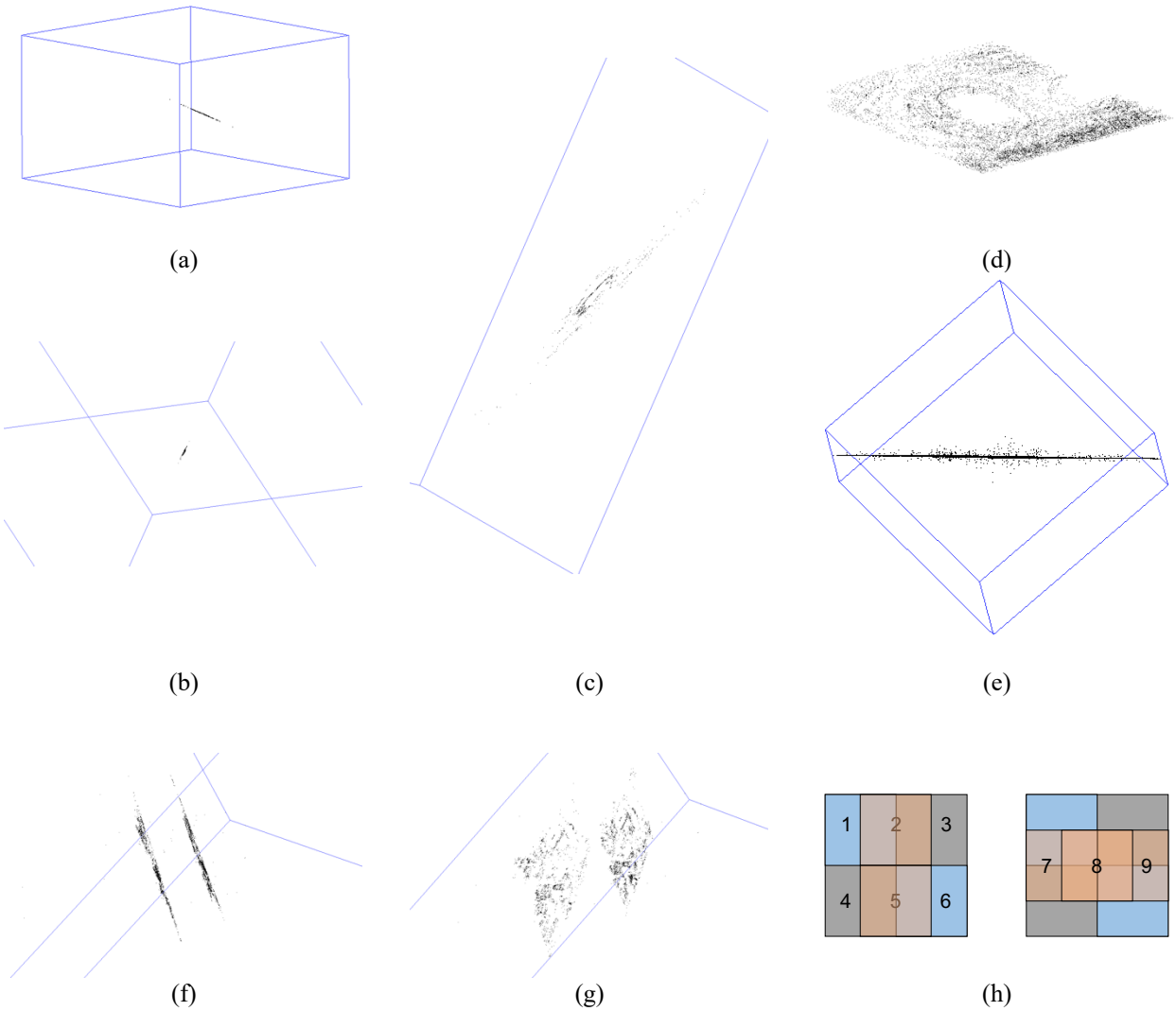


Figure 3. Error Strategy

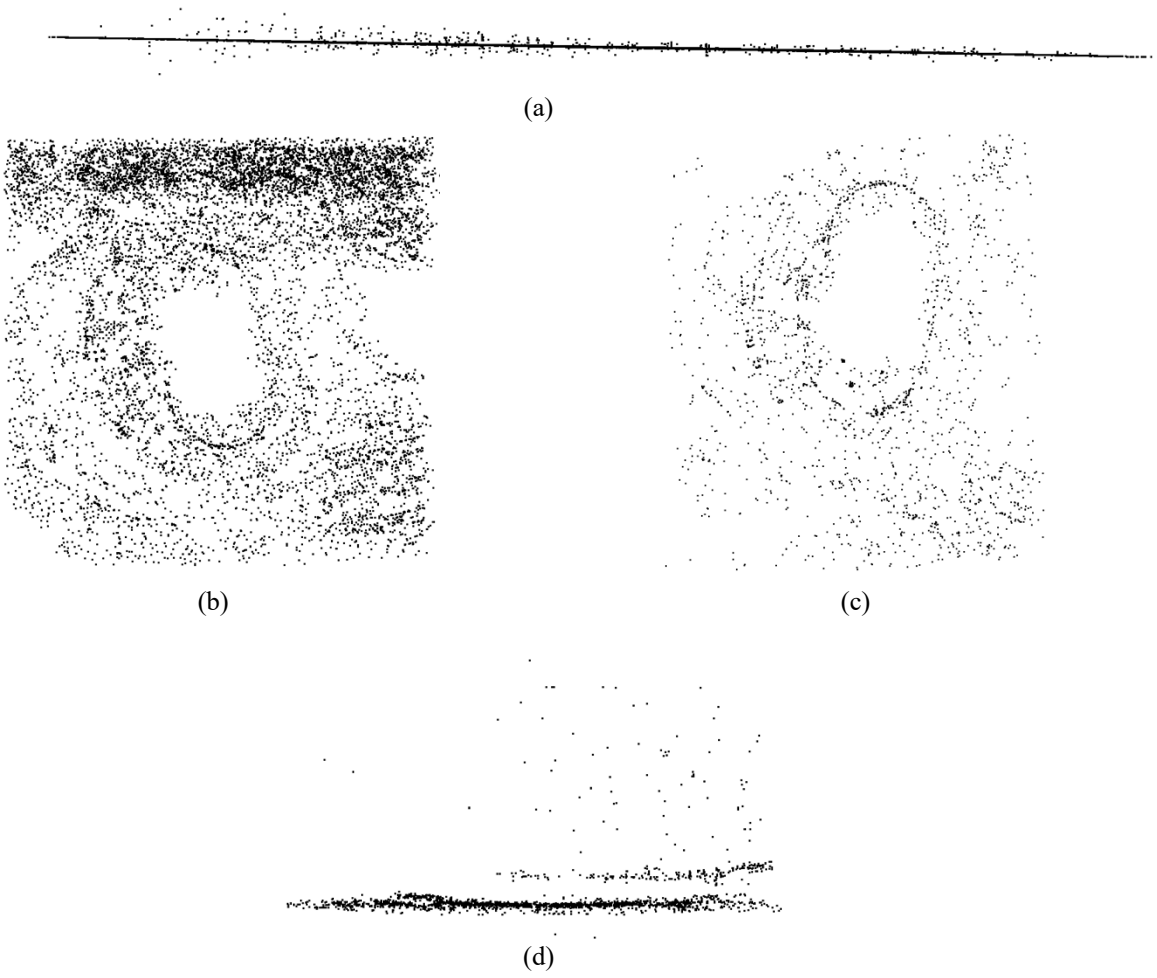
4.2 Noise and first Impression

The comparison on the quality of the point cloud generated from RGB in comparison to PAN is summarized in Table 2 and Figure 4. The panchromatic image is used to pansharpen the multispectral RGB, as described.

The result of PAN with an initial PCL of 282,991 points shows a clear tile boarder (Figure 4 (b)). The stadium in the model center is recognizable. Besides that, the solar panels (bottom left corner) are partially captured, whereas the street is completely blurred. The amount of noise (Figure 4 (a)) can be quantified with around 0.18% or 497 points, so that the clean PAN PCL consist of 282,494 points. The RGB with an initial PCL of 105,522 points (Figure 4 (c)) is much sparser than the PAN PCL by 62.71% or 177,469 points. The stadium in this model is recognizable. Besides that, the PCL is too sparse to identify more objects. The amount of noise (Figure 4 (d)) is around 42.89% or 45,256 points, so that the final RGB PCL consists of 60,266 points.

Table 2. Density and Noise of PAN and RGB based PCLs.

	Inital	Noise %	Noise absolut	Clean
PAN	282,991	0.18	497	282,494
RGB	105,522	42.89	45,256	60,266
Difference %	62.71	-42.71	-	21.33
Difference abs.	177,469	-	-44,759	222,228



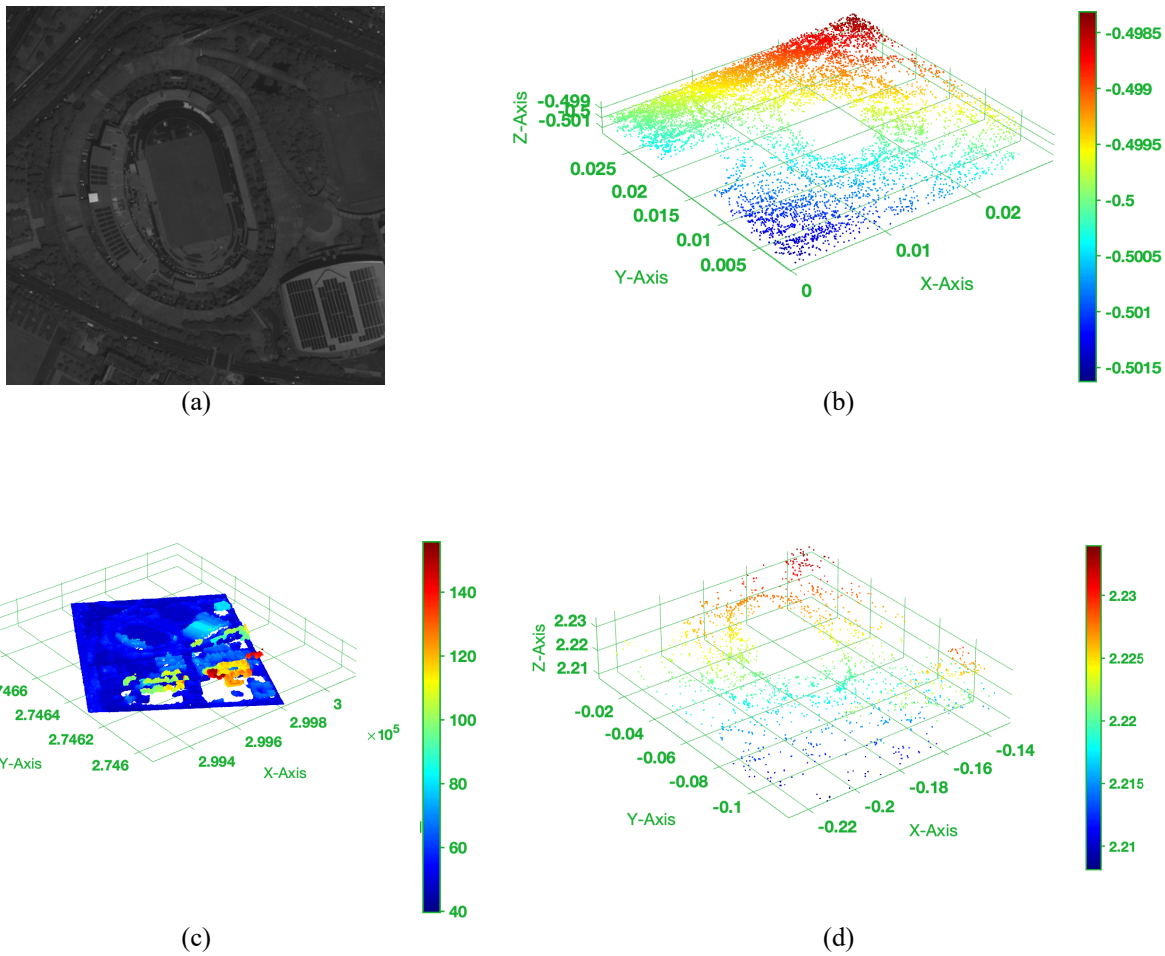
(a) (b) (c) (d)

Figure 4. Manual Postprocessing of PAN (a) with (b) and RGB (c) with (d).

4.3 Visible Inspection and Analysis

The visual inspection focus on shape, lack of data and reasonability. Figure 5 displays one of the input images as well as the ground truth (GT) and two PCLs generated by the method proposed in this study. The PAN image (Figure 5 (a)) shows the input area processed including the target. In contrast, the GT was generated by MicMac using tri-stereo pairs with RPC to reconstructed 3D Model of the entire overlapped scene. The target shown in Figure 5 (c) is a part of the digital surface model (DSM) of the study site generated using MicMac. The PCL shown in Figure 5 (b) is the result from RGB whereas the PAN based result is shown in Figure 5 (d).

In general, the density and scalability differences of the GT empathize the main differences of GT with the other two PCLs. The GT as well as the PCL generated by this study exhibit difficulties in effectively reconstruction of the secondary buildings surrounding the stadium. Overall, the combination with the nadir image seems to be better than forward and backward. Interestingly, for PAN and RGB are the combinations with the nadir image the best. So that, presented are PAN with nadir and forward and RGB nadir and backward as the best results.



(a) (b)

(c) (d)

Figure 5. Target PAN image (a) and GT PCL generated by MicMac (c) in comparison to our Studies PCLs PAN (b) and RGB (d).

5. CONCLUSIONS

This study proposes and demonstrates a method for reconstructing 3D point cloud models of buildings from VHR satellite images. In addition, the results were evaluated with ground truth data and the comparison between the PCLs from the panchromatic and the result from pansharpened RGB were also analyzed. Based on the results demonstrated in this paper, it is noticed that in general, a higher overlap seems to produce better results. This is conforming to a previous study for UAV-SfM (Nesbit and Hugenholtz, 2019). In addition, the nadir in combination with forward or backward view lead to better results for a two-view SfM on VHR satellite images tested in this study. The proposed method can help produce 3D point cloud models of buildings and other objects from very high-resolution satellite images and require few additional information. The results can further be processed to object-based models and used for subsequent applications.

The genuine misalignment of the input images needs to be addressed in both ways (Aslahishahri et al., 2021). Firstly, the alignment of the two images tiles should be minimized, so that a moving window can be narrower. Secondly, the band-to-band registration of the RGB can increase error and noise of the generated 3D model. Therefore, this issue must also be corrected during the preprocess. Future work will include image registration, adjustment of the moving window operation and studying image enhancement in order to improve the performance of the proposed method. Better image registration will increase correctness the overlapped area and indirectly improve the moving window operation. In addition, with a smaller moving window achieving a higher level of details is more likely. Furthermore, image enhancement strategies like histogram equalization can boost the level of detail.

Acknowledgements

This study was partially supported by the Ministry of Interior (MOI) of Taiwan (ROC) under project no. 112PL024A.

References

Aslahishahri, M., Stanley, K.G., Duddu, H., Shirtliffe, S., Vail, S., Bett, K., Pozniak, C., Stavness, I., 2021. From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops, 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 1312-1322.

Blaschke, T., Tiede, D., Heurich, M., 2004. 3D Landscape Metrics to Modelling Forest Structure and Diversity based on Laser Scanning Data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVI, 129-132.

Buyukdemircioglu, M., Kocaman, S., Kada, M., 2022. DEEP LEARNING FOR 3D BUILDING RECONSTRUCTION: A REVIEW. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2-2022, 359-366.

Bylina, B., Bylina, J., Chabudziński, Ł., Karpowicz, K., Klisowski, M., Oleszczuk, P., Potiopa, J., Stpiczynski, P., 2023. Fast slope algorithm with the use of vectorization and parallelization for multicore architectures. GeoInformatica, 1-31.

de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images.

Dippold, E., Tsai, F., 2022. Potential Exploration of Segmentation driven Stereo Matching of very high-resolution Satellite Imagery. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B5-2022, 67-72.

Dippold, E.J., Tsai, F., 2023. Two-View Structure-from-Motion with Multiple Feature Detector Operators, Remote Sensing.

Djuric, I., Vasiljević, I., Obradovic, M., Stojaković, V., Kicanovic, J., Obradovic, R., 2021. Comparative Analysis of Open-Source and Commercial Photogrammetry Software for Cultural Heritage.

Erdős, L., Bátori, Z., Tölgyesi, C., Kormoczi, L., 2014. The Moving Split Window (MSW) analysis in vegetation science – an overview. Applied Ecology and Environmental Research 12, 787-805.

Moncef Aharchi, Kbir, M.H.A., 2020. A Review on 3D Reconstruction Techniques from 2D Images, pp. 510-522.
Nesbit, P.R., Hugenholtz, C.H., 2019. Enhancing UAV–SfM 3D Model Accuracy in High-Relief Landscapes by Incorporating Oblique Images, Remote Sensing.

Rupnik, E., Daakir, M., Pierrot Deseilligny, M., 2017. MicMac – a free, open-source solution for photogrammetry. Open Geospatial Data, Software and Standards 2, 14.

Wang, J., Zhong, Y., Dai, Y., Birchfield, S., Zhang, K., Smolyanskiy, N., li, H., 2021. Deep Two-View Structure-from-Motion Revisited, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8953-8962.

Wu, C., Agarwal, S., Curless, B., Seitz, S.M., 2011. Multicore bundle adjustment, CVPR 2011, pp. 3057-3064.

Xiao, W., Cao, H., Tang, M., Zhang, Z., Chen, N., 2023. 3D urban object change detection from aerial and terrestrial point clouds: A review. International Journal of Applied Earth Observation and Geoinformation 118, 103258.

Yokoya, N., Yairi, T., Iwasaki, A., 2012. Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion. IEEE Transactions on Geoscience and Remote Sensing 50, 528-537.

Youssefi, D., Michel, J., Sarrazin, E., Buffe, F., Cournet, M., Delvit, J.M., Helguen, C.L., Melet, O., Emilien, A., Bosman, J., 2020. CARS: A Photogrammetry Pipeline Using Dask Graphs to Construct A Global 3D Model, IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 453-456.

Zawadzka, J.E., Harris, J.A., Corstanje, R., 2021. A simple method for determination of fine resolution urban form patterns with distinct thermal properties using class-level landscape metrics. Landscape Ecology 36, 1863-1876.

Zhou, H., Zhang, T., 2015. Two-view structure-from-motion algorithm. Proceedings of the World Congress on Intelligent Control and Automation (WCICA) 2015, 3345-3350.