3D Reconstruction from Multi-view Google Earth Satellite Stereo Images by Generating Virtual RPC based on 3D Homography-based Georeferencing

D. U. Seo [1], S. Y. Park [1] *

[1] Graduate School of Electronics and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea - sseodw@naver.com, sypark@knu.ac.kr

Commission II, WG II/1

KEY WORDS: Multi-view stereo matching, Multi-date satellite images, RPC Generating, Georeferencing, Google Earth.

ABSTRACT:

In this paper, we propose a method for performing 3D reconstruction by generating virtual RPC parameters from multi-view satellite stereo images provided by Google Earth (GE) software. In the multi-view stereo (MVS) image in a general case, after the pose and parameters of the camera are estimated, a dense 3D surface can be reconstructed. However, in the case of satellite images, it is not easy to obtain the original images with pose parameters of an area of interest. In the case of GE software, which can obtain images across the globe, the images provided are georeferenced and modified to fit the ground control point (GCP), so there is no camera model to explain the projection relationship. Therefore, the purpose of the proposed method is to perform 3D reconstruction by generating virtual camera parameters in modified satellite images obtained from GE software. In the proposed method, satellite images obtained from GE are estimated to be pinhole images using structure from motion (SfM) for initial reconstruction. After initial reconstruction, the 3D model is transformed from a distorted hexahedral space formed along a pixel ray to a UTM coordinate system metric space through a 3D homography-based georeferencing. A virtual rational polynomial camera (RPC) parameter is calculated through the satellite images and the 3D interspace correspondence point of UTM coordinates. The result is generated by virtual RPC and the MVS method using the RPC model. The reconstructed DSM using virtual RPC is improved over the initial reconstruction of the proposed process, and error measurement in the area with GT obtained significant results with an average of 1.366m on an MAE method.

1. INTRODUCTION

Recently, there are many studies in 3D surface reconstruction from satellite stereo images (Á. Gómez et al., 2022). Since it is performed through the camera geometry-based matching method, the Rational Polynomial Camera (RPC) model is required for stereo matching in the case of satellite stereo images. However, satellite images with RPC are not commonly available. The most accessible way to obtain satellite images of any region and various dates is to use Google Earth software (Google Earth Pro 7.3.6.9345, Accessed in 2023). Google Earth (GE) processes satellite images from various sources into an observable form and serves the public. The problem is that camera geometry such as RPC disappears during that processing. In other words, GE provides multi-view satellite images by multi-date for a specific location, but no camera geometry exists for each image. Therefore, the purpose of this paper is to generate virtual RPC coefficients for satellite images provided by GE and apply them to the previously introduced multi-view stereo method. In this paper, when the initial pinhole camera parameters are estimated from all stereo images, the 3D model is reconstructed using EnSoft3D (Lee et al., 2021), one of the cost volume-based multi-view stereo schemes. Even after creating a virtual RPC, we derive the final results using EnSoft3D, which is modified to the RPC model. EnSoft3D is a plane sweep stereo (PSS, Collins et al. 1996)-based MVS method that repeatedly refines the initial cost volume based on the consensus and visibility of object surface information at each point in time.

We generate virtual RPC models by using the relation between the metric 3D surfaces and their correspondences in the stereo images. The initial metric 3D reconstruction is done by using the perspective camera model (PCM) and an MVS method. The initial reconstruction is enhanced again by using the proposed virtual RPC models. Figure 1. shows the final height map of the proposed method. The proposed 3D reconstruction scheme consists of four steps. The first is to find virtual PCM by using a Structure from Motion (SfM) method. We use COLMAP (Schoenberger et al., 2016) as the SfM method and obtain virtual PCM in each multi-view image. In the second step, an MVS method called EnSoft3D is used for 3D reconstruction in the projective space using the virtual PCM. In the GE, the ground plane images are aligned and represented as orthoimages. However, the building structures which have high elevation are represented in different shapes in the multi-view images. For this reason, the 3D reconstruction using the virtual PCM is done in distorted hexahedral space according to the camera ray. In the third step, we correct this space by georeferencing. For georeferencing, 3D homography is employed to transform the

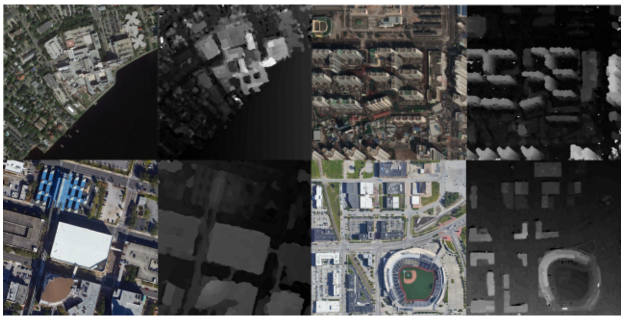

Figure 1. The result height maps of the proposed method. Input is cropped satellite images provided by GE.

* Corresponding author

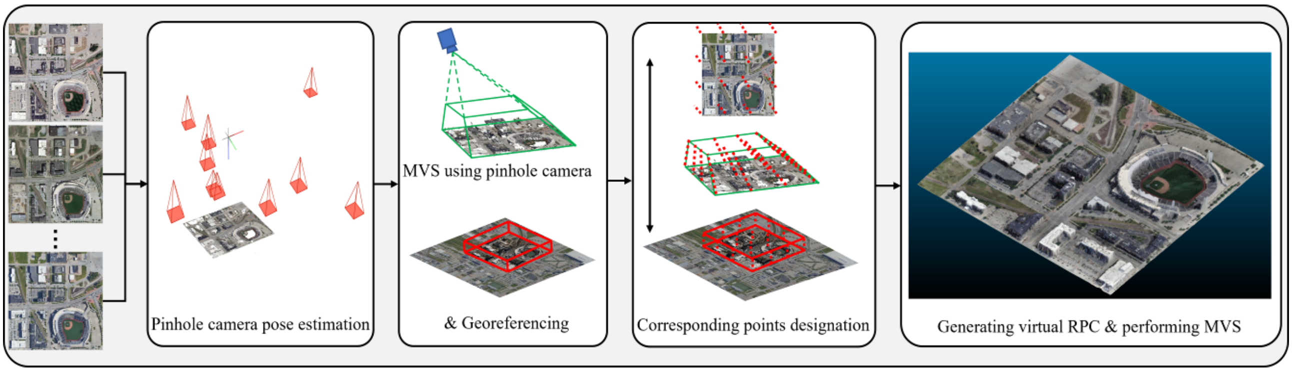

Figure 2. The proposed generating virtual RPC and 3D reconstruction scheme. The last result DSM of this scheme exists in UTM coordinates.

virtual distorted space to the georeferencing metric space of the Universal Transverse Mercator (UTM) coordinate system in GE satellite images. As the last step, virtual RPCs are generated for all images, and the 3D surface is refined by using the RPCs. This step begins with sampling a cuboid space with the WGS84 coordinate system and corresponding points in the image pixel coordinate system. Using the two groups of points, a virtual RPC is derived. With the RPC, surface reconstruction is performed again to obtain more accurate results. The pipeline of the proposed method is shown in Figure 2.

2. RELATED WORKS

Several methods for 3D reconstruction of multi-view satellite stereo images have been investigated. One well-known pipeline, S2P(Facciolo et al., 2017) rectifies input image pairs using RPC parameters and subsequently reconstructs a DSM using stereo-matching methods such as MGM(Facciolo et al., 2015). In the case of multi-view stereo images, the median altitude is computed for each cell from multiple pairwise models and generates a single DSM.

In addition to pairwise methods, certain studies have explored non-pairwise true multi-view stereo approaches that involve fitting the RPC model to the PCM without directly utilizing it. In research by Zhang (Zhang et al., 2019), 3D reconstruction was performed by replacing camera parameters with a fitted PCM in COLMAP and PSS, which was previously introduced in the field of computer vision. The results demonstrated that the fitted model was not much different from the results of S2P using RPC parameters. Fitting the RPC model to PCM, the proposed method of Zhang, is expected to be applicable to various studies.

There is also a paper that applies learning methods to previous studies and compares them (A. Gómez et al., 2022). In this paper, the author applied the stereo-matching method of S2P from MGM to GA-Net (F. Zhang et al., 2019), and the fitted camera parameter to CasMVSNet (X. Gu et al.,), not COLMAP. In comparison, it can be seen that the learning method performs better than the method using geometry, and the pairwise performs better than the true multi-view. There is also a study that applies RPC models to learning methods without modification of satellite images or camera models (J. Gao et al., 2021). In that paper, RPC parameters are transformed into tensors with a rank of 3 and a shape of 4x4x4 and it shows better performance in error measure than fitted PCM on large image size that over 5120 x 5120.

These studies basically require RPC parameters that describe the projection relationship with satellite images. However, as we explained in the intro, the satellite image of Google Earth, which is provided in a modified state, does not have RPC parameters. In some cases, it may be estimated by a PCM model using the SfM method, but there is a large error because of domain differences. To solve this problem, by the proposed method, we generate virtual RPC parameters to multi-view and multi-date satellite images that came from GE.

3. METHODOLOGY

3.1 Initial 3D Reconstruction

The depth value of the ground plane in SfM space must be accurately estimated in each view at the first 3D reconstruction. This is because the homography matrix used later for georeferencing is calculated based on that information. Therefore, we used EnSoft3D, one of the PSS-based true multi-view methods that have reliable performance.

When PCM parameters are calculated through COLMAP on multi-view and multi-date satellite images obtained from GE, PSS is performed to construct the initial cost volume. From that, consensus volumes and visibility volumes of each view are generated. The consensus volume contains the probability that which disparity value has the object's surface, and the visibility volume contains information for determining the occlusion area. EnSoft3D updates the cost volume and visibility weights by combining the information of the two volumes in one iteration. In addition, refinements are applied to the textureless region by using approximately estimated plane information. As the iteration process increases, the accuracy of the surface information stochastically selected at the multi-view increases. In this paper, the number of iterations is 10 times for all cases.

The camera pose estimated by the SfM method exists within a unique coordinate system, and reconstruction also proceeds within this coordinate system. Because of the use of inaccurate camera models, reconstruction exists in a hexahedral space distorted along the ray of the camera. Figure 3. shows some results of the initial 3D surface reconstruction. The red dotted line at the bottom of the figure represents the camera ray.

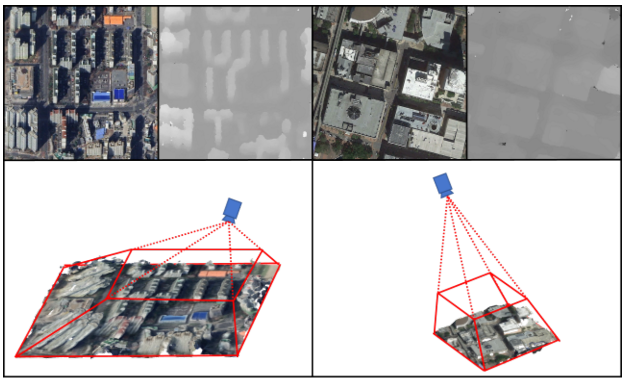

Figure 3. 2 cases of MVS result with estimated PCM.

3.2 Georeferencing

The RPC is a parameter used to describe the projection relationship between the wgs84 coordinate system and the pixel coordinates of the satellite image. To generate a virtual RPC, it is necessary to know the correspondence point between two coordinate systems. Therefore, we need to transform the 3D model reconstructed in the distorted SfM coordinate system in the previous part into the orthogonal space of the wgs84 coordinate system. In other words, what we already know is the transformation between the image pixel coordinates and the SfM coordinates, and what we need to know is the transformation between the SfM coordinate and the WGS84 coordinate. A 3D homography transform matrix is calculated for transformation between 3D coordinate systems (Lee et al., 2022). Eight pairs of 3D corresponding points in each coordinate system are needed for its calculation. In the SfM coordinate system, four ground vertices and four points are selected existing according to the camera ray at the corresponding vertices. In the WGS84 coordinate system, the coordinates of the ground surface can be obtained from the GE. The problem is the four upper points of the wgs84 coordinate system. The space in which the initial 3D information is obtained from the satellite image is expressed as a randomly generated coordinate system through the SfM method. Therefore, since the scale of the z-axis is an unknown parameter, the altitude of a specific structure in the satellite image must be known for accurate geographic reference. If there is no known altitude in satellite images, an arbitrary scale should be specified for georeferencing. After a scale for the wgs84 coordinate system is determined, a 4 x 4 transformation matrix is calculated using the following equation.

$$V_{SfM} = \left[v_{S,0}, v_{S,1}, \dots, v_{S,7}\right], V_{wgs} = \left[v_{w,0}, v_{w,1}, \dots, v_{w,7}\right] \quad (1)$$

$$H = SVD\left(V_{SfM}, V_{wgs}\right) = \begin{bmatrix} h_{01} & h_{02} & h_{03} & h_{04} \\ h_{05} & h_{06} & h_{07} & h_{08} \\ h_{09} & h_{10} & h_{11} & h_{12} \\ h_{13} & h_{14} & h_{15} & 1 \end{bmatrix} \quad (2)$$

$$P' = H \cdot \begin{bmatrix} p_x \\ p_y \\ p_z \\ 1 \end{bmatrix} \quad (3)$$

The V_{SfM} and V_{wgs} mean eight vertices in each coordinate system. SVD is singular value decomposition and H is 4 x 4 3D homography transform matrix. P' is a transformed 3D point by matrix H from distorted SfM coordinates (p_x, p_y, p_z). In Figure 4. it can be seen that the tilted buildings are transformed to be orthogonal to the ground. Additionally, we check in metric space (UTM) to see if the matrix is correctly calculated.

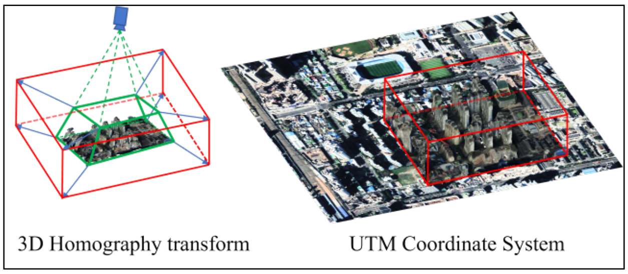

Figure 4. A distorted green hexahedron exists in the SfM coordinates system and a red one exists in the wgs84 coordinate system. For visualization, the 3D model is one more transformed into a matric coordinate system (UTM).

3.3 Virtual RPC Generating

If the 3D homography transform matrix generated by SVD is accurate, the 3D surface model converted to the UTM coordinate system will be represented in the orthogonal metric space. To generate virtual RPC, at least 39 correspondence points between the WGS84 coordinate system and the image pixel coordinate system are required. Thus, an orthogonal cuboid space is uniformly sampled in the UTM coordinate system and the sampled 3D points are converted to WGS84 to find their correspondences in the image. When converting to the WGS84 coordinate system, the zone number and zone letter in the UTM coordinates indicated by the GE are required. The conversion of sampled 3D points to the pixel coordinate system requires the camera parameters of each point of view estimated through COLMAP and the 3D homography transformation matrix calculated through SVD. The sampled 3D points in the UTM coordinate system are sent back to the distorted hexahedral space using the inverse matrix of the 3D homography and then projected into the pixel space of each point of view using the parameters of each pinhole camera. We sampled [10x10x10] points. This process is shown in Figure 5.

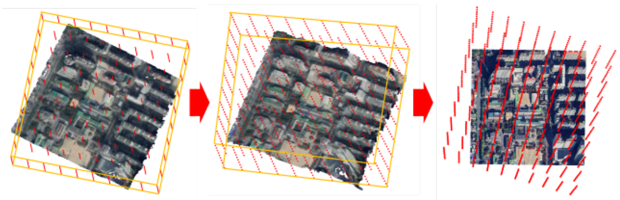

Figure 5. The first model exists in UTM coordinates system, second is in SfM coordinate system and last one is in image pixel coordinate system. All red dot is correspondence points for RPC calculation.

After the corresponding points in the two coordinate systems are determined, the points are normalized by the offset and scale suitable for each parameter. The offset and scale are determined considering the size of the GE satellite image and the distribution of WGS84 coordinates. After normalization, the virtual RPC is calculated by SVD. The processing is similar to equation (1,2,3) as follows:

$$V_{pixel} = \left[v_{p,0}, v_{p,1}, \dots, v_{p,99}\right], V_{wgs} = \left[v_{w,0}, v_{w,1}, \dots, v_{w,99}\right] \quad (3)$$

$$Virtual\ RPC = SVD_r\left(V_{pixel}, V_{wgs}\right) = \begin{bmatrix} s_{01} & s_{02} & \dots & s_{20} \\ 1 & s_{21} & \dots & s_{39} \\ l_{01} & l_{02} & \dots & l_{20} \\ 1 & l_{21} & \dots & l_{20} \end{bmatrix} \quad (4)$$

$$pixel_x = \frac{s_{01} + s_{02} \cdot lon' + \dots + s_{20} \cdot h'^3}{1 + s_{21} \cdot lon' + \dots + s_{39} \cdot h'^3} \cdot x_s + x_o$$
$$pixel_y = \frac{l_{01} + l_{02} \cdot lon' + \dots + l_{20} \cdot h'^3}{1 + l_{21} \cdot lon' + \dots + l_{39} \cdot h'^3} \cdot y_s + y_o \quad (5)$$

$s_{01\sim39}$ and $l_{01\sim39}$ are RPC parameters and lon', lat', and h' are normalized parameters of longitude, latitude, and height. x_s and y_s are scale parameters and x_o and y_o are offset parameters.

The generated virtual RPC enters the MVS, which created the initial 3D model, as an input with satellite images, and generates the final DSM. Inverse RPC parameters for transferring information from the image pixel coordinate system to the wgs84 coordinate system are also calculated through the process of equation (3,4).

https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1075-2023 | © Author(s) 2023. CC BY 4.0 License.

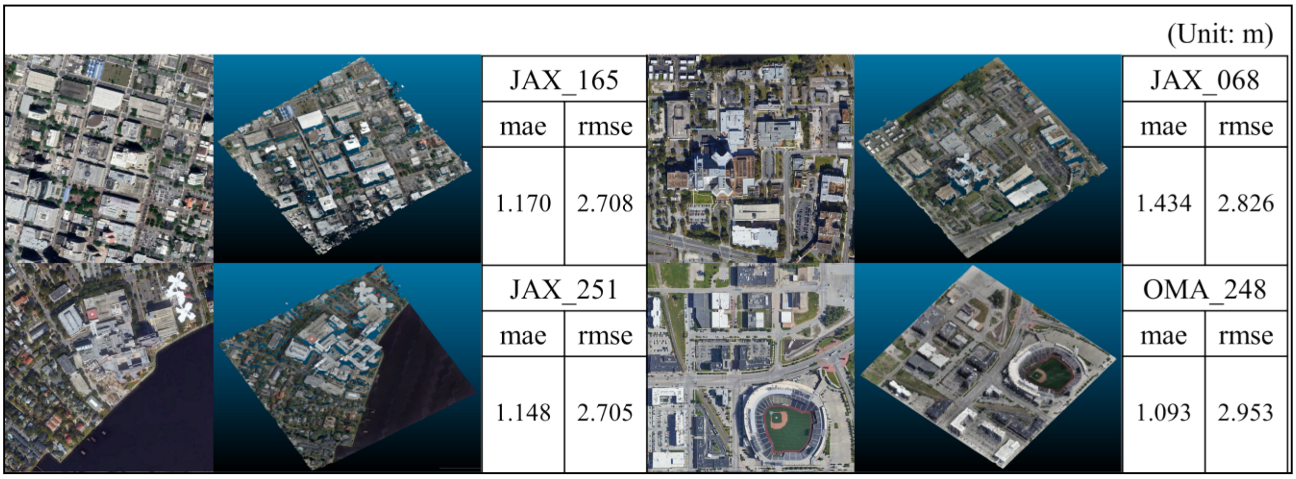

Figure 6. Results of the proposed method using GE satellite images of areas where data exists in DFC19. It's RGB satellite images, DSMs generated in the UTM coordinate system, and measurement errors in the order.

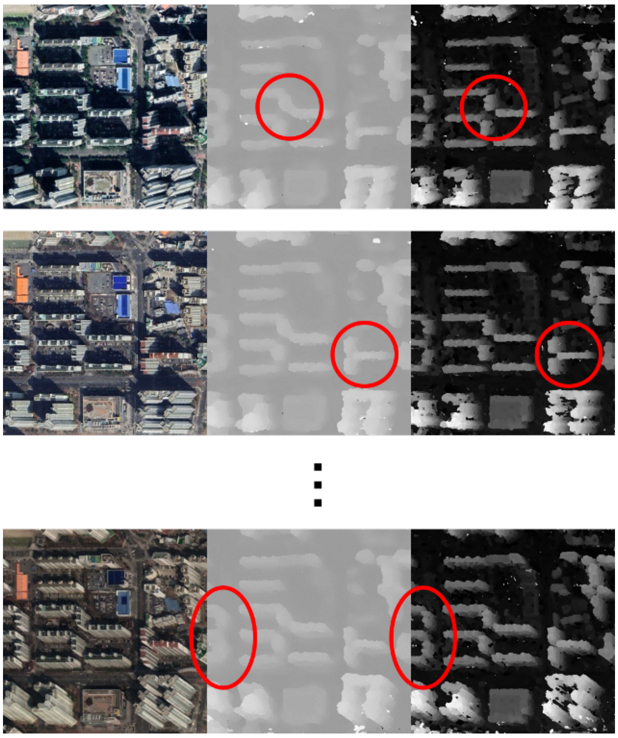

Figure 7. Comparison between the initial height map and the final height map. RGB satellite image on the left, initial height map in the middle, and final height map on the right.

4. EXPERIMENTS

This study started to generate virtual parameters in an environment without RPC parameters, but experiments were conducted in areas where GT exists to measure errors for performance judgment. In the experiment, DFC19(B. Le Saux et al., 2019) data was used. DFC19 data include multi-view satellite images and LiDAR data from Jacksonville and Omaha in the United States. Among them, only LiDAR data was used as GT. All input satellite images were obtained by cropping from a GE and were resized to 1024 x 1024. And ten input images were used in all cases. As described in the previous section, COLMAP is used as the SfM method and EnSoft3D is used as the MVS method. During the 3D homography conversion, the scale was obtained from the corresponding LiDAR data in the case of DFC19 data with GT, and if there is no GT, the height of one of the buildings in the image was determined and designated through search. When the final MVS was performed after the virtual RPC was created, the interval of the planes in the PSS was set to 1 m.

Figure 6. shows the results of the proposed method, and Figure 7. shows the performance of the virtual RPC compared to the calculated PCM results.

On average, the MAE method was measured at 1.211m and the RMSE method was measured at 2.798m. In addition, it can be seen that the virtual RPC parameter generated by the proposed method shows an improved performance than the PCM parameter calculated by COLMAP at the red circle in Figure. 7.

5. CONCLUSION

In this paper, we propose a method to reconstruct a 3D surface by generating a virtual RPC from a satellite image without camera parameters. Although the SfM method was performed to estimate the camera parameters from images that could be obtained from GE, the reconstructed model existed in a distorted space. Therefore, we moved the model from a distorted space to a space orthogonal to the surface using a 3D homography transformation matrix calculated using SVD. In addition, points obtained by sampling the rectangular orthogonal space in the WGS84 coordinate system were projected into the pixel coordinate system of the satellite image with a 3D homography transform matrix and PCM parameters to set a corresponding point. A virtual RPC was created using the set corresponding point, and the MVS was finally performed by applying the generated virtual RPC. As a result of 3D surface reconstruction using virtual RPC, the tilting problem of the building is solved. Experimental results show that the proposed RPC generation method can enhance the performance of 3D reconstruction from multi-view satellite stereo images. Figure 8. Show it. The proposed method can be used for any multi-view satellite stereo image without RPC.

6. LIMITATION

There are several limitations to the method we propose. First, there is the scale problem described in the georeferencing section. When estimating PCM using COLMAP from a satellite

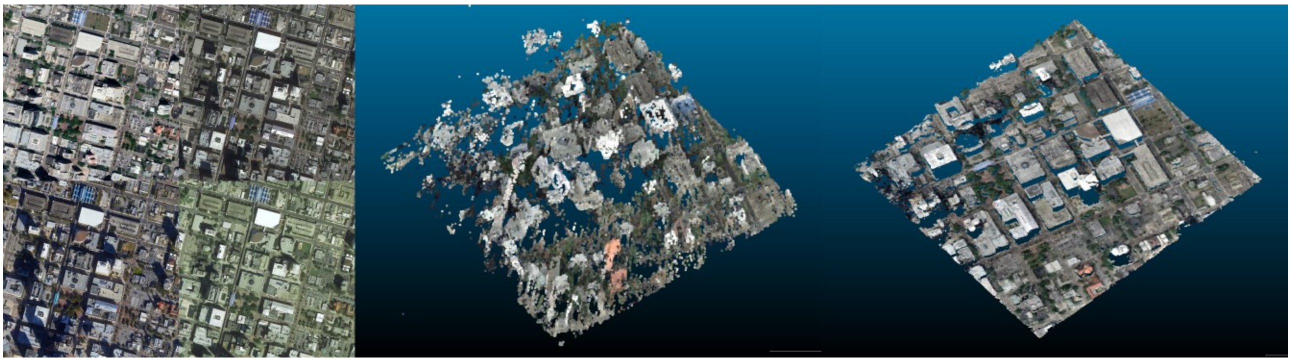

Figure 8. Comparison of results. The lefts are rgb satellite images obtained from GE, the middle is COLMAP (SfM + stereo) result, and the right is a result of the proposed method. When the stereo method is performed with PCM, more errors appear than the virtual RPC model.

image cropped in a GE, the scale of the Z-axis is arbitrarily specified because it is calculated as an inaccurate camera model. Figure 9 shows that.

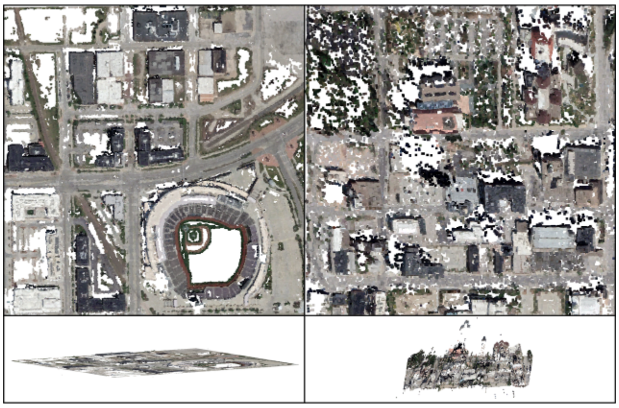

Figure 9. The above two images are orthogonal images of the model reconstructed through COLMAP in each case, and the image below is a side view.

On the left, it can be seen that the structure in the satellite images is almost reconstructed to the ground, and on the right, on the contrary, it is reconstructed to a higher height than the original height. Therefore, for accurate transformation, it is necessary to know the altitude of at least one structure in the satellite image.

The second limitation is the problem of the satellite image itself obtained from GE. The satellite image provided by GE after processing is a modified image based on GCP. Therefore, if there are many GCPs in the satellite image and a large deformation occurs. In other words, when there is a large amount of GCP in the satellite image and a large deformation occurs, the final result is produced like Above Ground Level (AGL), not DSM.

In future studies, we would like to research a method to estimate the scale in the WGS84 coordinate system or in the SfM coordinate system and a way to rectify distortion in the image to overcome the problems. If these methods are studied, it is expected that more accurate DSMs can be generated from all images obtained from GE.

ACKNOWLEDGMENTS

This work has been supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT)(No. 2021R1A2C2009722) and supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1A6A1A03043144).

REFERENCES

Á. Gómez, G. Randall, G. Facciolo, and R.G. von Gioi, 2022: “An experimental comparison of multi-view stereo approaches on satellite images,” 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 707-716, 2022.

Google Earth Pro 7.3.6.9345, Accessed in 2023. [online]: https://earth.google.com/web/

Lee, M. J., Um, G. M., Yun, J., Cheong, W. S., Park, S. Y. 2021: Enhanced Soft 3D Reconstruction Method with an Iterative Matching Cost Update Using Object Surface Consensus. *Sensors*, 21(19), pp. 6680.

Collins, R. T., 1996. A space-sweep approach to true multi-image matching. Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition

Schönberger, J. L., Zheng, E., Frahm, J. M., Pollefeys, M. 2016. Pixelwise view selection for unstructured multi-view stereo. *Proceedings of the European Conference on Computer Vision* (ECCV), Amsterdam, Netherlands.

Facciolo, G., De Franchis, C., & Meinhardt-Llopis, E. 2017. Automatic 3D reconstruction from multi-date satellite images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

Zhang, K., Snavely, N., & Sun, J. 2019. Leveraging vision reconstruction pipelines for satellite imagery. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops.

Gabriele Facciolo, Carlo de Franchis and Enric Meinhardt. 2015. MGM: A Significantly More Global Matching for Stereovision. Proceedings of the British Machine Vision Conference (BMVC), pages 90.1-90.12. BMVA Press.

Feihu Zhang, Victor Prisacariu, Ruigang Yang, and Philip H.S. Torr. 2019. GA-Net: Guided Aggregation Net for End To-End Stereo Matching. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 185–194, Long Beach, CA, USA, June 2019. IEEE.

Xiaodong Gu, Zhiwen Fan, Siyu Zhu, Zuozhuo Dai, Feitong Tan, and Ping Tan. 2020. Cascade Cost Volume for High-Resolution Multi-View Stereo and Stereo Matching. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Pages 2495-2504.

J. Gao, J. Liu, and S. Ji, "Rational Polynomial Camera Model Warping for Deep Learning Based Satellite Multi-View Stereo Matching," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 6128-6137, doi: 10.1109/ICCV48922.2021.00609.

Lee, M. J., Park, S. Y. 2022. 3D Surface Reconstruction from Multi-view and Multi-data Google Earth Satellite Images with 3D Homography-based Projective Reconstruction. ISPRS Congress (2022 edition)

B. Le Saux, N. Yokoya, R. Hansch, and M. Brown, "2019 IEEE GRSS Data Fusion Contest: Large-Scale Semantic 3D Reconstruction [Technical Committees]", IEEE Geoscience and Remote Sensing Magazine, 2019.