Model-Based MVS Point Reconstruction of Texture-Less Regions with Epipolar Constraints

Yuichiro Yamaguchi, Masafumi Nakagawa

Shibaura Institute of Technology, 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan
Email: me20130@shibaura-it.ac.jp; mnaka@shibaura-it.ac.jp

KEY WORDS: Structure from Motion; Multi View Stereo; Camera direction constraints; BIM

ABSTRUCT: Point clouds are acquired with Structure from Motion/Multi-View Stereo (SfM/MVS) and laser scanning for Building Information Modeling (BIM). Although the SfM/MVS can generate dense point clouds, it is not easy to reconstruct texture-less regions because the SfM/MVS is based on feature-based image matching. Thus, in metal bridge measurements, point clouds are not generated in many texture-less regions such as the plane of the girder. Therefore, we propose a model-based MVS methodology with epipolar constraints using the intrinsic parameters and extrinsic parameters estimated with SfM processing. Our point cloud reconstruction approach consists of SfM, texture-less region selection with sparse point cloud back-projection, and dense point cloud generation with model-based MVS. We selected metal bridge girders as measured objects. Through our experiment, we confirmed that our methodology can reconstruct point clouds, even if measured regions are texture-less.

1. INTRODUCTION

Recently, point clouds have been used for 3D modeling of structures and ground surfaces for building information modeling (BIM). BIM can improve the efficiency of construction projects with 3D models and information related to the planning, investigation, design, construction, and maintenance. Terrestrial laser scanners, mobile mapping systems, and unmanned aerial vehicles (UAVs) are generally used for 3D acquisition for BIM. Other point cloud acquisition methodologies are structure from motion (SfM) (Tomasi and Kanade, 1992) and multi-view stereo (MVS) (Seitz et al., 2006) with many images acquired from various positions and angles. Although feature-based image matching requires textures and clear edges in images, the SfM/MVS can generate dense point clouds with a lower cost than laser scanning. Feature-based image matching has the possibility to be robust against scale and brightness changes among images with feature descriptors, such as scale-invariant feature transform (David, 2004). The SfM/MVS can be applied for 3D data acquisition in bridge inspection (Kimoto, 2018). However, missing areas often remain in point cloud generation in metal bridge measurements because of texture-less regions. Therefore, we proposed a methodology to estimate missing areas caused by texture-less regions in point cloud generation. We focus on a model-based MVS processing with epipolar constraints using the intrinsic parameters and extrinsic parameters estimated with SfM processing.

2. METHODOLOGY

Our proposed methodology is shown in Figure 1. First, images are acquired from various points and angles. Next, conventional SfM/MVS processing is applied to estimate camera parameters and initial point clouds. Then, missing points of texture-less regions are generated with model-based MVS processing.

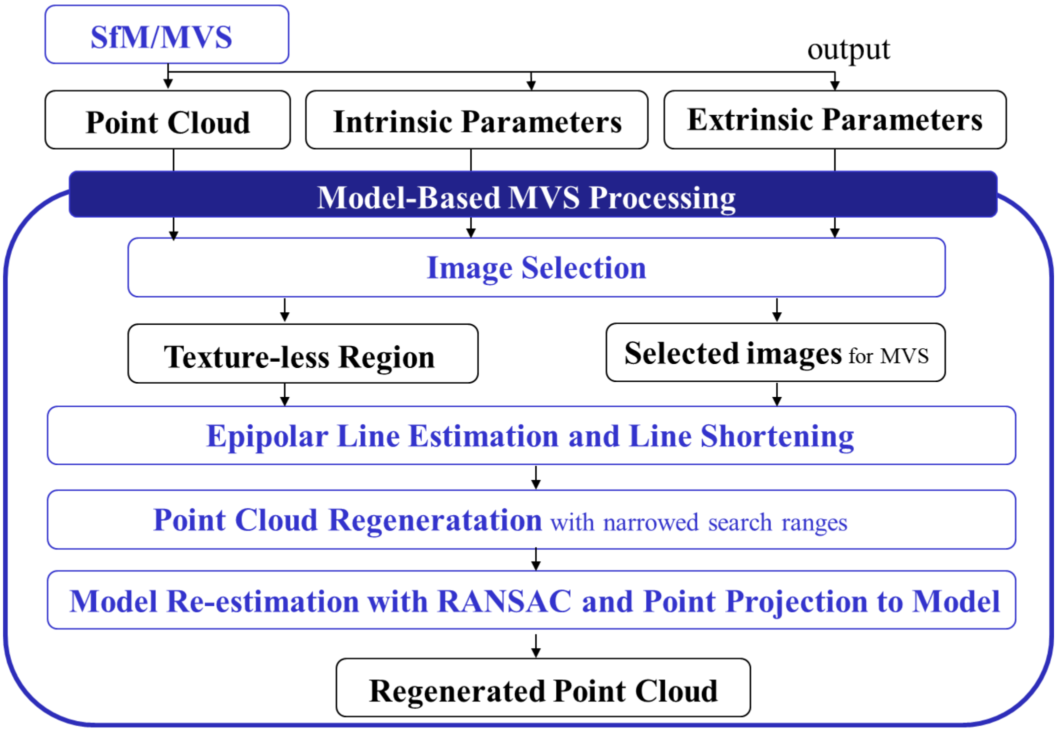


Figure 1. Proposed methodology

The model-based MVS processing consists of the image selection, epipolar line estimation, epipolar line shortening, and model re-estimation with random sample consensus (RANSAC).

2.1 Image selection

In image selection, images including missing areas are selected from all acquired images. First, missing areas are selected from dense point clouds generated by conventional SfM/MVS processing. Next, images including the missing areas are selected from all acquired images with point cloud back-projection based on Eq.1.

$$sx = A[R|t]X_w \quad \text{(Eq.1)}$$

s is a scale parameter, A are intrinsic parameter, [R|t] are extrinsic parameters,
x is a pixel coordinate point, X_w is a world coordinate point.

Figure 2 shows the image selection methodology. When all vertex points of missing areas are back-projected onto an image, the image is used for the next processing. Next, the selected image is used as a base image to search corresponded images to regenerate point clouds.

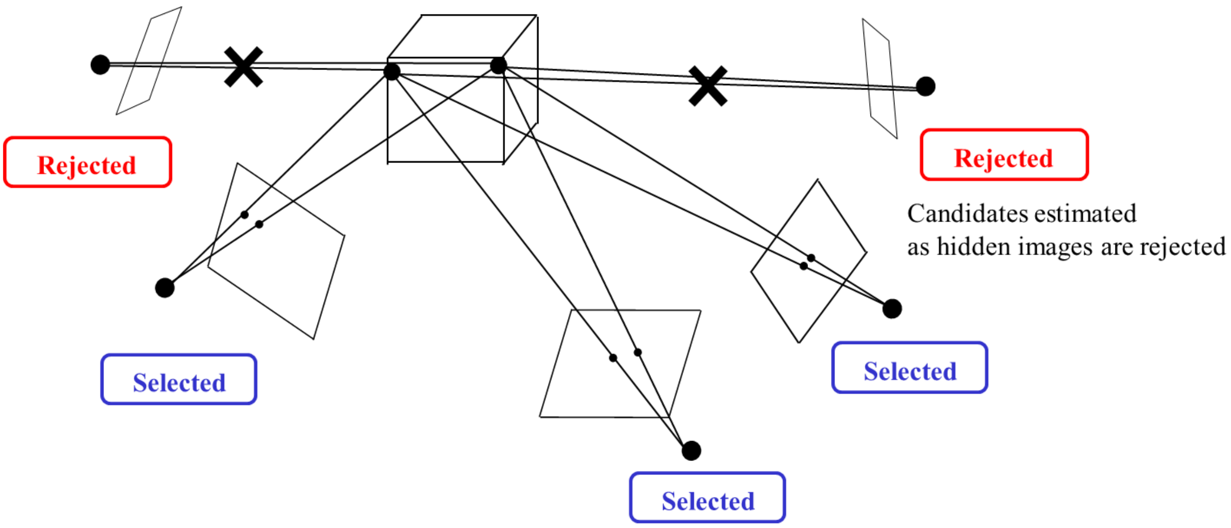


Figure 2. Image selection by back-projection of point clouds

2.2 Epipolar line estimation

The fundamental matrix F is derived by intrinsic and extrinsic parameters estimate with the SfM processing. Generally, features and corresponding points in images are used for the fundamental matrix estimation. Although the accuracy of the estimation becomes low when an image holds texture-less regions, intrinsic and extrinsic parameters of each camera can be used as the initial values. Eq.2 shows the fundamental matrix.

$$F = K'^{-T}[t_x]RK^{-1} \quad \text{(Eq.2)}$$

K is an intrinsic matrix of base image,
t_x is a skewed symmetry matrix of the translation vector between cameras,
R is a rotation matrix between cameras, K' is an intrinsic matrix of selected images.

2.3 Epipolar line shortening

A 3D model (plane) is generated from point clouds around in the targeted area. The search range for image matching is determined with the area around the model. Figure 3 describes that the epipolar line shortening reduce the search range of the matching to about 10 pixels (the number of pixels on the shortened search range along an epipolar line). Thus, even if the targeted areas are texture-less regions, point clouds are generated roughly.

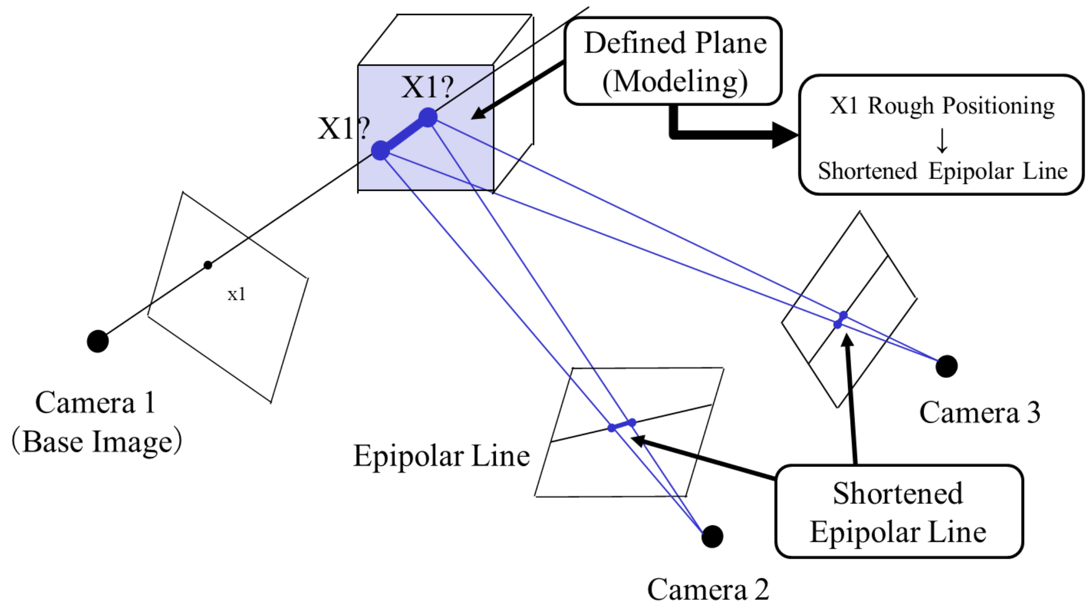
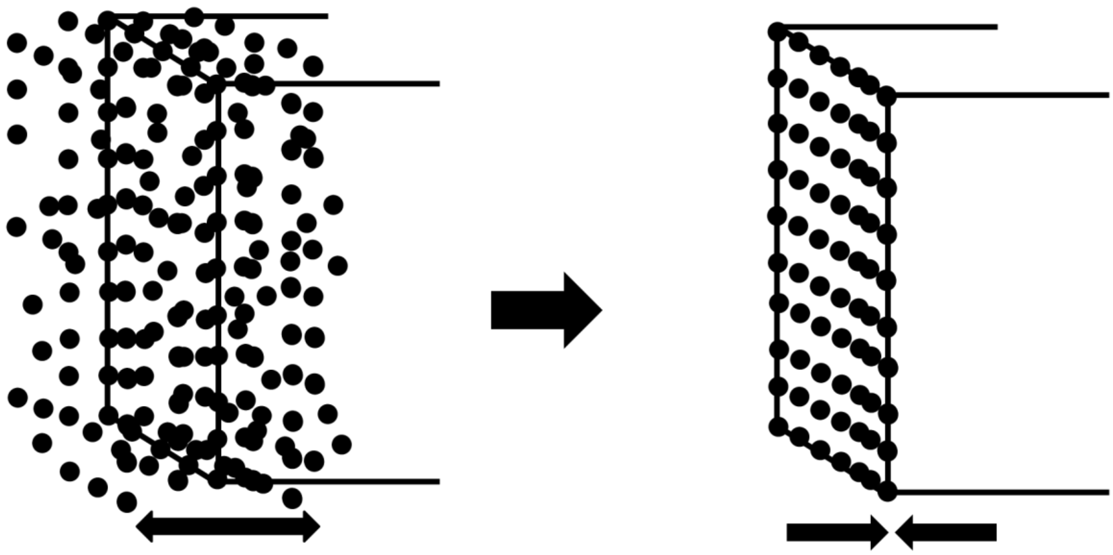


Figure 3. Geometric constraints with 3D model in image matching

2.4 Model estimation with RANSAC

On the shortened epipolar line, image matching is performed on all pixels and point clouds are generated. For all generated point clouds, the model is re-estimated with RANSAC (Fischler, 1981). Furthermore, a point cloud projection onto the re-estimated plane is applied for point cloud regeneration. Figure 4 shows that the model estimation with RANSAC and point cloud projection onto the estimated plane.



Generated all point clouds on each epipolar line

Merge additional point clouds into SfM/MVS point clouds, re-estimate a plane with RANSAC, and project additional point to the estimated plane

Figure 4. Model estimation with RANSAC and point cloud projection onto the estimated plane

The above-mentioned processing is called as model-based MVS. The point cloud generation is evaluated by the number of generated point clouds for all the pixels in the selected area of the image with the point cloud generation rate, as described in Eq.3.

$$Generation\ rate\ [\%] = \frac{The\ number\ of\ generated\ point\ cloud}{The\ number\ of\ pixels\ of\ targeted\ regions} \times 100 \quad (Eq.3)$$

3. EXPERIMENTS

A metal girder bridge was selected as a measured object, as shown in Figure 5. The measured object includes flat surfaces, edges, and texture-less regions. Images were acquired from 24 positions with various directions using a hand-held digital camera (Table 1). The acquired images were used for SfM/MVS processing using VisualSfM.

Figure 5. Measured object (metal girder bridge)

Table 1. Hand-held digital camera (DSC-HX60V, SONY)

Image size	2 million pixels (1920*1080)
Image sensor	1/2.3 type, "Exmor R" CMOS
Focal length	f = 4.3-129mm

4. RESULTS

4.1. Conventional SfM/MVS Processing Results

Table 2 and Figure 6 show conventional SfM/MVS processing results.

Table 2. Conventional SfM/MVS processing results

The number of images	The number of point clouds	Point cloud generation rate
421	1,917,751	27.8%

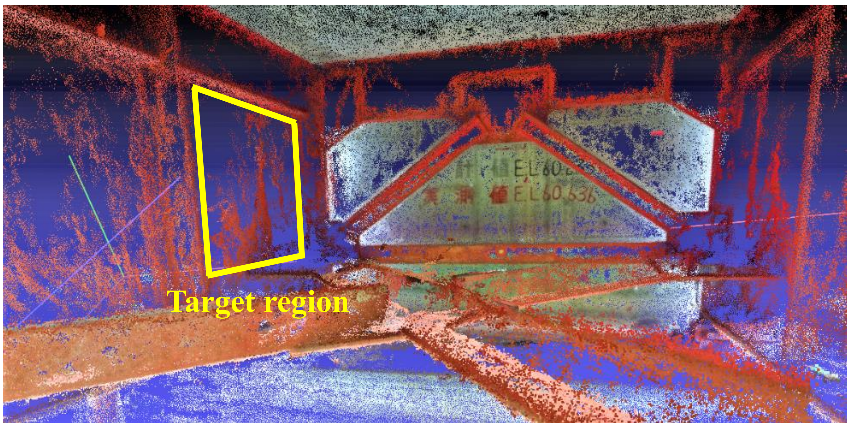


Figure 6. Conventional SfM/MVS processing results

4.2. Proposed Methodology Processing Results

In the image selection processing, the targeted region shown in Figure 6 was selected as a missing area, and 82 images were selected for point cloud regeneration. In the epipolar line estimation, a base image was selected from the selected images, and the other images were selected as reference images. The search range was narrowed to 10 pixels with epipolar constraints. Then, we performed image matching along the epipolar line with a shorter search range using geometric constraints and planes estimated from RANSAC. The results after point cloud regeneration are shown in Table 3 and Figure 7.

Table 3. Proposed methodology processing results

The number of images	The number of point clouds	Point cloud generation rate
82	63,614	95.8%

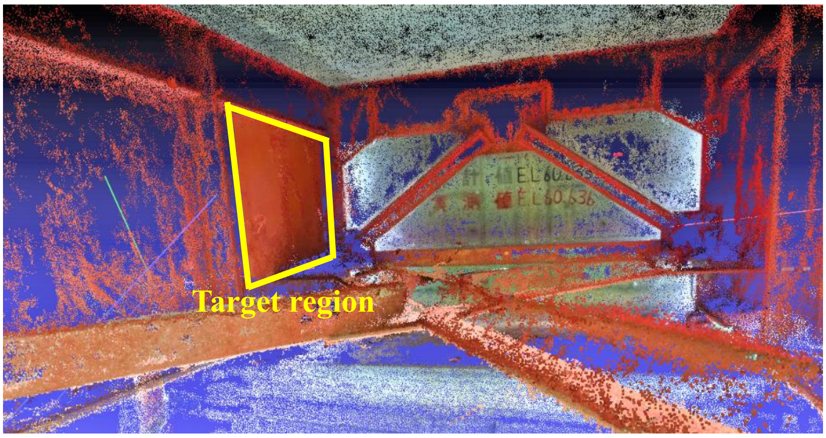


Figure 7. Regenerated point clouds

5. DISCUSSION

The conventional SfM/MVS processing reconstructed point clouds enough around edges and corners of girders. However, point clouds were missed around the main girder planes because of non-textured regions for image matching and SfM/MVS processing.

On the other hand, in our proposed methodology, model-based MVS reconstructed point clouds enough around the main girder planes. In the image selection process, images are selected automatically. Camera parameters were used for back projection processing in image selection processing. The camera parameters are estimated with bundle adjustment in the conventional SfM/MVS (Lourakis, 2009). However, the estimation accuracy was low because of insufficient feature and corresponding points. Thus, some unnecessary images were selected.

Next, Figure 8 shows a part of the results of the epipolar line estimation. Although the epipolar lines were estimated with intrinsic and extrinsic camera parameters obtained from the SfM processing, epipolar lines were wrongly estimated.

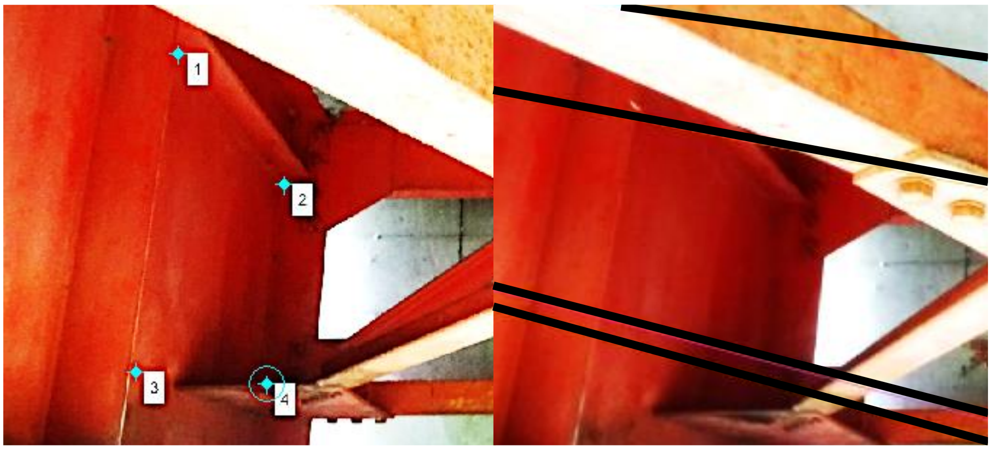


Figure 8. Estimated epipolar lines (Left: base image, right: reference image)

In addition, search ranges were shortened with plane estimation using RANSAC. Figure 9 shows the regenerated point clouds by image matching with geometric constraints. We have confirmed that point clouds were regenerated successfully even if targeted areas are texture-less regions. However, some points were excluded as false matching points due to the error of epipolar line estimation. Thus, the point cloud generation rate was 95.39%.

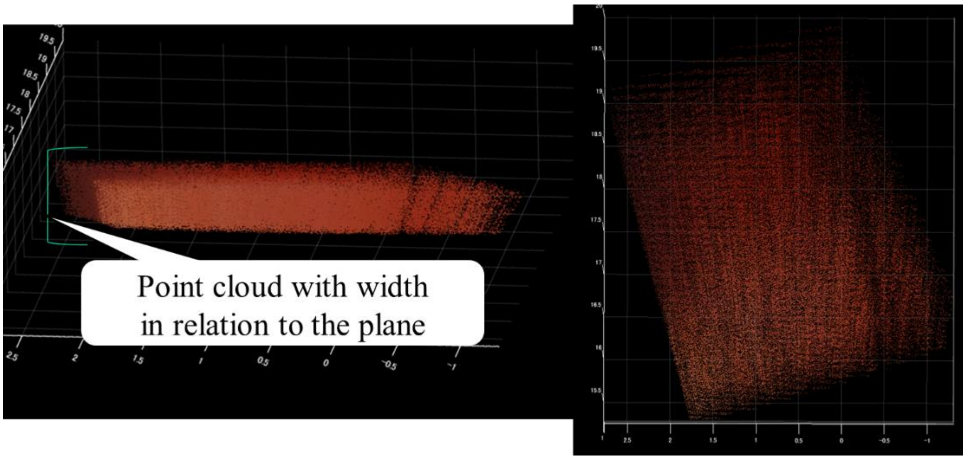


Figure 9. Regenerated point clouds

In this study, although a plane was used as a 3D model, a sphere and rectangular object can be used as additional 3D models to improve the performance of point cloud reconstruction.

6. CONCLUSION

We proposed a methodology to generate additional point clouds for texture-less regions of measured objects with model-based MVS. Through experiments on point cloud acquisition of metal girder bridge, we confirmed that the proposed methodology can reconstruct point clouds in texture-less regions such as the main girder flats.

REFERNCES

David, G. L., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), pp.91-110.

Fischler, M. and Bolles, R. 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Communications of the ACM, 24(6), pp. 381-395.

Kimoto, K., Matsuda, H., 2018, Comparison and Verification of Utilizing Method Using SfM in Inspection of Medium and Short Span Bridges. Journal of the Japanese Society for Experimental Mechanics, Volume 17, Issue 4, pp.290-297.

Lourakis, M. and Argyros, A. 2009. SBA: A software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software (TOMS), 36(1), article No.2, pp.1-30.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R., 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, pp.519-528.

Tomasi, C., and Kanade, T. 1992. Shape and motion from image streams under orthography: a factorization method. International Journal of Computer Vision, 9(2), pp.137-154.