Point Cloud Reconstruction of Texture less Regions with Topology Constraints between Corresponding Points

Yuichiro Yamaguchi, Masafumi Nakagawa

Shibaura Institute of Technology, 3-7-5, Toyosu, Koto-ku, Tokyo 135-8548, Japan
Email: ah16076@shibaura-it.ac.jp; mnaka@shibaura-it.ac.jp

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

Abstract: Point clouds are generated with structure from motion/multi-view stereo (SfM/MVS) or laser scanning for building information modeling (BIM). Although the SfM/MVS can generate dense point clouds, it is not easy to reconstruct point clouds in texture-less regions because the SfM/MVS is based on feature point-based image matching. Thus, we propose a methodology to generate point clouds of all pixels in texture-less regions with the topology constraints among corresponding images. In this study, we use the back projection of point clouds to understand the topology of the corresponding points to improve the performance of corresponding points estimation and mismatching rejection with the epipolar constraints. We conducted an experiment on the 3D modeling of a metal bridge. Through our experiment, we confirmed that our methodology can reconstruct point clouds, even if targeted images contain texture-less regions.

1. INTRODUCTION

Structure from motion/multi-view stereo (SfM/MVS) can acquire dense point clouds (Tomasi and Kanade, 1992) (Seitz et al., 2006). However, because SfM/MVS applies image matching methodologies such as scale-invariant feature transform (David, 2004), which uses feature descriptors that are robust to changes in scale and luminance between images, it cannot generate sufficient point clouds in feature-poor regions of the image (hereinafter referred to as texture less regions) (Kimoto et al., 2018).

As previous studies on point cloud generation in texture less regions, correspondence points acquisition using contrast-enhanced images has been proposed (Ley et al., 2016). Contrast-enhanced images use contrast enhancement by the Wallis filter, which is a kind of high-pass filter, and these images are processed to expand the area where point clouds can be generated. However, because this methodology increases the number of feature points by contrast enhancement, it is impossible to acquire corresponding points for all pixels in texture less regions, resulting in a point cloud with missing points. In addition, there are issues such as an increase in false matching because of the non-uniformity of the image histogram.

In the simultaneous localization and mapping process, the direct method (Engel et al., 2014) and method using the extrema of image curvature (Yokozuka et al., 2019) have been proposed as

previous studies for the dense point cloud generation. In these methods, instead of using feature points, the corresponding points are obtained by using the luminance of each pixel and the extrema of curvature. Therefore, it is possible to generate a dense point cloud even from objects with texture less regions. However, both methods require an approximate 99% overlap rate among frames, making it difficult to adapt them to image groups that are not moving images.
To generate point clouds of all pixels in texture less regions, we propose a method to understand the topology of the corresponding points by back-projecting point clouds and estimating the corresponding points in texture less regions for the point cloud generation.

2. METHODOLOGY

Our proposed method 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 topology constraints and epipolar constraints.

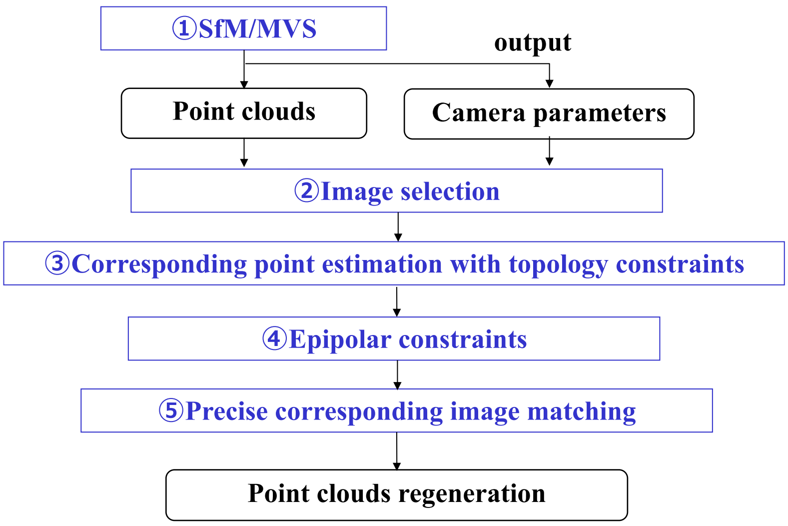


Figure 1. Proposed methodology

First, images are acquired using a handheld digital camera, and point clouds are then generated by conventional SfM/MVS. In the process of image selection, images are selected by point cloud back projection. In the process of corresponding point estimation with topology constraints, the corresponding points among selected images are estimated with topology constraints. In the process of epipolar constraints, camera parameters estimated by conventional SfM/MVS are used for epipolar constraints. Finally, the corresponding points are determined with topology and epipolar constraints. Using those points, point clouds in texture less regions are regenerated.

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: scale parameters, A: intrinsic parameters, [R|t]: extrinsic parameters,
x: pixel coordinate points, X_w: world coordinate points.

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 further processing. Next, the selected image is used as a base image to search corresponded images to regenerate point clouds

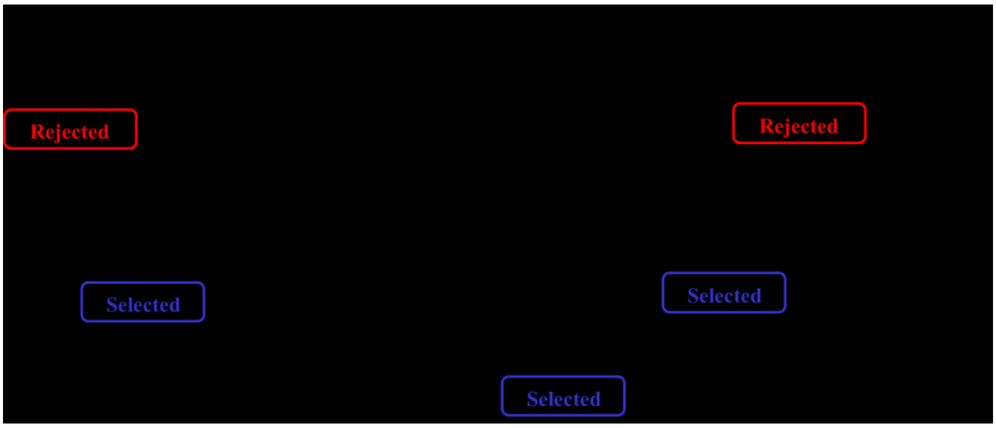


Figure 2. Image selection by back-projection of point clouds

2.2. Corresponding Point Estimation with Topology Constraints

Among the selected images, an arbitrary image is defined as a base image, and others are defined as reference images.

First, point clouds generated by conventional SfM/MVS are back-projected onto the base image. Next, all point clouds projected onto the base image are back-projected onto the reference images. Thus, pixels of the base image, pixels of the reference images, and the point clouds are combined. These combinations of images and point clouds can be used to understand how the object is seen from different viewpoints. By understanding how the object is seen, the correspondence between the image and point clouds can be roughly estimated, even for texture less regions. Therefore, we estimate the rough position of the point clouds and use it as a constraint to improve the accuracy of the precise point clouds generation.

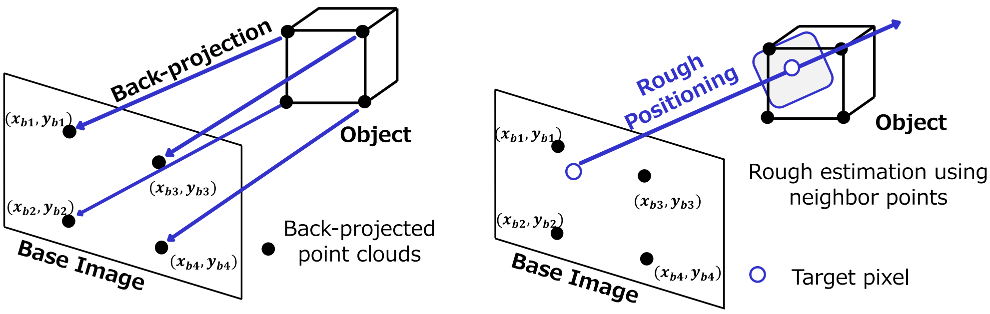
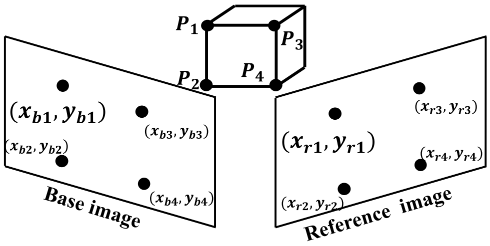


Figure 3. Rough estimation based on the combination between the image and point clouds

Corresponding points are obtained by the combination of point clouds, the base image, and the reference image. Using the corresponding points, the differences in pixel values between the base image and the reference image are computed as D_n (Figure 4). In this study, the corresponding points in texture less regions are estimated with D_n around the target pixel.



Point clouds	Base image	Reference image	The difference of pixel values
P_1	(x_{b1}, y_{b1})	(x_{r1}, y_{r1})	$[x_{r1} - x_{b1}, y_{r1} - y_{b1}]$
P_2	(x_{b2}, y_{b2})	(x_{r2}, y_{r2})	$[x_{r2} - x_{b2}, y_{r2} - y_{b2}]$
P_3	(x_{b3}, y_{b3})	(x_{r3}, y_{r3})	$[x_{r3} - x_{b3}, y_{r3} - y_{b3}]$
P_4	(x_{b4}, y_{b4})	(x_{r4}, y_{r4})	$[x_{r4} - x_{b4}, y_{r4} - y_{b4}]$

The difference of pixel values D_n [A, B] : $(x_{rn}, y_{rn}) \rightarrow (x_{bn} + A, y_{bn} + B)$

$$D_1 = [x_{r1} - x_{b1}, y_{r1} - y_{b1}],\ D_2 = [x_{r2} - x_{b2}, y_{r2} - y_{b2}],\ \dots$$

Figure 4. The difference of pixel values between the base image and the reference image

The target pixel's D_n takes a value close to the surrounding pixel's D_n. Therefore, we set constraints such that the target pixel's D_n takes a value corresponding to the positional relationship around the target pixel. Such constraints are defined as topology constraints, and the corresponding points are estimated with the constraints (Figure 5). In addition, the target pixel's D_n is calculated using Eq. 2 and Eq. 3.

$$D_n = \sum_{k=1}^{n} D_k \times bias_k \ (Eq.2), \qquad bias_n = \frac{\sum_{k=1}^{n} R_k}{R_n^2 \times \sum_{k=1}^{n} bias_k} \ (Eq.3)$$

D_n: the difference in pixel values for the target pixel,
D_k: the difference in pixel values for the surrounding pixels,
$bias_k$: the bias according to the distance from the target pixel to the surrounding pixels,
R_k: the distance from the target pixel to the surrounding pixels

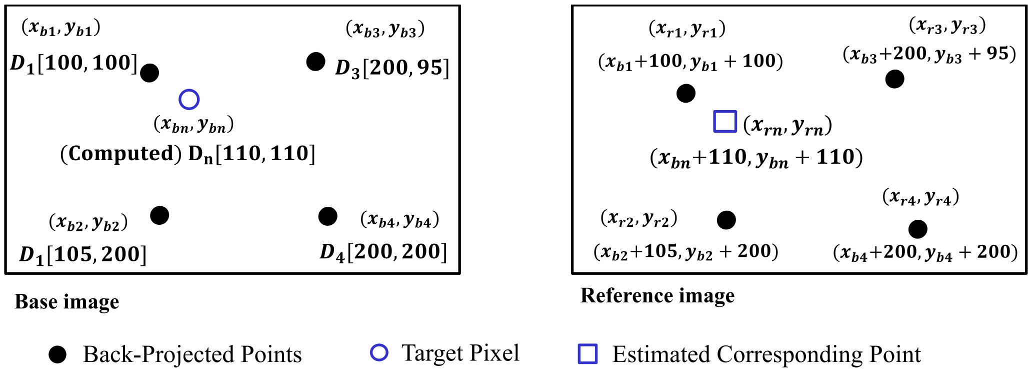


Figure 5. Corresponding point estimation with topology constraints

2.3. Epipolar constraints

For each pixel in the texture less region, epipolar lines are computed with the epipolar constraints. The camera parameters estimated by SfM/MVS are used to compute the epipolar lines.

2.4. Precise corresponding image matching

The precise corresponding points are determined by using the estimated correspondence points with topology constraints and the epipolar lines. First, the nearest neighbor point of the estimated corresponding point is selected on the epipolar line. Next, pixels on the epipolar line that exist around that point are selected as candidates for the corresponding point. For all the candidates, the similarity to the target pixel is calculated with zero mean normalized cross-correlation (ZNCC), and the pixel with the highest similarity is determined as the corresponding point (Figure 6). The additional point clouds are generated with the determined corresponding points. The point clouds whose re-projection error is less than 0.1 pixel are considered the final additional point clouds.

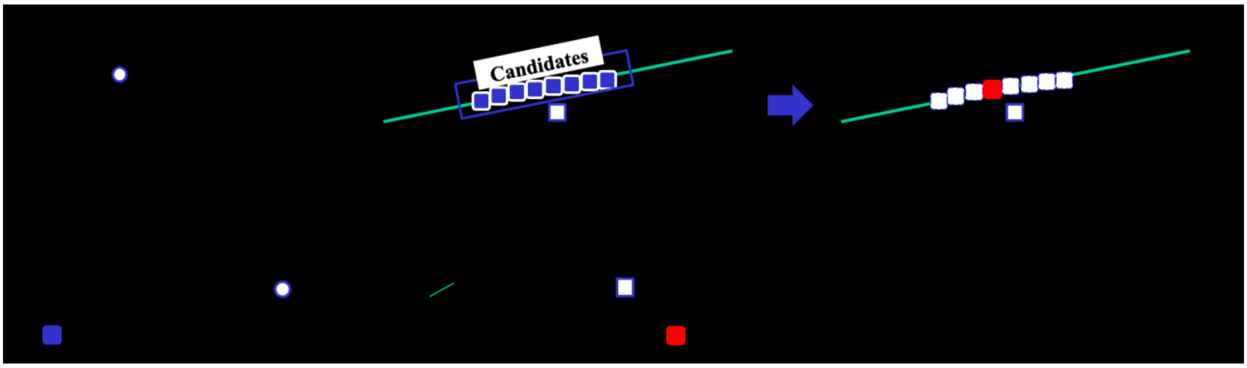


Figure 6. Precise corresponding image matching

3. EXPERIMENTS

Part of a metal bridge truss was selected as a measured object for the proposed methodology (Figure 7). The images were acquired with a handheld digital camera, and VisualSfM was used for conventional SfM/MVS processing. The surface of the truss was selected as a texture less region and point cloud generation was attempted. To evaluate the accuracy, we extracted the plane part of the point cloud, estimated the plane, and fit the extracted point cloud with the estimation result to evaluate the residue.

Figure 7. Measured object

4. RESULTS

The processing results of the proposed methodology are shown in Table 1.

Table 1. Processing results of the proposed methodology

The number of target pixels	The number of additional point clouds	The processing time	RMSE
2,516,734	1,231,126	483(sec)	0.94 mm

The target area of the proposed methodology was the plane of the metal bridge (Figure 8, left), and 2,516,734 pixels in the area were processed. The reference images selected by the image selection process (Figure 8, right) were used to generate point clouds. The point cloud whose re-projection error was less than 0.1 pixel was considered the final additional point cloud, and 1,231,126 points were generated. The processing time from the image selection process to the point cloud generation process was 483 seconds, and the root mean square error with the plane estimation result was 0.94 mm.

Figure 8. Target area for processing (left), selected reference image (right).

The comparison results between the point cloud acquired with the conventional SfM/MVS and the additional point cloud generated by the proposed methodology are shown in Figure 9.

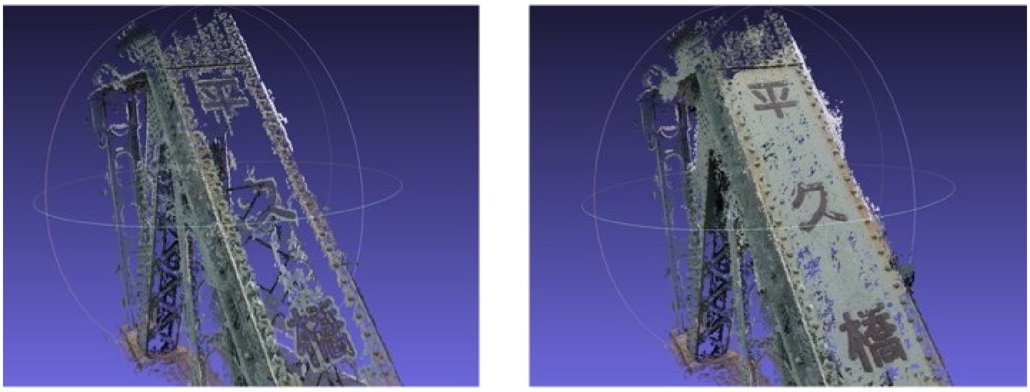


Figure 9. Conventional SfM/MVS point cloud generation result (left) and generated point clouds with the proposed methodology (right)

5. DISCUSSION

In the corresponding points estimation processing with the topology constraints, the target pixel's D_n was computed with a weight inversely proportional to the square of the distance from the target pixel to the surrounding pixels. The equation was set based on the topology constraints that the target pixel's D_n should take a value close to the surrounding pixel's D_n. By searching the nearest neighbor points of the estimated corresponding points, the number of candidate points on the epipolar line was reduced to approximately 10 points. Such processing not only reduced the mismatching by ZNCC, but also reduced the time required to determine the corresponding points to about 1/100.

In the precise correspondence point determination process, the filtering by re-projection error and the accuracy verification results confirm that the proposed methodology can determine the corresponding points in the texture-less region. However, when image rotations exist between the base image and the reference image, accurate image matching becomes difficult because conventional ZNCC is used. In our experiment, 52% of the processed pixels were mismatched. Therefore, template matching using changeable templates should be developed to improve the robustness against image rotation in image matching.

In future works, triplet matching methodology will be developed instead of stereo matching with ZNCC to improve the robustness of image matching.

6. CONCLUSION

In this study, we proposed a methodology to generate additional point clouds by using corresponding points estimation based on epipolar and topology constraints for texture less regions where point clouds are not generated by conventional SfM/MVS. Through experiments, it was confirmed that the proposed methodology can be applied to texture less regions such as metal bridge planes. In future work, we will develop a template matching based on triplet matching to improve the accuracy and robustness of image matching.

REFERNCES

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

Engel, J., Schöps, T., Cremers, D., 2014, LSD-SLAM: Large-Scale Direct Monocular SLAM, European Conference on Computer Vision (ECCV), pp.834-849.

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.

Ley, A., Hänsch, R,, Hellwich, O., 2016, Reconstructing White Walls: Multi-view, Multi-shot 3D Reconstruction of Textureless Surfaces, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., III-3, pp.91-98.

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.

Yokozuka, M., Oishi, S., Thompson, S., Banno, A., 2019, VITAMIN-E: VIsual Tracking and MappINg With Extremely Dense Feature Points, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.9641-9650.