Journal of Mathematics and Computing Science, 2023,
Volume 9, No 1, 58-65

A New Framework for 3D Point Cloud Reconstruction of Geometric Object from Multi-View Images

Intan Syaherra Ramli[1*], Rahmita Wirza O.K Rahmat[2], Ng Seng Beng[3], Hizmawati Madzin[4] and Puteri Suhaiza Sulaiman[5]

[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Cawangan Negeri Sembilan, Kampus Kuala Pilah, Malaysia
[2,3,4,5]Multimedia Department, Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, Malaysia

Authors' email: intan3885@uitm.edu.my*, rahmita@upm.edu.my, ngsengbeng@upm.edu.my, hizmawati@upm.edu.my and psuhaiza@upm.edu.my

*Corresponding author

Received 10 April 2023; Received in revised 25 May 2023; Accepted 13 June 2023
Available online 27 June 2023

Abstract: This paper proposed a new framework to enhance the 3D point cloud reconstruction from multi-view images. The accurate 3D point cloud is crucial to produce a 3D model SURF ace which can be used in many applications. There were still challenges in the existing approach to reconstruct the accurate 3D point clouds as it involves various camera orientations. This paper proposed a simplified experimental setup based on pure camera translation to capture the images from four viewpoints. The geometric object measurement is used as the ground truth data to measure the accuracy. Besides, the optical flow feature match method, perspective projection and image transformation were used to reconstruct the 3D point cloud. The result shows that all objects have an average RMSD value of 1.24mm in width and 1.26mm in depth. These values are lower than the previous method based on the images captured using known rotation, indicating higher accuracy. In future, the generated 3D point cloud can be used to create the 3D model SURF ace.

Keywords: 3D points cloud, Optical Flow, Multi-View Images

1 Introduction

3D model reconstruction is crucial for virtual and augmented reality applications in various fields, such as engineering and architecture. The conventional method used software such as 3D Studio Max, Unity 3D and AutoCAD to reconstruct the 3D model. However, this process is tedious as it requires training and skills for novices [1]. Meanwhile, the Coordinate Measuring Machine (CMM) uses a contact-based approach which interferes with the object directly to reconstruct the 3D point cloud [2]. Recently, research have been using noncontact-based methods such as image-based modelling, which is more accessible and more suitable for novices. This approach only requires sets of images captured using a standard camera sensor. In other research, the volumetric approach was used to extract the voxels from the images. This method is based on silhouette extraction, which still has an issue with concave objects [3], [4].

In previous research, feature-based methods from multi-view focuses on the 2D feature point extracted from the images. This approach is high in demand as it can reduce time complexity compared to image stitching approach in 3D reconstruction. Recently, various software used the Structure from Motion (SfM) based on feature points, such as Agisoft and VisualSfM[5],[6] to reconstruct 3D point cloud from multi-view images. However, previous methods still face accuracy issue as the input images captured might be affected by noise and outliers. Furthermore, this process requires many images to reconstruct an accurate 3D point cloud which involves optimization and iteration process that could increase complexity [7]–[9]. Generally, the feature-based framework consists of image acquisition, feature extraction and matching, camera calibration estimation and 3D model enhancement [10].

A Image Acquisition Setup

Previous 3D point cloud reconstruction process faces challenges with the input images that contain noise and outliers. Furthermore, uncertainty regarding image relation and unknown camera parameters contribute to the complexity of the process. Therefore, previous research has introduced image acquisition setup to simplify the process and to overcome the issue related with image uncertainty and unknown camera parameters [11] [12][13]. These approaches involve prior knowledge about camera parameter, which could simplify the 3D reconstruction process. The object is placed at a static position which is parallel to the camera centre. Then, a turn table is used to capture the object in rotation motion [13]. These approaches can produce 3D point cloud but are always affected by the large degree of rotation and error that could affect accuracy. Besides, there are also research that employed an automatic motor to rotate objects. However, this approach could be affected by the speed of the motor during the image-capturing process [11]. Furthermore, some previous studies also require more than one camera to capture the images, which is tedious to set up.

B Feature Extraction and Matching

Feature extraction is the process of selecting the desired feature points from the images captured. At this stage, reliable points are vital to represent the real object. Previous research [14]–[17] used Harris Corner Detection, speeded-up Robust Feature (SURF), and Scale Invariant Feature Transform (SIFT) to extract the features from images. SIFT and SURF are the most popular methods as these methods are robust in changing translation, rotation, and lighting [18][19]. Meanwhile, Good Feature to Track uses the eigenvalues method that is also robust for image rotation and translation [17]. Besides, the corresponding feature points are matched based on the similarities. Previous studies employed Fundamental Matrix (F-Matrix), Zero Mean Normalized Cross Relation (ZNCC) and clustering technique [9] to find a feature match. The challenge with the existing method is when handling huge images simultaneously. Therefore, clustering technique using a union cluster can minimize the time complexity. In other studies, feature match can also be obtained by using Lucas-Kanade (LK) Optical Flow [20] [21] [22]. This method is suitable for sparse feature point tracking that can speed up the process. Furthermore, the pyramidal Lucas Kanade manages to track the feature points from the highest to the lowest level, which can solve the large displacement between two feature points.

C Camera Calibration

Camera calibration process relates between images and estimates of camera parameters based on corresponding feature points. These parameters are essential to produce accurate 3D point cloud. The challenges in this step are when the relations between images are uncertain as the images are captured from different orientations. Thus, the self-calibration method is flexible but is always affected by accuracy and time complexity as the camera parameters are unknown [23] [24]. Previous methods rely on calibrated objects, checkerboard patterns and calibration platforms whereby it is cumbersome to estimate the unknown camera parameter such as rotation and translation [25]. Thus, camera calibration can be simplified with prior knowledge about camera parameters such as known rotation, translation, and focal length [26].

D 3D Point Cloud Reconstruction

3D point cloud is reconstructed based on feature points and camera parameters estimation obtained from previous steps. This process aims to produce a complete 3D point cloud representing the 3D model. The challenge in this process is to approximate the accurate depth based on 2D feature points. This process becomes more complex when it requires more than two images to approximate the 3D point simultaneously. Previous research used a triangulation method based on F-Matrix and Bundle Adjustment (BA) to enhance 3D point cloud reconstruction[6]. However, this process is tedious and time-consuming as it involves iteration and optimization. Furthermore, existing methods become more challenging when images are captured from various orientations and scales [27][28].

Based on preliminary research, this study proposed a framework to enhance 3D point cloud reconstruction from multi-view images based on geometric objects. The following section in this paper will describe the proposed framework, result, and conclusion.

2 Proposed Framework

The proposed framework starts with the setup for image acquisition based on viewpoints, feature extraction and matching, camera calibration from camera translation and 3D point cloud reconstruction. Figure 1 shows the proposed research framework for this study.

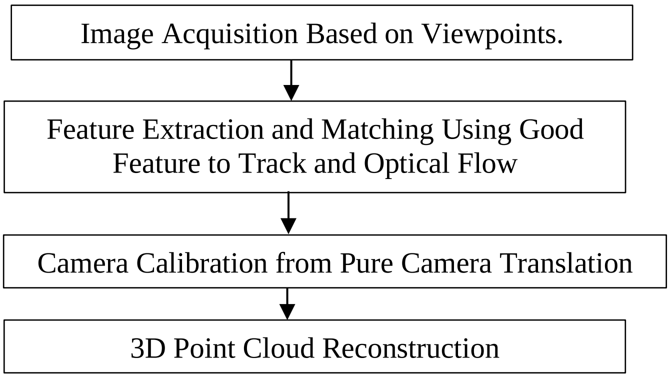


Figure 1: Proposed Framework

A Image Acquisition Based on Viewpoints

In this research, the setup to acquire sets of images is based on four viewpoints. It was labelled as front, right, left, and right viewpoints, as shown in Figure 2. For each viewpoint, two images were captured using pure camera translation. Therefore, eight images were used to reconstruct the 3D point cloud. In this setup, the object was placed at the centre of the setup, and the smartphone camera was used to capture the images. The set up also used a black background to reduce noise and to ensure only the desired features from objects are extracted. The initial process starts at viewpoint 1. It is followed by viewpoint 2, viewpoint 3 and viewpoint 4. In this research, geometric objects were used as the ground truth data obtained from the width, height, and depth measurements, as shown in Figure 3.

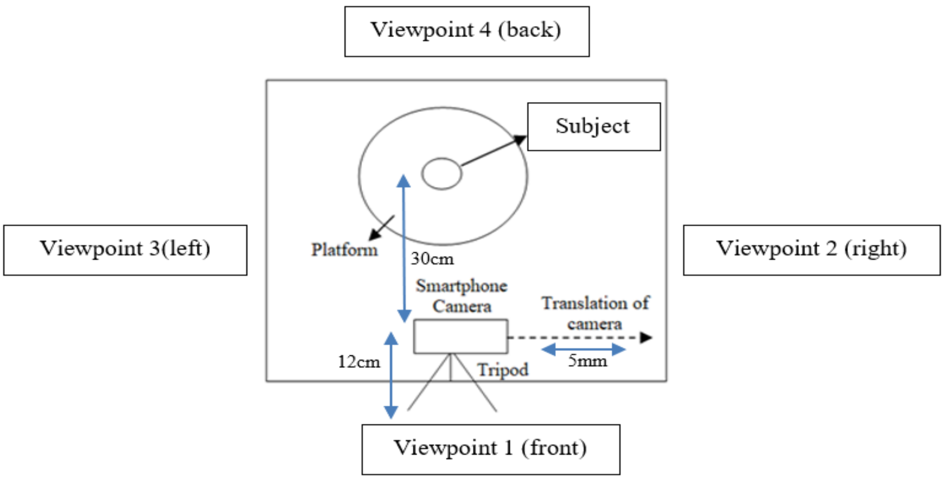


Figure 2: Image Acquisition Set Up

Subject		**1**	**2**	**3**
				
Measurement(mm)	Width	62.02	58.00	58.00
	Height	100.00	155.00	93.00
	Depth	57.50	58.00	58.00

Figure 3: Geometric Object

B Feature Extraction and Matching Using Good Features to Track and Optical Flow

The features were extracted from the sets of images using the Good feature to Track. This method improves a previous method proposed by Shi-Tomasi, which is based on corner detection [17]. It selects the corner region in the situation where the λ1 and λ2 are above the minimum value. Next, the selected features were matched with the corresponding feature points between images using Lucas Kanade's optical flow feature match. This method assumes that corresponding features have similar motion and constant brightness. The similarity can be verified using the inverse matric with a corner detector, as shown in Equation 1.

Furthermore, this approach is also based on the displacement between points. The value of *fx* and *fy* are image gradients. The vector (*u*,*v*) and magnitude obtained from this step can be used to estimate camera calibration parameters in the next step. Figure 4 shows the optical flow feature pattern when pure camera translation was used to capture the second image. The extracted 2D feature points (*x*,*y*) were used as the input in the following steps.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i {f_{x_i}}^2 & \sum_i f_{x_i} f_{y_i} \\ \sum_i f_{x_i} f_{y_i} & \sum_i {f_{y_i}}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{x_i} f_{t_i} \\ -\sum_i f_{y_i} f_{t_i} \end{bmatrix} \tag{1}$$

(a) Optical Flow Pattern	(b) Optical Flow Direction
	

Figure 4: Pure Camera Translation

C Camera Calibration from Camera Translation

Camera calibration estimates camera position and other parameters such as translation and distance. This study introduces a simple approach for camera calibration based on pure camera translation which reduces the outlier as corresponding feature points are obtained based on similar and small translations. This approach is straightforward as compared to other methods, but in some conditions, a translation error could occur. Based on Figure 5, the optical flow pattern is not straight as in the optical flow direction for pure camera translation as stated in Figure 4. Thus, the gradient analysis is used to determine the translation position either pure or error translation. Next, the centroid between the two images was rectified. The translation between two images and the distance between the object to the camera were calculated based on focal length equations which involve world and image views. The translation and distance parameters are essential for the next step to produce the 3D point cloud.

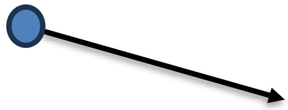
Optical Flow Pattern	Camera Direction
	

Figure 5: Camera Translation Error

D 3D Point Cloud Reconstruction

This step used perspective projection and triangulation to estimate the depth between two images based on feature points and camera parameters obtained from the previous steps. This process is based on the sketch from pure translation and image relation between world and image views from the controlled setup. The ratio between the actual object height and image height is used as the scaling factor. Next, the merging process connects all the 3D point cloud to produce a complete 3D point cloud. Since the images were captured using the controlled setup, image transformation from each viewpoint based on rotation and translation are used to merge the 3D point cloud. Figure 6 shows the depth approximation based on two images from each view and the merging 3D point cloud.

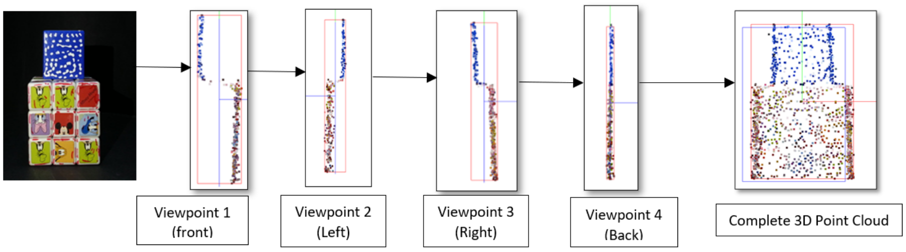


Figure 6: 3D Point Cloud Reconstruction

3 Results and Discussion

This experiment captured three sets (*N*) of images from each geometric object to validate the proposed framework. The actual measurement of the object, such as width, height, and depth value, were considered ground truth datasets. After that, the detected 3D point cloud was measured based on the boundary dimension of width, height, and depth. Root Mean Square Deviation (RMSD) calculates the difference between actual and predicted 3D point cloud, as shown in Equation 2. The result was also compared with the previous method based on known rotation angle and trigonometry equation [13]. In this experiment, the lower the RMSD value indicated, the higher the accuracy.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(Actual - Predicted)^2}{N}} \quad (2)$$

Table 1 shows the average RMSD value for the proposed method is below 1.24mm in width and 1.27mm in depth, respectively. The result indicates that the difference between the predicted 3D point cloud reconstruction based on RMSD value is low compared to the actual measurement. In this experiment, the height measurement is constant since the parameter is used as the scaling factor. Furthermore, the average of RMSD for Method 1 is 4.83mm in width and 2.83mm in depth, respectively. Based on Figure 7, the proposed method yielded the lowest RMSD for Object 3 as the object's structure is cuboid and the feature extraction becomes more accurate as the images were captured using pure camera translation approach. Besides, Object 1 yielded the highest RMSD for the proposed method due to the uneven structure of the object with the tapered end structure. Therefore, the result shows that the accuracy of the 3D point reconstructed also relies on the structure of the objects. Overall, the proposed method's accuracy is higher than Method 1 for all objects as it shows a lower average of RMSD value.

Table 1: Average RMSD for Proposed Method and Method 1

Average RMSD (mm)			
Proposed Method (Pure Translation)		Method 1 (Known Rotation)	
Width	Depth	Width	Depth
1.24	1.27	4.83	2.83

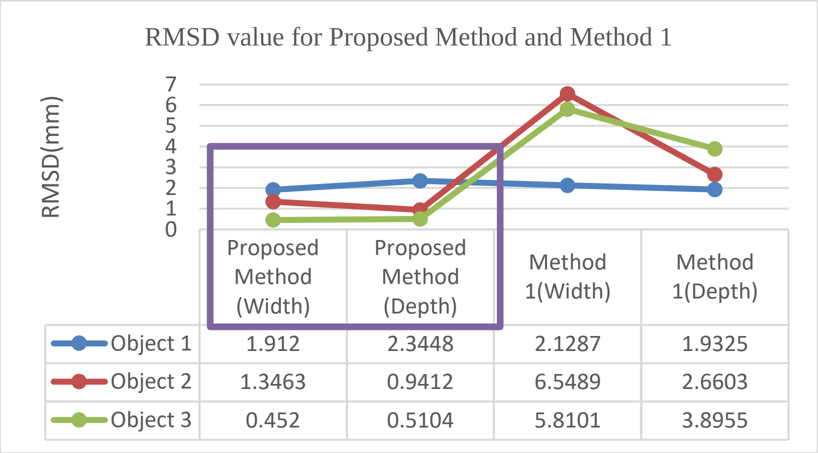


Figure 7. RMSD Comparison between Proposed Method and Method 1 for All Objects

4 Conclusion

This paper proposed a new framework to enhance the 3D point cloud reconstruction from multi-view images. It was simplified using images captured based on pure camera translation at only four viewpoints. The camera calibration process is also less complex as it only requires eight images with prior knowledge about camera translation position and optical flow feature match. Besides, the framework also utilizes perspective projection, triangulation, and image transformation equations based on images captured using simplified camera pure translation to produce a complete and accurate 3D points cloud. As a result, the proposed framework yielded an accurate 3D point cloud with the average of RMSD value for the width of 1.24mm and 1.26mm in depth for all objects. The average of RMSD value is also lower than the previous method that used a known rotation angle to produce 3D point clouds. The result shows that the proposed framework could enhance 3D point cloud accuracy for geometric objects. In the future, the framework also can be applied to complex and different types of objects. It can also be compared to the existing VSfM approach to enhance performance. Furthermore, the SURF ace of the 3D model can be constructed using this accurate 3D point cloud, and it can be applied to various virtual and augmented reality applications.

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or private sectors.

References

[1] I. S. Ramli, R. Wirza, O. K. Rahmat, N. S. Beng, P. S. Sulaiman, and H. Madzin, "Camera Rotation and Translation Recovery using Optical Flow Based Method for 3D Reconstruction from Un-Calibrated Images Article Info," *Journal of Advanced Science and Engineering Research*, vol. 7, no. 2, pp. 35–44, 2017.

[2] M. Ma, H. Shao, J. Zhang, X. Wang, and G. Li, "A Calibration Method of Anamorphic Lens Camera Based on Virtual 3D Target," In *2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2019, pp. 205–210.

[3] P.-L. Lai and A. Yilmaz, "Projective Reconstruction of Building Shape from Silhouette Images Acquired From Uncalibrated Cameras", Commission III/2." [Online]. Available: http://dpl.ceegs.ohio-state.edu/

[4] H. M. Nguyen, B. C. Wünsche, P. Delmas, C. Lutteroth, and W. van der Mark, "High resolution 3d content creation using unconstrained and uncalibrated cameras." in *2013 6th International Conference on Human System Interactions (HSI)*, 2013, pp. 637–644.

[5] S. Bianco, G. Ciocca, and D. Marelli, "Evaluating the performance of structure from motion pipelines," *J Imaging*, vol. 4, no. 8, Aug. 2018.

[6] S. Jiang, C. Jiang, and W. Jiang, "Efficient structure from motion for large-scale UAV images: A review and a comparison of SfM tools," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167. Elsevier B.V., pp. 230–251, Sep. 01, 2020.

[7] S. El Hazzat, A. Saaidi, and K. Satori, "Multi-view passive 3D reconstruction: Comparison and evaluation of three techniques and a new method for 3D object reconstruction," in *International Conference on Next Generation Networks and Services, NGNS*, IEEE Computer Society, Dec, pp. 194–201, 2014.

[8] S. Hadfield, K. Lebeda, and R. Bowden, "Stereo reconstruction using top-down cues," *Computer Vision and Image Understanding*, vol. 157, pp. 206–222, Apr. 2017.

[9] M. Cao *et al.*, "Fast and robust feature tracking for 3D reconstruction," *Opt Laser Technol*, vol. 110, pp. 120–128, Feb. 2019.

[10] Z. Ma and S. Liu, "A review of 3D reconstruction techniques in civil engineering and their applications," *Advanced Engineering Informatics*, vol. 37. Elsevier Ltd, pp. 163–174, Aug. 01, 2018.

[11] A. H. Ahmadabadian, R. Yazdan, A. Karami, M. Moradi, and F. Ghorbani, "Clustering and selecting vantage images in a low-cost system for 3D reconstruction of texture-less objects," *Measurement (Lond)*, vol. 99, pp. 185–191, Mar. 2017.

[12] Z. Cheng, H. Li, Y. Asano, Y. Zheng, and I. Sato, "Multi-view 3D Reconstruction of a Texture-less Smooth SURF ace of Unknown Generic Reflectance." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16226-16235, 2021.

[13] Ng Seng beng, "Depth Value Approximation of 2d Complex-shaped Objects for 3d Modelling Using Optical Flow and Trigonometry," Universiti Putra Malaysia, 2015.

[14] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2003.

[15] J. Tushar, S. Kulbir, and A. Aditya, “A Review and Comparison of Multi-view 3D Reconstruction Methods,” *Journal of Engineering Research*, vol. 5, no. 3, 2017.

[16] J. Sun and H. B. Gu, “Research on Camera Calibration and Object Location based on a Planar Pattern of a Rectangle [J],” *Journal of Chinese Computer Systems*, vol. 29, no. 9, pp. 1740–1744, 2008.

[17] J. Shi and C. Tomasi, “Good Features to Track,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[18] S. A. K. Tareen and Z. Saleem, “A Comparative Analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK,” in *International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1–10, 2018.

[19] Paul M., Karsh R.K, and Ahmed Talukdar F., “Image Hashing Based on Shape Context and Speeded Up Robust Features (SURF),” pp. 464–468, 2019.

[20] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski, “A Database and Evaluation Methodology for Optical Flow,” *A Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[21] K. Blachut and T. Kryjak, “Real-Time Efficient FPGA Implementation of the Multi-Scale Lucas-Kanade and Horn-Schunck Optical Flow Algorithms for a 4K Video Stream f,” *Sensors 22*, vol. 13, p. 5017, 2022.

[22] R. Rahmat and S. Azmi, “Depth Value Deduction Using Optical Flow for Reverse Engineering,” in *Proceeding of IADIS International Conference Computer Graphics, Visualization, Computer Vision, and Image Processing*, pp. 245–248, 2012.

[23] W. Li, S. Shan, and H. Liu, “High-precision method of binocular camera calibration with a distortion model.,” *Appl Opt*, vol. 56, no. 8, pp. 2368–2377, 2017.

[24] A. J. Perez, J. C. Perez-Cortes, and J. L. Guardiola, “Simple and precise multi-view camera calibration for 3D reconstruction,” *Comput Ind*, vol. 123, Dec. 2020.

[25] A. Distante and C. Distante, *Handbook of image processing and computer vision: Volume 3: From pattern to object*. Springer International Publishing, 2020.

[26] M. Yang, X. Chen, and C. Yu, “Camera calibration using a planar target with pure translation,” *Appl Opt*, vol. 58, no. 31, p. 8362, Nov. 2019.

[27] M. Alndiwee, M. M. Al-Mahairi, and R. Hamdan, “A Novel Bundle Adjustment Approach based on Guess-aided and Angle quantization Multi-Objective Particle Swarm Optimization (GAMOPSO) for 3D reconstruction applications.,” *IEEE Access*, 2022.

[28] Q. Demoulin, F. Lefebvre-Albaret, A. Basarab, D. Kouamé, and J.-Y. Tourneret, “Constrained Bundle Adjustment Applied to Wing 3d Reconstruction with Mechanical Limitations,” 2020.