

Fast 3D Reconstruction of Power Transmission Lines Based on Sequential Images

Xinyan Qin
 College of Mechanical and Electrical Engineering
 Shihezi University
 Shihezi, China
 qinx@shzu.edu.cn

Yujie Zeng
 College of Mechanical and Electrical Engineering
 Shihezi University
 Shihezi, China
 20212009014@stu.shzu.edu.cn

Jin Lei*
 College of Mechanical and Electrical Engineering
 Shihezi University
 Shihezi, China
 jinlei@shzu.edu.cn

Bo Li
 College of Mechanical and Electrical Engineering
 Shihezi University
 Shihezi, China
 20202109065@stu.shzu.edu.cn

Jie Zhang
 College of Mechanical and Electrical Engineering
 Shihezi University
 Shihezi, China
 zhangjie77@stu.shzu.edu.cn

Abstract—The 3D reconstruction can visualize the inspection scene of power transmission lines (PTLs), providing data support for PTL intelligent inspection. To address the low-efficiency and large-computation of the 3D reconstruction methods for PTL, this paper proposes an efficient reconstruction method based on the sequential images of PTLs. Firstly, collect multi-view sequential images of PTLs by the Flying-walking Power Line Inspection Robot (FPLIR). Secondly, obtain the initial sparse point clouds data by the Structure from Motion (SFM) algorithm. Thirdly, remove the redundant point clouds based on the camera trajectories to acquire the point clouds data of PTLs. Lastly, obtain the dense point clouds of PTLs data and spatial surface model by the Multiple View Stereo (MVS) algorithm. The experimental results show that our proposed method allows processing 70 images within 48 minutes, obtaining the PTL model with a mean reprojection error of about 1.2 pixels. The proposed method provides the foundation for a refined and intelligent inspection of PTL.

Keywords—Fast 3D Reconstruction, Sequential images, PTL, SFM, MVS

I. INTRODUCTION

PTL is the infrastructure of the country and the main carrier of long-distance high-voltage power transmission, which is vital for people's daily life and economic development. However, PTL is affected by sandstorms, temperature, vegetation growth, etc. Therefore, it needs regular inspections and maintenance. The existing PTL inspection methods mainly include manual inspection and robot inspection. Compared with manual inspection, robot inspection is safer and more efficient. Besides, building a digital transmission operation and maintenance system is the primary task for the current power supply department. And 3D reconstruction is a major channel for realizing the digital system. The obtained 3D data, on the one hand, can be used for predicting the distance between objects [1], planning and obstacle avoidance in online inspection [2], etc.;

on the other hand, it can provide data support for the selection, construction management, and maintenance of PTL [3].

These methods currently used for the 3D reconstruction of PTLs are mainly active and passive. The active method is represented by LiDAR, which has high data acquisition efficiency, high modeling quality, high measurement accuracy, and low weather requirements. But it is expensive and unsuitable for large-scale applications. The passive method is represented by vision, which has low-cost, good-flexibility, 3D visualization effect. But compared to the active method, it has the disadvantage of lower model accuracy [4, 5]. Therefore, to promote the 3D reconstruction technology of PTLs for large-scale applications, we choose the image-based 3D reconstruction method.

Due to the slender, evenly distributed, and repetitive texture of the power line structure, it is difficult for feature extraction and matching [1, 6, 7]. The background of PTL is mainly sky, which is an unbounded scene causing reconstructed vignettes of distant objects [8, 9]. These reasons cause difficulties in reconstructing PTLs using disordered images. The current methods for reconstructing PTLs using disordered images always firstly segment structures to extract PTLs, and then reconstruct them [6, 8, 10, 11]. Such methods are complex and time-consuming.

To solve the above problems, we propose a fast 3D reconstruction method for PTL based on sequential images. The proposed method mainly includes five steps:

- 1) Acquire multi-view sequential images of PTLs by FPLIR.
- 2) Calculate the initial sparse point clouds in the captured scene from acquired sequential images, while calculating the internal camera parameters.
- 3) Remove redundant points, obtaining sparse point clouds of PTLs.

*Corresponding author

4) Estimate the depth map and densify the reconstruction of PTLs.

5) Obtain the final digital model of PTLs by segmenting the mesh, optimizing the dense point clouds, mapping the texture, and repairing the surface joints.

II. SYSTEM OVERVIEW

A. MVS Method

The MVS method is based on the parallax principle, also known as the triangulation principle. It uses a single camera or multiple cameras to capture objects from multiple views, recovering the geometric information of the spatial objects based on the captured multiple 2D images.

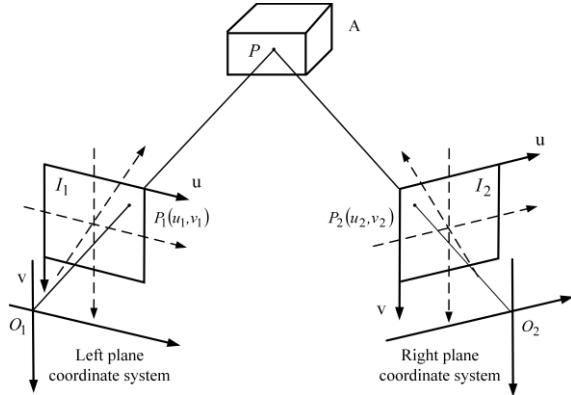


Fig. 1. MVS working principle.

As shown in Fig. 1, the images I_1 and I_2 are captured from two positions O_1 and O_2 of object A. It is assumed that any point P on the surface of object A is projected as point $P_1(u_1, v_1)$ and $P_2(u_2, v_2)$ on the images I_1 and I_2 respectively. M_1 and M_2 are the projection matrix of I_1 and I_2 . The spatial coordinates of point P can be solved by 3D coordinate transformation expressed as follows:

$$Z_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \begin{bmatrix} M_{11} \\ M_{12} \\ M_{13} \end{bmatrix} X \quad (1)$$

$$\begin{bmatrix} M_{13}u_1 - M_{11} \\ M_{13}v_1 - M_{12} \end{bmatrix} X = 0 \quad (2)$$

where X is the homogeneous coordinates in the world frame. Similarly, the projection matrix M_2 can be obtained by:

$$\begin{bmatrix} M_{23}u_2 - M_{21} \\ M_{23}v_2 - M_{22} \end{bmatrix} X = 0 \quad (3)$$

According to (2) and (3), then get:

$$\begin{bmatrix} M_{13}u_1 - M_{11} \\ M_{13}v_1 - M_{12} \\ M_{23}u_2 - M_{21} \\ M_{23}v_2 - M_{22} \end{bmatrix} X = 0 \quad (4)$$

The spatial coordinate of point P can be obtained by solving (4). However, in practical applications, since the data always includes noises, the least squares method is used to calculate the coordinates of point P .

B. Sequence Image Acquisition Platform

The vision acquisition system in this paper consists of a camera, a tripod head, a developed FPLIR, and an onboard computer, as shown in Fig. 2. The inspection process of FPLIR includes taking off, approaching the ground line, landing on the ground line, rolling along the ground line, crossing obstacles, flying over the tower head, leaving the ground line, and landing on the ground. The data can be acquired during the FPLIR inspection of the PTL. The inspection process of FPLIR is shown in Fig. 3.

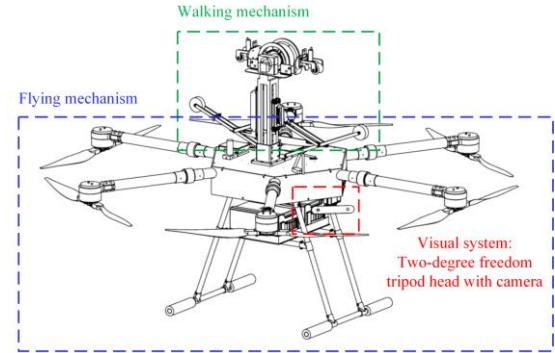


Fig. 2. Schematic diagram of FPLIR.

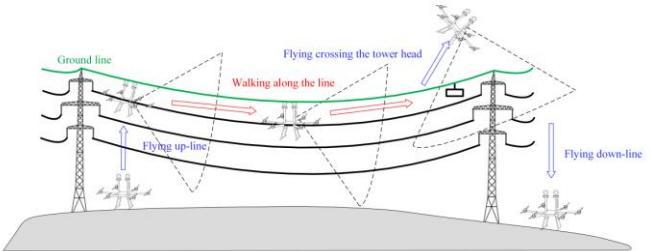


Fig. 3. Inspection process of FPLIR.

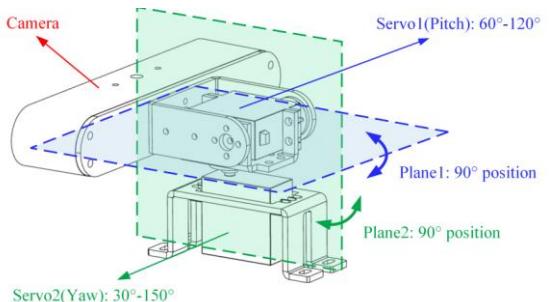


Fig. 4. Schematic diagram of the visual system.

Multi-view images are vital for 3D reconstruction, but PFLIR can only obtain single-view images while rolling along the ground line. To acquire multi-view sequential images of PTLs, we use a tripod head with two degrees of freedom (DOF) to control camera motion. The tripod head controls the motion of the camera in pitch and yaw from 60° to 120° and from 30° to 150° , respectively. These angles are distributed symmetrically according to the shown position, as shown in Fig. 4. The main parameters of the camera and the tripod head are listed in Table 1.

TABLE I. THE MAIN PARAMETERS OF THE CAMERA AND THE TRIPOD HEAD

Device	Angle of view (°)
Camera	110° (H) $\times 70^\circ$ (V) $\times 120^\circ$ (D)
Servo1 (pitch)	From 60° to 120°
Servo2 (yaw)	From 30° to 150°

C. Algorithm Flow

In this paper, the 3D reconstruction method of PTL based on sequential images includes the following steps: (1) Acquire multi-view sequential images of PTLs during rolling along the ground line; (2) Extract and match features by the sequential images; (3) Use the SFM algorithm to calculate the initial sparse point cloud information, recovering the camera's poses; (4) Remove the redundant parts of the initial sparse point clouds, obtaining the sparse point clouds of PTLs; (5) Calculate the depth information from the obtained camera poses and PTLs sparse point clouds by using the MVS algorithm, as well as do dense reconstruction; (6) Reconstruct the mesh, optimizing the mesh surface reconstruction; (7) Perform texture mapping, obtaining the final visualization model of PTLs. The overall algorithm flow of the method is shown in Fig. 5.

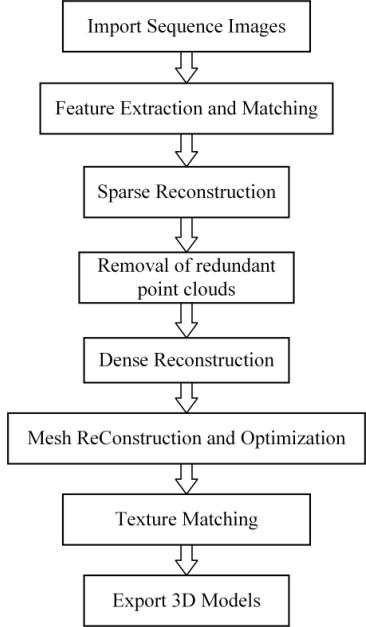


Fig. 5. Overall algorithm flow of our method.

III. POINT CLOUD ACQUISITION AND PROCESSING

Firstly, we select the Scale Invariant Feature Transform (SIFT) algorithm [12] for feature extraction. Secondly, the Random Sample Consensus (RANSAC) [13] is used to eliminate false match points. Thirdly, the initial 3D point cloud is obtained using the SFM, while the Bundle Adjustment (BA) method is used to adjust the position of all beams and the camera. Finally, the redundant points are removed from the outputting sparse point clouds. The processing flowchart is shown in Fig. 6.

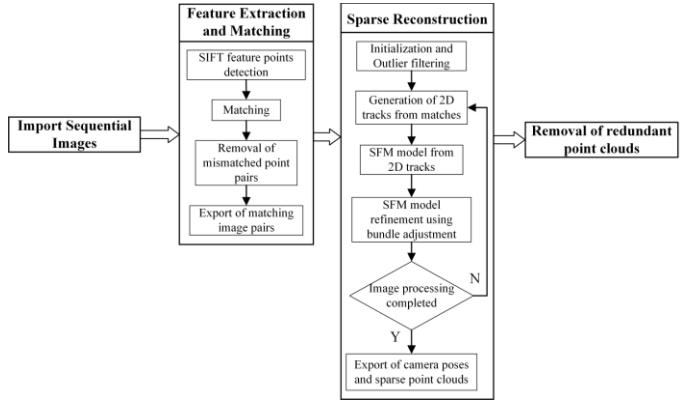


Fig. 6. Flowchart of point cloud acquisition and processing.

A. Image Feature Extraction and Matching

Since the surface of PTL has a metallic luster or is pure black, it causes difficulties extracting and matching the line feature points of PTLs in images. The SIFT descriptor is light-invariant, perspective-invariant, perspective-invariant and rotation-invariant, improving the reflection problem effectively. Therefore, we select SIFT descriptor to extract features, increasing the number of feature points of PTLs. In addition, the SIFT descriptor is distinctive, informative, and efficient. In SIFT algorithm, Gaussian difference and image convolution are used for feature points detection as follows:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (5) \\ = L(x, y, k\sigma) - L(x, y, \sigma)$$

where L is the image scale space after convolution. $I(x, y)$ is the pixel value of the image at position (x, y) . The two-dimensional Gaussian kernel function $G(x, y, \sigma)$ is expressed as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} = 0 \quad (6)$$

where σ is the mean squared deviation of the Gaussian distribution.

Since the matching of SIFT descriptors is only based on appearance without considering the geometric relationship between images, it cannot guarantee that the counterpart feature points can be mapped to the same 3D field point. We use the RANSAC method and geometric relations to eliminate mismatched points, computing a reliable basis matrix and

single-strain matrix between images. Due to the fact that the input is sequential images of PTLs, the matching rule of images is set to sequential matching, which saves the matching time and reduces the false matching rate. The effect of feature detection is shown in Fig. 7.

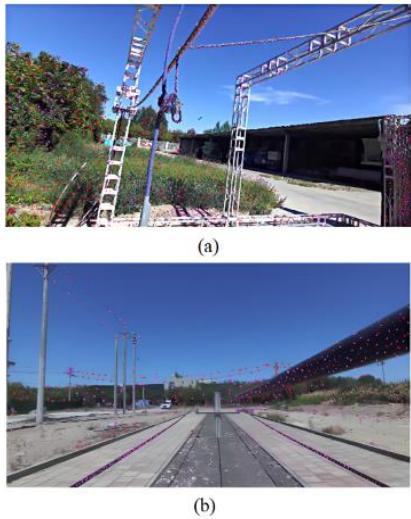


Fig. 7. The effect of feature detection (some purple or red points are the detected feature).

B. Sparse Reconstruction

According to the triangulation principle, the coordinates of the spatial points can be obtained from multiple images. The solution of the spatial point $P(X, Y, Z)^T$ at the corresponding image coordinate $P(u, v)^T$ belongs to Perspective-n-Point (PnP) problem. Therefore, the camera internal K and external $[R, T]$ parameters can be calculated by the following:

$$s \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = K[R, T] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (7)$$

$$= \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} t_1^T \\ t_2^T \\ t_3^T \\ 1 \end{pmatrix} P$$

Equation (7) can be simplified as:

$$P^T t_1 - P^T t_3 u_1 = 0 \quad (8)$$

$$P^T t_2 - P^T t_3 u_1 = 0 \quad (9)$$

Assuming that there are N feature points, a system of linear equations can be listed as follows:

$$\begin{pmatrix} P_1^T & 0 & -u_1 P_1^T \\ 0 & P_1^T & -v_1 P_1^T \\ \vdots & \vdots & \vdots \\ P_N^T & 0 & -u_1 P_N^T \\ 0 & P_N^T & -v_1 P_N^T \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 0 \quad (10)$$

Our image data are all captured by the same camera with known internal parameters, improving the accuracy and reducing the time consumption in the computational phase. The 3D coordinates used in (7) are some estimated numbers that are different from the actually observed ones. Therefore, the 3D coordinates of feature points and camera pose need the BA optimization to minimize the errors. For n images and m feature points, the objective function of BA is as follows:

$$\min_{P_i, M_j, i=1, j=1} \sum_{i=1}^n \sum_{j=1}^m D(m_{ij}, M_j P_i)^2 \quad (11)$$

where M_j is the projection matrix of the frame i image, and D is the Euclidean distance.

We use the Levenberg-Marquardt (LM) method to optimize for the minimum error. It combines the Steepest Descent method and the Gauss-Newton method, which has the characteristics of fast convergence, not easy to fall into local minimum extrema, and good stability. The initial sparse point clouds are shown in Fig. 8. The rough outlines of the experimental frame and PTLs can be seen.

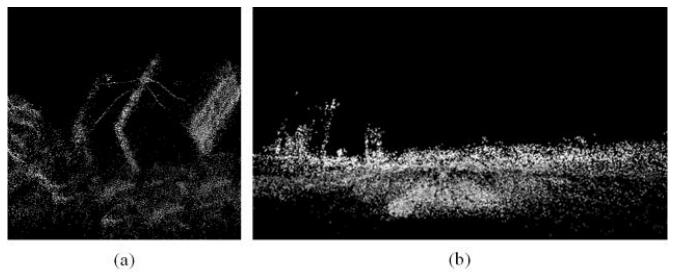


Fig. 8. Initial sparse point clouds.

C. Removal of Redundant Points

Many redundant points exist in the built sparse point cloud based on sequential images, increasing the processing-storage space, reducing operation rate, and wasting computational resources. FPLIR rolls along the ground line. Therefore, the camera trajectories obtained can be used to remove redundant points.

We construct sparse point clouds of the PTLs by adding camera trajectories to the initial sparse point clouds. Suppose there are N points in the original sparse point clouds, and any point of them is $P_i(a, b, c)$. Based on the motion trajectories of the camera, the constraints threshold in the x_r, y_r, z_r directions are $\sigma_{x_{r_1}}, \sigma_{x_{r_2}}, \sigma_{y_{r_1}}, \sigma_{y_{r_2}}, \sigma_{z_{r_1}}$ and $\sigma_{z_{r_2}}$, respectively. The threshold value is determined according to the actual needs. The region of interest can be calculated as follows:

$$\begin{cases} x_{r_{\min}} - \sigma_{x_{r_1}} \leq a \leq x_{r_{\max}} + \sigma_{x_{r_2}} \\ \text{mean}(y_r) - \sigma_{y_{r_1}} \leq b \leq \text{mean}(y_r) + \sigma_{y_{r_2}} \\ \text{mean}(z_r) - \sigma_{z_{r_1}} \leq c \leq \text{mean}(z_r) + \sigma_{z_{r_2}} \end{cases} \quad (12)$$

where $\text{mean}(y_r)$ is the mean value of y coordinate; $\text{mean}(z_r)$ is the mean value of z_r coordinate; $x_{r_{\min}}$ is the minimum value of x_r coordinate, and $x_{r_{\max}}$ is the maximum value of x_r coordinate. The points in the sparse point clouds dissatisfied (12) will be removed. The results are shown in Fig. 9. The outlines of the experimental frame and PTLs are more definite than Fig. 8.

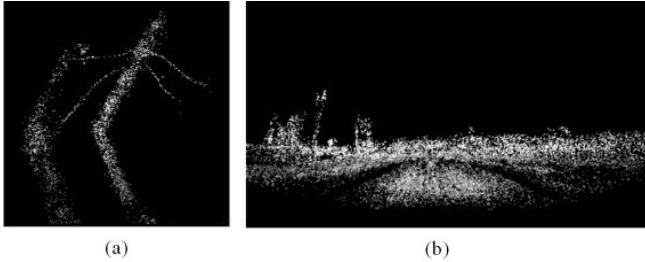


Fig. 9. Sparse point clouds after removal.

IV. DENSE RECONSTRUCTION AND POINT CLOUDS MODELIZATION

A. Depth Map Estimation

Depth map estimation is crucial for dense reconstruction, and can usually be solved using semi-global matching (SGM/tSGM) [14] or PatchMatch [15] proposed by C. Barnes. Compared with these two methods, the SGM/tSGM uses the classical Fronto-parallel window model. The PatchMatch uses the new Slanted Support Window model to calculate sub-pixel parallax, handle slanted surfaces, and better fit surfaces. The PatchMatch method is more effective for PTLs reconstruction. Therefore, we use the PatchMatch method to estimate the depth map.

B. Mesh Reconstruction and Optimization

Due to the small number of PTLs sparse point clouds, the method based on point clouds fitting (e.g., Poisson reconstruction) is less effective in reconstruction. To effectively recover the surface information of PTLs, mesh segmentation is required before the surface reconstruction. This paper chooses the tetrahedral mesh segmentation. The method divides the convex packets of the point clouds into tetrahedra whose vertices are all the point clouds. The weight of the tetrahedron is calculated from the visual information of each point, using the minimum cut method [16] to find the partition surface with the minimum weight.

C. Texture Matching

Texture mapping calculates the correspondence between grid vertices and image pixels. The Loop belief propagation (LBP) algorithm solves the optimal view of each surface by computing the Markov Random Field (MRF) energy function. The LBP is a message-passing algorithm. A node transmits a message to an adjacent node only after receiving all incoming messages. The label state of the current MRF is updated by

passing information between nodes, obtaining the optimal view of each surface. Since the images of the same object have different lightness and darkness from different viewpoints, the camera position estimation will deviate, leading to gaps in the surface connection. The color correction includes global and local methods. For the same patch, the color correction values in the texture region covered for each face are obtained by interpolating the correction values of the three vertices. The correction values are applied to all patch texture matches, finishing the global color correction. However, global fusion does not eliminate the local seams. Poisson fusion [17] is used near the seam to correct the local color, allowing a natural transition between the surfaces.

V. EXPERIMENTS

A. Test Data Sets

We collected two data sets at the test field and the power supply company, which have 71 and 68 RGB images, respectively. The data set acquisition uses a calibrated image acquisition platform with every 6° a frame and a resolution of 1280×720 . Fig. 10 (a) and (b) show some images of the two data sets.

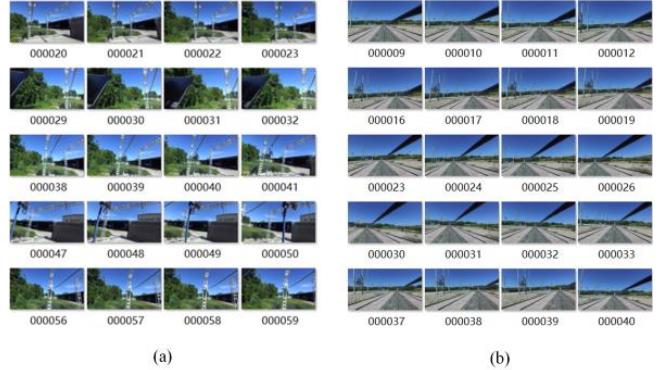


Fig. 10. Some sequential images of PTLs (a) Data Set I; (b) Data Set II.

B. Results and Analysis

The experiments were conducted on a computer with an I9-12900K CPU, a 3090 GPU, and an operating system of Ubuntu 18.04.

The main parameters are listed in Table 2 for the 3D reconstruction of two data sets. The accuracy of PTLs reconstruction can be indirectly reflected by the number of points, the number of observations, and the mean track length. In two trials, the mean reprojection error and reconstruction time were less than 1.22 pixels and 50 minutes, respectively, which satisfied our expectation of speediness.

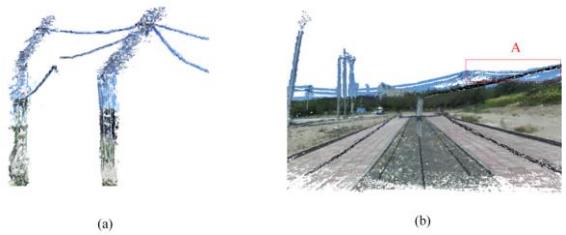


Fig. 11. Point clouds after dense reconstruction (a) Data Set I; (b) Data Set II.

Fig. 11 shows the point clouds effect after dense reconstruction. Due to the motion trajectories of the camera cannot be used, the detailed information of some towers and PTLs in the distance is lost, which causes the edge blurring in area A of Fig. 11(b). The result shows that our method only can well-reconstruct PTLs at close distances. However, the reconstruction capability needs to be improved for PTLs at far distances without camera motion trajectories.

TABLE II. THE MAIN PARAMETERS OF THE CAMERA AND THE TRIPOD HEAD

Parameters	Data Set I	Data Set II
Images	71	68
Points	24176	22462
Observations	195818	160836
Mean track length	8.09969 m	7.16036 m
Mean reprojection error	1.21607 pix	1.2158 pix
Reconstruction Time	49.055 min	46.501 min

Applying texture mapping to the Data Set II, the obtained 3D model was shown in Fig. 12.



Fig. 12. final 3d reconstruction model of Data Set II.

VI. CONCLUSION

This paper proposes an efficient reconstruction method using sequential images of PTLs collected by the developed FPLIR. The method can be used for large-scale 3D reconstruction of PTLs, improving the integrity and efficiency of reconstruction. Firstly, the SIFT algorithm is adopted for feature extraction based on the characteristics of PTL. Secondly, the motion trajectories of the camera are used to remove the redundant points in the sparse point clouds, improving the reconstruction efficiency. Finally, the PatchMatch method calculates the depth map, and its new Slanted Support Windows model is used to fit the power line surfaces better. The experiment shows that our method can quickly reconstruct PTLs. It processed and calculated 70 images within 50 minutes. And the average reprojection error of the obtained model is around 1.2 pixels. The reconstruction results meet the expected requirements in speed and accuracy.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (grant nos. 62063030 and 62163032), the

Financial Science and Technology Program of the XPCC (grant nos. 2021DB003, 2022CB011, and 2022CB002-07), and the High-Level Talent Project of Shihezi University (grant nos. RCZK2018C31 and RCZK2018C32).

REFERENCES

- [1] E. Pastucha, E. Punia, A. Ścisłowicz, P. Ćwiąkała, W. Niewiem, and P. Więcek, "3D reconstruction of power lines using uav images to monitor corridor clearance," *Remote Sensing*, vol. 12, no. 22, pp. 2072-4292, 2020.
- [2] C. Shuai, H. Wang, W. Zhang, P. Yao, and Y. Qin, "Binocular vision perception and obstacle avoidance of visual simulation system for power lines inspection with UAV," 2017: IEEE, pp. 10480-10485.
- [3] J. S. Álvares, D. B. Costa, and R. R. S. de Melo, "Exploratory study of using unmanned aerial system imagery for construction site 3D mapping," *Construction Innovation*, pp. 1471-4175, 2018.
- [4] Y. Hao, Y. C. Tan, V. C. Tai, X. D. Zhang, E. P. Wei, and S. C. Ng, "Review of key technologies for warehouse 3D reconstruction," *Journal of Mechanical Engineering and Sciences*, vol. 16, no. 3, pp. 9142-9156, 2022.
- [5] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects," *Sensors*, vol. 13, no. 8, pp. 11007-11031, 2013.
- [6] W. Huang, S. Jiang, S. He, and W. Jiang, "Accelerated Multi-View Stereo for 3D Reconstruction of Transmission Corridor with Fine-Scale Power Line," *Remote Sensing*, vol. 13, no. 20, p. 4097, 2021.
- [7] Z. Haopeng, W. Quanmao, Z. Wei, W. Junfeng, and J. Zhiguo, "Sequential-image-based space object 3D reconstruction," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 42, no. 2, pp. 273-279, 2016.
- [8] F. Ganovelli, L. Malomo, and R. Scopigno, "Reconstructing power lines from images," in 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), 2018: IEEE, pp. 1-6.
- [9] M. Tancik et al., "Block-nerf: Scalable large scene neural view synthesis," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 8248-8258.
- [10] X. Li, F. Zhang, G. Li, D. Xiang, and C. Yang, "Research on 3D Reconstruction Technology of Power Line Based on Image Semantic Segmentation," in *Journal of Physics: Conference Series*, 2021, vol. 2095, no. 1: IOP Publishing, p. 012015.
- [11] S. Jiang and W. Jiang, "UAV-based oblique photogrammetry for 3D reconstruction of transmission line: Practices and applications," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 401-406, 2019.
- [12] G. Lowe, "Sift-the-scale invariant feature transform," *Int. J.*, vol. 2, no. 91-110, p. 2, 2004.
- [13] D. A. Forsyth and J. Ponce, *Computer vision: a modern approach*. prentice hall professional technical reference, 2002.
- [14] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, vol. 2: IEEE, pp. 807-814.
- [15] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.
- [16] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images," in Proceedings eighth IEEE international conference on computer vision. ICCV 2001, 2001, vol. 1: IEEE, pp. 105-112.
- [17] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in ACM SIGGRAPH 2003 Papers, 2003, pp. 313-318.