

A 3D Reconstruction and Relocalization Method for Humanoid Welding Robots

Peng Chi , Zhenmin Wang , Haipeng Liao , Ting Li , and Qin Zhang

Abstract—Welding robots represent pivotal equipment in intelligent welding for manufacturing and maintenance. Presently, most welding robots are stationary single-arm units, exhibiting limited flexibility and efficiency, thereby compromising welding quality and productivity. Consequently, there is an urgent need to develop a new generation of humanoid welding robots (HWR) endowed with autonomous mobility and dual-arm collaborative capabilities. Key to this advancement are pose estimation and three-dimensional (3D) reconstruction methods, which traditionally focus on mapping and navigating unfamiliar environments, often struggling to adapt to the routine welding and maintenance scenes of large-scale equipment. This paper introduces a novel approach to 3D reconstruction and relocalization tailored for HWR, facilitating rapid localization of welding areas and transmission of point cloud maps. Initially, a vision-based 3D reconstruction system is proposed, encompassing pose estimation, 3D reconstruction, and target detection, enabling self-localization and precise targeting for HWR. Subsequently, a novel method for 3D point cloud map segmentation based on 2D features and 3D point clouds matching is introduced to expedite the transmission of point cloud maps. Finally, a relocalization and point cloud map updating method grounded in prior knowledge is proposed, facilitating seamless welding operations by HWR in routine maintenance scenes. The effectiveness and superiority of the proposed methodology are validated through comparative tests with existing methods using actual HWR.

Index Terms—3D reconstruction, simultaneous localization and mapping, relocalization, humanoid welding robots.

I. INTRODUCTION

In manufacturing, welding is a crucial process in sectors such as aerospace, automotive, shipbuilding, construction, and electronics [1], [2]. However, manual welding is plagued by inefficiencies, inconsistent quality, and hazardous conditions, exacerbating the shortage of skilled welders. Researchers are thus focusing on advancing intelligent welding manufacturing and automated maintenance through welding robots [3], [4], [5],

Received 16 November 2024; accepted 18 February 2025. Date of publication 24 February 2025; date of current version 14 March 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFB3407703, in part by the National Natural Science Foundation of China under Grant U23A20625, Grant U2141216, and Grant 52375334, in part by the Natural Science Foundation of Guangdong Province under Grant 2023B1515250003 and Grant 2023A1515012112, and in part by Shenzhen Science and Technology Program under Grant KJZD20230923114614029. The associate editor coordinating the review of this article and approving it for publication was Dr. Thomas Maugey. (*Corresponding author: Qin Zhang.*)

Peng Chi, Zhenmin Wang, Haipeng Liao, and Ting Li are with the School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: mechipeng@mail.scut.edu.cn; wangznm@scut.edu.cn; haipengliao_scut@163.com; melt@mail.scut.edu.cn).

Qin Zhang is with the School of Computer Science & Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: cszhangq@scut.edu.cn).

Digital Object Identifier 10.1109/LSP.2025.3544967

[6]. Most current welding robots are fixed-position, single-arm systems, which lack flexibility and automation, limiting their ability to handle complex, small-scale, and customized welding tasks [2], [4]. With the rise of humanoid robotics, dual-arm robots offer a promising solution to enhance productivity and flexibility [7], [8]. Furthermore, traditional robots struggle with autonomous maintenance and replacing humans in hazardous environments, making humanoid welding robots (HWR) with dual arms and mobility a key future direction [9].

HWR, as mobile robots, rely on 3D reconstruction and Simultaneous Localization and Mapping (SLAM) to enable real-time environmental feedback, enhancing welding accuracy and efficiency [3], [10]. 3D reconstruction methods include Structure from Motion (SfM) [11], [12], Multiple View Stereo (MVS) [13], [14], Neural Radiance Fields (NeRF) [15], [16], and 3D Gaussian Splatting (3DGS) [17], [18]. While SfM, MVS, and NeRF offer high accuracy, they are computationally expensive and require high-quality input images. In contrast, 3DGS, based on initial point cloud data, is still underdeveloped, limiting its use in welding robots [19], [20], [21]. Currently, welding robots often rely on structured light cameras and traditional methods [22], [23], which face challenges such as high cost, complexity, and scalability issues.

SLAM methods can generate sparse or dense maps. Sparse maps use feature points for 3D coordinates, whereas dense maps create comprehensive models of all visible surfaces and objects, useful for capturing detailed environmental geometry [24], [25], [26], [27]. Dense maps, while accurate, generate large amounts of data, making them difficult to integrate with human intervention in welding tasks, thereby reducing flexibility and efficiency.

Despite significant progress in exploring unknown environments, existing research often overlooks the relocalization and map updating needed for routine maintenance tasks [28], [29]. Given the complexity of welding environments and the prevalence of manual intervention, this paper focuses on vision-based 3D reconstruction and relocalization techniques tailored to HWR applications. The contributions are as follows:

- A vision-based 3D reconstruction system encompassing pose estimation, 3D reconstruction, and target detection is proposed, which can achieve self-positioning and precise target localization for HWR.
- A novel 3D point cloud map segmentation method based on 2D features and 3D point clouds matching is proposed to expedite the transmission of 3D point cloud maps.
- A relocalization and point cloud map update method grounded in prior knowledge is proposed, which can enable HWR to conduct welding operations seamlessly within routine maintenance scenes.

This manuscript extends our previous conference paper [29], with Sections II-D and II-E drawing from the original work.

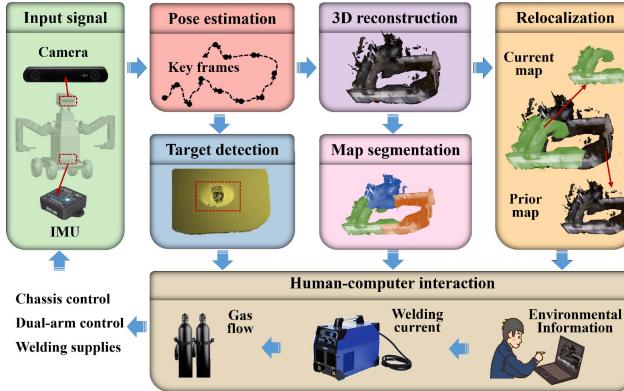


Fig. 1. Overview of the proposed 3D reconstruction and relocalization method for HWR.

Building on the previous framework, we replace expensive LIDAR with RGB-D cameras and introduce a vision-based detection method for welding areas. Additionally, the original 3D point cloud-based local map matching strategy has been revised to a fusion approach that integrates 2D features with 3D point clouds, significantly improving segmentation accuracy.

II. METHODOLOGY

A. Overview of the Method

This section presents a 3D reconstruction and relocalization method for HWR, designed for repetitive tasks with human intervention. As shown in Fig. 1, the method utilizes IMU and camera data (color images and depth information) to used as the input signals. Pose estimation is then applied to determine the positions of the HWR. Due to the low movement speed of the HWR, key frame image poses and IMU data are stored throughout the process. During a single movement, these key frames are used for target detection, 3D reconstruction, and map segmentation, improving both estimation accuracy and map transmission speed. The key frame poses also aid in the relocalization process, reducing pose drift and initialization errors in repetitive scenarios. Additionally, human intervention controls are integrated for welding current, gas flow, chassis, and dual-arm management.

B. Pose Estimation and 3D Reconstruction

Current SLAM methods primarily rely on sparse mapping techniques, such as image feature point matching and tracking, with ORB-SLAM3 [30] being a notable example. The extrinsic parameters linking the camera and IMU data are denoted as $R_{\text{camera}}^{\text{imu}}$ and $T_{\text{camera}}^{\text{imu}}$. The initial IMU frame establishes the world coordinate system, serving as the origin for the point cloud map and pose estimation, denoted as Θ^{world} . To address IMU data drift, various filtering techniques, such as the multiplicative extended Kalman filter (MEKF) [31], have been explored to improve data quality and reliability. IMU state propagation is then used to estimate the state at time t , where the state is defined as:

$$S_t = [R_t \ T_t \ V_t \ b_t] \quad (1)$$

where S_t is the state at time t , R_t and T_t are the rotation and translation parameters relative to the world coordinate system Θ^{world} at time t , V_t represents the velocity and b_t is the bias of IMU.

Key frames from pose estimation are utilized for 3D reconstruction, which are identified based on a pose difference threshold of 0.2 m or 0.05 rad , validated through practical experimentation. Given that pose estimation operates within the IMU coordinate system, the pose of image frames necessitates temporal interpolation and consideration of $R_{\text{camera}}^{\text{imu}}$ and $T_{\text{camera}}^{\text{imu}}$. The interpolation method is defined as:

$$\begin{aligned} T_k &= T_i + (k - i)(T_j - T_i)/(j - i) \\ R_k &= R_i + (k - i)(R_j - R_i)/(j - i) \end{aligned} \quad (2)$$

where T_k and R_k are the pose of the image frames at time k ($k \in [i, j]$).

Based on the depth map of the RGB-D camera, the 3D coordinates corresponding to the pixel points on the color map can be calculated:

$$\frac{d}{s_d} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

where (X, Y, Z) is the coordinates in the image coordinate system. s_d is the depth scaling factor. (u_0, v_0) the image center in the pixel coordinate system. f_x, f_y are the inverses of the physical dimensions of each pixel in the x and y directions of the image plane. d is the corresponding depth value. u, v are the pixel value in the image.

Assume feature points $p_m = (u_m, v_m)$ and $p_n = (u_n, v_n)$ in key frames K_m and K_n , these points are mapped to the normalized coordinate system of the camera using (3). Then the 3D coordinates can be associated through the triangulation:

$$P_i = \text{Triangulate}(p_m, p_n, R_n^m, T_n^m) \quad (4)$$

where $P_i = (X_i, Y_i, Z_i)$ is 3D coordinate of feature points. R_n^m, T_n^m are the rotation and translation parameters of K_m and K_n .

The relative poses of all key frames and the poses of the 3D points are globally optimized by minimizing the following error function:

$$\min_{P_i, R_i^j, T_i^j} \sum_{i,j} \|z_{ij} - \psi(P_i, R_i^j, T_i^j)\|^2 \quad (5)$$

where $z = (f_x \cdot X/Z, f_y \cdot Y/Z)$ is the observation value projected onto the pixel plane by the key frame K_i . $\psi(P_i, R_i^j, T_i^j)$ is the predicted projection of the 3D point P_i onto key frame K_j .

The global 3D point cloud map is constructed using the spatial position of key frames K_i in the Θ^{world} and the corresponding 3D coordinates P_i of pixels within the images.

C. Target Detection

In this study, YOLOv10 [32], recognized as a state-of-the-art method, is adopted and retrained specifically as a detection model for welding regions. However, given the intricate and varied nature of welding environments, our empirical findings indicate that automated operations cannot consistently ensure

accuracy, necessitating periodic human intervention. The delineation of the welding area for detection $R_t^{welding}$ and $T_t^{welding}$ involves transforming the results obtained in the camera coordinate system to the IMU coordinate system using R_{camera}^{imu} and T_{camera}^{imu} . Subsequently, these coordinates are mapped to the Θ_{world}^{camera} based on the outcomes of pose estimation.

D. Map Segmentation

Traditional SLAM methods maintain a global map, which becomes computationally expensive as the number of data points increases. To address this, we transition from a global map to dynamically loading local maps via key frame matching. Map segmentation is based on the total number of points in the global map and the distance traveled by the robot. Our experiments show that segmenting the map after robot movements of approximately 15–20 m optimizes performance.

To mitigate potential frame loss during segmentation, we use a fusion method that combines 2D feature matching with 3D point cloud matching, ensuring accurate pose estimation across adjacent local maps. Specifically, the segmentation process involves matching ORB keypoints between the last frame before segmentation and the first frame of the new map to determine relative parameters for image matching. Point clouds for these frames are then extracted, and initial parameters are fed into the NDT (Normal Distributions Transform) point cloud matching method [33]. This step refines the parameters for accurate relative point cloud alignment between the two frames.

E. Relocalization

Relocalization involves recalibrating the parameters $R_{current}^{prior}$, $T_{current}^{prior}$ between the world coordinate systems Θ_{world}^{prior} , $\Theta_{world}^{current}$, which can be decomposed as follows:

$$\Gamma = [\tau_x, \tau_y, \tau_z, \theta_x, \theta_y, \theta_z] \quad (6)$$

where τ_x, τ_y, τ_z are translation parameters and $\theta_x, \theta_y, \theta_z$ are rotation parameters.

Traditional NDT-based relocalization requires manual selection of the initial parameters. In contrast, our method automatically provides initial values for NDT method. To match the prior map, we use a point cloud map shifted by 2 m as a reference. The IMU magnetometer helps estimate relative angles θ_z . When GPS data is available, absolute coordinates τ_x, τ_y, τ_z are used for relocalization, ensuring accurate position estimates in meters. If GPS data is unavailable, key frame sampling and NDT matching are performed between the current point cloud and the prior map with the key frames. Key frames are typically sampled within twice the moving distance of the current point cloud and the initial parameters are set to the pose of the current key frame. Three high-resolution poses are selected for precise matching to achieve repositioning in the relative coordinate system.

III. EXPERIMENTAL RESULTS

A. Experimental Setups

To validate the efficacy of our proposed method, we conducted experiments using an actual HWR system. The hardware platform employed includes a HWR equipped with a ZED 2i camera, Handsfree A9 IMU, Ublox M8T (GPS), and an edge computing

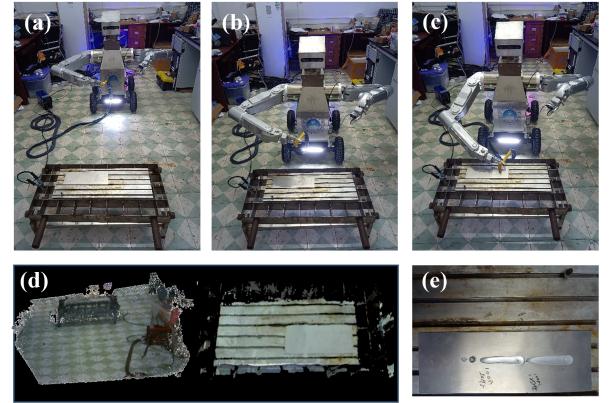


Fig. 2. Verification of HWR in laboratory scene. (a) Movement process of HWR. (b) HWR stops in front of the area to be welded. (c) HWR uses robotic arm to position welding gun. (d) 3D reconstruction results of environment during movement and area to be welded. (e) Welding operation test of HWR.

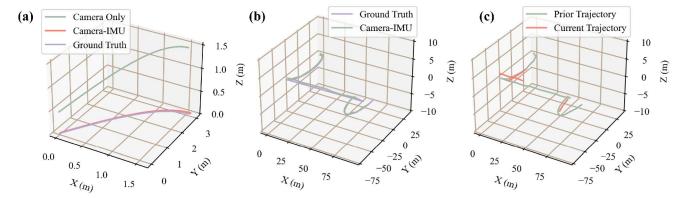


Fig. 3. Pose estimation trajectory results. (a) Laboratory scene (b) Outdoor scene (c) Outdoor relocalization scene.

device (Nvidia AGX). This setup enables the implementation of algorithms such as pose estimation, target detection, 3D reconstruction, and map segmentation. The generated point cloud map and the detected regions earmarked for welding are transmitted via a wired network connection. Based on this data, control signals for the chassis and dual-arm robot, as well as parameters for the welding process, are manually configured and deployed. The experiment included both laboratory (indoor welding) and outdoor scenes. The ground truth for the laboratory scene is obtained using the LiDAR SLAM [29], while the ground truth for the outdoor scene is based on high-precision RTK (Real-Time Kinematic). For the 3D reconstruction results, the fixed-size steel plate is used as the ground truth for comparison.

B. Pose Estimation and 3D Reconstruction

Target detection is not the primary focus of this paper; thus, the evaluation centers on pose estimation and 3D reconstruction accuracy. As shown in Fig. 2, the HWR welding operation test was conducted in a laboratory scenario. Fig. 2(a), (b), and (c) illustrate the workflow of the HWR, while Fig. 2(d) shows the motion process and 3D reconstruction results of the welding area. Fig. 2(e) presents the actual welding result (welding without filler wire). Fig. 3(a) demonstrates the difference between the estimated trajectory and the true value in the laboratory scene. Given the limited indoor space, the outdoor scene was used to evaluate the pose estimation and 3D reconstruction accuracy of the proposed method. Fig. 3(b) shows the difference between the estimated and true trajectories in the outdoor scene.

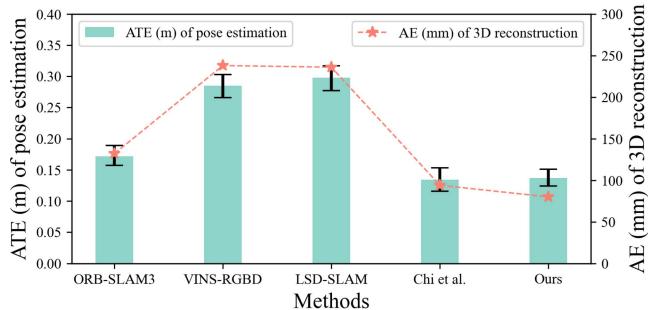


Fig. 4. Comparison of ORB-SLAM3 [30], VINS-RGBD [34], LSD-SLAM [25], Chi et al. [3] and our method in outdoor scene. Evaluation indicators include Absolute Trajectory Error (ATE) of pose estimation and Absolute Error (AE) of 3D reconstruction.

TABLE I
TIME COST (MS) OF ORB-SLAM3 [30], VINS-RGBD [34], LSD-SLAM [25],
CHI ET AL. [3] AND OUR METHOD IN OUTDOOR SCENE

Methods	[30]	[34]	[25]	[3]	Ours
Time cost (ms)	158.2±45.7	56.0±28.5	72.9±32.8	358.2±94.3	162.4±19.1

TABLE II
AVERAGE TIME COST OF MAP TRANSMISSION IN OUTDOOR SCENE, WHERE
THE SEGMENTATION THRESHOLD OF POINTS IS 5×10^5

Size ($\times 10^6$)	0.1	0.9	1.7	2.5	3.8
Direct (s)	1.25	3.95	9.67	26.79	48.95
Segmentation (s)	1.25	2.32	2.74	2.65	2.38

We compare the Absolute Trajectory Error (ATE) for pose estimation and the Absolute Error (AE) for welding area 3D reconstruction, using [30], [34], [25], and [3] as benchmark methods. Additionally, dense mapping is applied to [30] and [25] for 3D reconstruction. As shown in Fig. 4, our method ranks second in ATE for pose estimation, after [3], but outperforms all others in AE for welding area 3D reconstruction. This success is attributed to the incorporation of target detection, which focuses the method on the welding area, improving performance. As shown in Table I, in terms of computational efficiency, our method is slightly less efficient than [3], mainly due to the integration of object detection. However, it remains suitable for HWR applications, as the robot's motion speed does not exceed 5 m/s.

C. Map Segmentation

This section evaluates the effectiveness of our proposed map segmentation method in both laboratory and outdoor environments. Table II compares the transmission times for direct map transmission and segmented map transmission, with the segmentation threshold set at 5×10^5 points, aligned with the robot's operational speed of 5 m/s. Notably, our real-time segmentation method does not compromise map accuracy. Table III compares the pose estimation results before and after applying local map segmentation. The results show that the deviation in global pose estimation after segmentation ranges from -2.34% to -3.75% . Although local map integration does not significantly affect global pose estimation, its application is

TABLE III
INFLUENCE OF THE PROPOSED MAP SEGMENTATION METHOD ON THE (ATE
(M)) (POSE ESTIMATION) AND AE (MM) (3D RECONSTRUCTION) IN
LABORATORY AND OUTDOOR SCENES

Methods	ATE (m)		AE (mm)	
	Laboratory	Outdoor	Laboratory	Outdoor
Without segmentation	0.138	0.341	80	208
With segmentation	0.142	0.349	83	214
Difference(%)	-2.89	-2.34	-3.75	-2.88

TABLE IV
AVERAGE TIME COST (S) OF THE PROPOSED RELOCALIZATION ALGORITHM IN
LABORATORY AND OUTDOOR SCENES

Scenes	Average time cost (s)		
	only GPS	only IMU	GPS and IMU
Outdoor	12.64 ± 2.83	16.24 ± 2.68	1.57 ± 0.67
Laboratory	\	15.37 ± 2.39	\

crucial for human-computer interaction and data transmission efficiency.

D. Relocalization and Map Update

The experimental conditions for evaluating the robot relocalization method follow those outlined in Section III-C. Fig. 3(c) shows the prior trajectory and three current trajectories of the outdoor scene. Table IV presents the relocalization times using GPS alone, IMU alone, and both sensors across various scenes. The results show significant variability in relocalization time when relying solely on GPS. Incorporating IMU data improves both the consistency and efficiency of the process, thanks to the use of initial angular information. Notably, relocalization using only IMU data takes an average of 18.92 s to 23.98 s, demonstrating its feasibility for HWR applications requiring autonomous relocalization.

IV. CONCLUSION

In this paper, we present a 3D reconstruction and relocalization method for HWR operating in environments requiring human intervention, demonstrating strong performance. We first utilized camera and IMU data for accurate pose estimation and 3D reconstruction, with a focus on target detection in the welding area. To address the challenge of slow point cloud map transmission, we introduced a novel map segmentation approach combined with local map matching. For repetitive tasks, we proposed a relocalization method that leverages prior knowledge to improve efficiency. Our HWR system was implemented and validated through comprehensive experiments, confirming its effectiveness. Future work will focus on integrating path planning for mobile robots and trajectory planning for dual-arm robots into the system to further enhance its capabilities.

ACKNOWLEDGMENT

The authors give thanks to the National Key Laboratory for Remanufacturing, Army Academy of Armored Forces for supporting their research.

REFERENCES

- [1] Q. Guo et al., "Progress, challenges and trends on vision sensing technologies in automatic/intelligent robotic welding: State-of-the-art review," *Robot. Comput.-Integr. Manuf.*, vol. 89, 2024, Art. no. 102767.
- [2] P. Chi et al., "Multicamera-IMU calibration system for underwater welding scenes," *IEEE Sensors J.*, vol. 23, no. 21, pp. 26486–26494, Nov. 2023.
- [3] P. Chi et al., "Low-latency visual-based high-quality 3-D reconstruction using point cloud optimization," *IEEE Sensors J.*, vol. 23, no. 17, pp. 20055–20065, Sep. 2023.
- [4] W. Fang and X. Tian, "A novel model-based welding trajectory planning method for identical structural workpieces," *Robot. Comput.-Integr. Manuf.*, vol. 89, 2024, Art. no. 102772.
- [5] Y. Ma, J. Fan, S. Zhao, F. Jing, S. Wang, and M. Tan, "From model to reality: A robust framework for automatic generation of welding paths," *IEEE Trans. Ind. Electron.*, vol. 71, no. 12, pp. 16153–16164, Dec. 2024.
- [6] Z. Lu, J. Fan, Z. Hou, S. Deng, C. Zhou, and F. Jing, "Automatic 3D seam extraction method for welding robot based on monocular structured light," *IEEE Sensors J.*, vol. 21, no. 14, pp. 16359–16370, Jul. 2021.
- [7] P. Chi, Z. Wang, H. Liao, T. Li, X. Wu, and Q. Zhang, "Application of artificial intelligence in the new generation of underwater humanoid welding robots: A review," *Arif. Intell. Rev.*, vol. 57, no. 11, 2024, Art. no. 306.
- [8] H. Su, A. Mariani, S. E. Ovur, A. Menciassi, G. Ferrigno, and E. De Momi, "Toward teaching by demonstration for robot-assisted minimally invasive surgery," *IEEE Trans. Automat. Sci. Eng.*, vol. 18, no. 2, pp. 484–494, Apr. 2021.
- [9] P. Chi, Z. Wang, H. Liao, T. Li, X. Wu, and Q. Zhang, "Towards new-generation of intelligent welding manufacturing: A systematic review on 3D vision measurement and path planning of humanoid welding robots," *Measurement*, vol. 242, 2025, Art. no. 116065.
- [10] P. Chi et al., "Research on image-based 3D reconstruction method in underwater welding scenes," in *Proc. Int. Conf. Sens., Meas. Data Analytics Era Artif. Intell.*, 2023, pp. 1–6.
- [11] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4104–4113.
- [12] B. U. Meinen and D. T. Robinson, "Mapping erosion and deposition in an agricultural landscape: Optimization of UAV image acquisition schemes for SFM-MVS," *Remote Sens. Environ.*, vol. 239, 2020, Art. no. 111666.
- [13] Y. Yao, Z. Luo, S. Li, T. Shen, T. Fang, and L. Quan, "Recurrent MVSNet for high-resolution multi-view stereo depth inference," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 5525–5534.
- [14] R. Chen, S. Han, J. Xu, and H. Su, "Point-based multi-view stereo network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1538–1547.
- [15] K. Deng, A. Liu, J. -Y. Zhu, and D. Ramanan, "Depth-supervised NeRF: Fewer views and faster training for free," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12882–12891.
- [16] B. Roessle, J. T. Barron, B. Mildenhall, P.P. Srinivasan, and M. Nießner, "Dense depth priors for neural radiance fields from sparse input views," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12892–12901.
- [17] T. Wu et al., "Recent advances in 3D Gaussian splatting," *Comput. Vis. Media*, vol. 10, pp. 613–642, 2024.
- [18] W. Chen and L. Liu, "Deblur-GS: 3D Gaussian splatting from camera motion blurred images," *Proc. ACM Comput. Graph. Interactive Techn.*, vol. 7, no. 1, pp. 1–15, 2024.
- [19] K. Zhang, M. Yan, T. Huang, J. Zheng, and Z. Li, "3D reconstruction of complex spatial weld seam for autonomous welding by laser structured light scanning," *J. Manuf. Processes*, vol. 39, pp. 200–207, 2019.
- [20] N. Jia, Z. Li, J. Ren, Y. Wang, and L. Yang, "A 3D reconstruction method based on grid laser and gray scale photo for visual inspection of welds," *Opt. Laser Technol.*, vol. 119, 2019, Art. no. 105648.
- [21] Y. Ma et al., "An efficient and robust complex weld seam feature point extraction method for seam tracking and posture adjustment," *IEEE Trans. Ind. Informat.*, vol. 19, no. 11, pp. 10704–10715, Nov. 2023.
- [22] L. Yang, Y. Liu, J. Peng, and Z. Liang, "A novel system for off-line 3D seam extraction and path planning based on point cloud segmentation for arc welding robot," *Robot. Comput.-Integr. Manuf.*, vol. 64, 2020, Art. no. 101929.
- [23] Z. Tan et al., "A welding seam positioning method based on polarization 3D reconstruction and linear structured light imaging," *Opt. Laser Technol.*, vol. 151, 2022, Art. no. 108046.
- [24] R. A. Newcombe et al., "KinectFusion: Real-time dense surface mapping and tracking," in *Proc. 10th IEEE Int. Symp. Mixed Augmented Reality*, 2011, pp. 127–136.
- [25] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. 13th Eur. Conf. Comput. Vis.*, 2014, pp. 834–849.
- [26] R. A. Newcombe, D. Fox, and S. M. Seitz, "DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 343–352.
- [27] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. J. Davison, "ElasticFusion: Dense SLAM without a pose graph," in *Proc. Robot. Sci. Syst.*, 2015.
- [28] P. Chi, H. Liao, Q. Zhang, X. Wu, J. Tian, and Z. Wang, "Online static point cloud map construction based on 3D point clouds and 2D images," *Vis. Comput.*, vol. 40, no. 4, pp. 2889–2904, 2024.
- [29] P. Chi, Z. Wang, H. Liao, X. Wu, J. Tian, and Q. Zhang, "Robot localization and reconstruction based on 3D point cloud," in *Proc. 32nd IEEE Int. Conf. Robot. Hum. Interactive Commun.*, 2023, pp. 520–525.
- [30] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [31] W. Wang and P. G. Adamczyk, "Comparison of Bingham filter and extended Kalman filter in IMU attitude estimation," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8845–8854, Oct. 2019.
- [32] A. Wang et al., "YOLOv10: Real-time end-to-end object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2025, vol. 37, pp. 10794–108011.
- [33] M. Magnusson, "The three-dimensional normal-distributions transform: An efficient representation for registration, surface analysis, and loop detection," Ph.D. dissertation, Örebro universitet, Örebro, Sweden, 2009.
- [34] Z. Shan, R. Li, and S. Schwerdfeger, "RGBD-inertial trajectory estimation and mapping for ground robots," *Sensors*, vol. 19, no. 10, 2019, Art. no. 2251.