Opto-Electronic Advances

CN 51-1781/TN ISSN (Print) 2096-4579 ISSN (Online) 2097-3993

Scene-level passive polarization 3D imaging

Xin Wang, Pingli Han, Xiyuan Luo, Qianqian Liu, Tong Zhang, Xue Dong, Meng Xiang, Jinpeng Liu, Yanyan Liu and Fei Liu*

Citation: Wang X, Han PL, Luo XY, Liu QQ, Zhang T, Dong X, Xiang M, Liu JP, Liu YY, Liu F. Scene-level passive polarization 3D imaging. *Opto-Electron Adv* **9**, 250267 (2026).

https://doi.org/10.29026/oea.2026.250267

Received: 8 October 2025; Accepted: 20 January 2026; Published online: 10 February 2026

OEA | Opto-Electronic Advances

Letter

February 2026, Vol. 9, No. 2

Open Access

Scene-level passive polarization 3D imaging

Xin Wang[1†], Pingli Han[1†], Xiyuan Luo[1], Qianqian Liu[1], Tong Zhang[1], Xue Dong[1], Meng Xiang[1], Jinpeng Liu[1], Yanyan Liu[2] and Fei Liu[1,3*]

Abstract: Scene-level passive 3D imaging under natural conditions is highly challenging yet urgently demanded. Polarization 3D provides possibility but impeded by two major obstacles brought by natural large scenes: discontinuities of multiple targets and dynamic reconstruction. This study proposed a scene-level passive polarization 3D imaging method, integrating binocular stereo and polarization. We abstract the discontinuous targets reconstruction into a minimization problem. The pixel-level normal direction from polarization and the absolute scale information from binocular stereo vision then work as mutual constraints for iterative optimization of the problem. By iterating for final solution, the challenge of reconstructing discontinuous targets was tackled, and true depth was also recovered. The true depth then provides a reference for solving inter-frame scale inconsistencies which hinders dynamic reconstruction by the designed scale normalization strategy which globally aligns multi-view measurement data. Scene-level 3D structure was finally reconstructed through multi-frame point cloud fusion. We showcase wide-scene, high-accuracy passive video reconstructions on natural field scenes. Our passive polarization stereo represents a major advancement in scene-level 3D imaging and may find broad applications in fields requiring passive 3D imaging solutions.

Keywords: 3D imaging; polarization; natural scenes

DOI: 10.29026/oea.2026.250267 | CSTR: 32247.14.oea.2026.250267

Citation: Wang X, Han PL, Luo XY et al. Scene-level passive polarization 3D imaging. *Opto-Electron Adv* **9**, 250267 (2026).

3D imaging is experiencing the transition from reconstructing individual objects to the comprehensive capture of entire scenes[1]. Well-developed 3D imaging techniques have difficulties in managing large scenes due to reliance on controlled active illumination and scanning methods, including structured light, light detection and ranging (LiDAR), stereo vision, and deep learning[2–4]. Passive scene-level 3D imaging remains to be a highly challenging task. Polarization 3D imaging estimates 3D structures based upon polarization information of reflected light[5–7]. It has original advantages in passive long-distance conditions and high precision reconstruction. Major challenges to reconstruct natural scenes are the inevitable π ambiguity of azimuth component of the surface normal[8,9], and the discontinuity of multiple targets[10,11]. Researchers have explored addressing ambiguity utilizing a stereo depth map as a guiding surface[12,13]. Neural networks are also employed to resolve polarization ambiguities for high-accuracy reconstruction, such as human faces[14–18]. But structure reconstruction still majorly relies on surface normal integration, resulting in limitations in discontinuity and pseudo-depth information. Accordingly, to implement scene-level 3D imaging, we need a feasible solution to fully harness the passive and high-precision potential inherently offered by polarization. The Tian's group employed binocular stereo vision to develop a joint 3D reconstruction model for true depth recovery[19]. This model assumes the 3D shape from polarization imaging always has a scaling and offset relationship with real depth. But the linear relationship is limited to single continuous objects and won't work for large scenes with multiple independent targets. Cui handled surface normal ambiguity and reconstructed 3D shape of single texture-less object incorporating polarization and multi-view imaging[20]. Therefore, multiple discontinuous targets in natural scenes put a great challenge to polarization 3D reconstruction. Besides, for scene-level dynamic 3D reconstruction, dynamic camera

Received: 8 October 2025

Accepted: 20 January 2026

Published online: 10 February 2026

[1]School of Optoelectronic Engineering, Xidian University, Xi'an 710071, China; [2]National Key Laboratory of Electromagnetic Space Security, Tianjin 300308, China; [3]State Key Laboratory of Electromechanical Integrated Manufacturing of High-Performance Electronic Equipments, Xidian University, Xi'an 710071, China.

†These authors contributed equally to this work.

*Correspondence: F Liu, E-mail: feiliu@xidian.edu.cn

path and frames fusion are evidently necessary as illustrated in the Supplemental videos 1 and 2. It brings further challenge of scale inconsistencies between frames where cumulative errors and inaccurate camera pose estimation may lead to discontinuities or jumps in the final scene 3D.

In this letter, we present a passive scene-level non-integral polarization 3D imaging method shown in Fig. 1. It addresses the challenges of scene discontinuity and inter-frame scale inconsistencies that have long plagued scene-level 3D reconstruction. Heterogeneous information, the pixel-level surface normal from polarization and the absolute scale information from binocular stereo vision, were unified by joint iteration for reconstructing accurate 3D structure. The discontinuous scene reconstruction was abstracted into a minimization problem, where the absolute depth and normal direction work as mutual constraints for iterative optimization. The Alternating Direction Method of Multipliers (ADMM) decomposes the original problem into simpler subproblems and optimizes them alternately, ensuring the algorithm's efficient convergence and stability[21]. The iteration solution tackles challenges of discontinuous targets in large-scale reconstruction, enabling simultaneous high-precision and true-depth 3D reconstruction. The true depth then provides a reference for solving inter-frame scale inconsistencies by the designed scale normalization strategy which globally aligns multi-view measurement data. Scene-level 3D structure was finally reconstructed through point cloud fusion.

Scene-level high-precision 3D structures require both global depth from stereo vision and detailed textures from polarization. One problem of polarization 3D is the π ambiguity, making surface normal correction necessary. Please see correction algorithm in the Supplemental information. We focus on the discontinuity problem. The 3D reconstruction is abstracted into a minimization problem expressed by the energy function in Eq. (1), which includes three energy terms: contour fidelity term E_1, texture fidelity term E_2, and low-rank term E_3,

$$E = E_1 + \xi E_2 + \eta E_3 \ , \tag{1}$$

where ξ and η are balance parameters adjusting the weight of contour and texture constraints. The term $E_1 = ||\boldsymbol{D} - \boldsymbol{S}||_2^2/2$ constrains the contour of the reconstruction result by introducing depth $\boldsymbol{D}$ from stereo vision. It's a least-square problem between the surface $\boldsymbol{S}$ to be reconstructed and the stereo vision depth. It's described by the L_2 norm, which can measure the error between the predicted and true values. The energy term $E_2 = ||\boldsymbol{N}_{\rm p} - \nabla \boldsymbol{S}||_2^2/2$ ensures detailed textures with polarization normal $\boldsymbol{N}_{\rm p}$ as another constraint, where ∇ denotes the normal of the surface $\boldsymbol{S}$.

When imaging, the camera always faces the target scene, which means the reconstructed depth should consistently be positive. A positive-value discrimination function F was then formulated as Eq. (2),

$$F(\boldsymbol{S}) = \begin{cases} 1 & \boldsymbol{S} > 0 \\ \boldsymbol{S} & \boldsymbol{S} \le 0 \end{cases} . \tag{2}$$

It evaluates the depth values and penalize non-positive ones.

We apply it to surface $\boldsymbol{S}$ and get a low-rank matrix $\boldsymbol{FS}$. Then this low-rank problem is described by the nuclear norm $E_3 = ||\boldsymbol{FS}||_*$.

Till now, the function in Eq. (1) is formulated into Eq. (3),

$$E = \frac{1}{2}||\boldsymbol{D} - \boldsymbol{S}||_2^2 + \frac{\xi}{2}||\boldsymbol{N}_P - \nabla \boldsymbol{S}||_2^2 + \eta||\boldsymbol{FS}||_* \ . \tag{3}$$

This framework integrates both the coarse depth from stereo vision and the detailed surface normal from polarization, making scene-level 3D reconstruction possible. It converts the correction of surface normal field and the integration problem into a least-squares minimization constraint and iterative optimization problem. Yet, this problem is inherently nonlinear and ill-posed, posing challenges for straightforward solutions. Therefore, we introduce Lagrange multipliers to link the constraint conditions with the original problem. By the ADMM algorithm, the model is decomposed into a primary problem and several subproblems in Eq. (4), where L_i (i=1, 2) are Lagrange multipliers. Please refer to the supplement document for detailed solution process and parameters selection. By solving the optimization problem iteratively for its analytical solution,

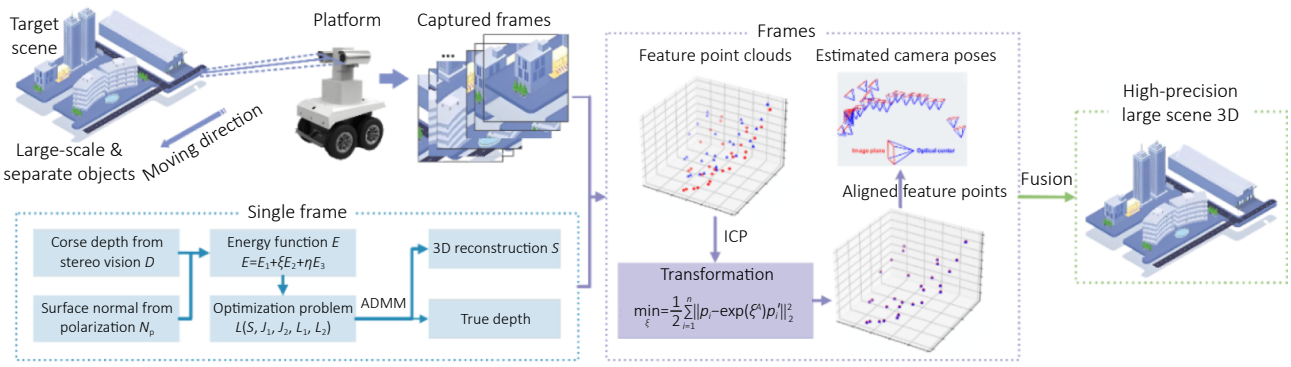

Fig. 1 | Overview of the proposed method.

when the iteration error falls below the predefined threshold, we get the reconstructed high-precision scene-level 3D.

$$L(\boldsymbol{S},\ \boldsymbol{J}_1,\ \boldsymbol{J}_2,\ \boldsymbol{L}_1,\ \boldsymbol{L}_2) = \frac{1}{2}||\boldsymbol{D}-\boldsymbol{S}||_2^2 + \frac{\xi}{2}||\boldsymbol{N}_P-\boldsymbol{J}_1||_2^2 + \eta||\boldsymbol{J}_2||_* - \langle \boldsymbol{L}_1, \boldsymbol{J}_1-\nabla \boldsymbol{S}\rangle - \langle \boldsymbol{L}_2, \boldsymbol{J}_2-\boldsymbol{F}\boldsymbol{S}\rangle + \frac{\rho}{2}(||\boldsymbol{J}_1-\nabla \boldsymbol{S}||_2^2 + ||\boldsymbol{J}_2-\boldsymbol{F}\boldsymbol{S}||_2^2)\ . \tag{4}$$

Limited by imaging field-of-view, scene-level reconstruction requires dynamic imaging by scanning as shown by the attached Supplemental video 1. Two major challenges come along: the dynamic reconstruction and multi-frame data fusion. Both make point cloud registration necessary. It's completed by the Iterative Closest Point (ICP)[22] to first calculate reliable camera pose parameters and further accurate point match. It's finally expressed by Eq. (5):

$$\min_{\xi} = \frac{1}{2}\sum_{i=1}^{n}||\boldsymbol{p}_i - \exp(\xi^{\wedge})\boldsymbol{p}_i'||_2^2\ , \tag{5}$$

where $\xi^{\wedge}$ is the pose expressed by Lie algebra, $\boldsymbol{p}_i$ indicates a feature point. Please see Algorithm S3 in the Supplemental information for details.

The designed and fabricated scene-level polarization 3D imaging system is shown in Supplemental video 2. The platform (baseline length B = 100 mm) consists of two polarization cameras equipped with Sony polarized sensors. It is capable of video imaging at 25 frames per second at full frame (2448 × 2048 pixels). Figure 2(a) and 2(c) provide an example of the reconstructed natural large scene. It's a typical natural scene containing typical targets including plants, buildings, and road surfaces. For analyses of the polarimetric characteristics of these targets, please refer to the supplement information. Figure 2(b) is reconstruction before feature-point aligning. For a whole large scene, each 3D point cloud we acquire corresponds to the camera's local coordinate system at the respective capture positions. If we directly fuse these point clouds, the resulting point clouds will not be aligned with each other, as shown in Fig. 2(b). Therefore, by estimating the camera's pose parameters by ICP (in Supplemental information), the point clouds of different views were aligned into a common coordinate system and we then get accurate 3D of the natural scene in Fig. 2(c).

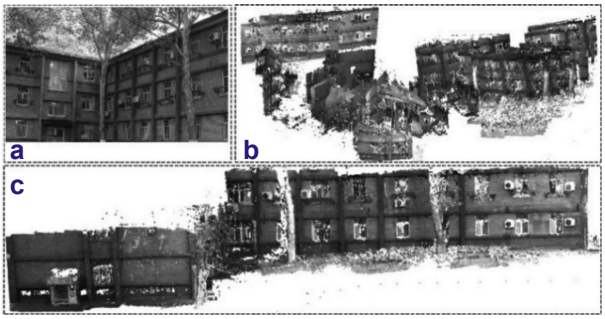

Fig. 2 | Illustration of a reconstructed natural scene. (a) Directly captured image of partial scene. (b) Unaligned scene point clouds. (c) Finally aligned scene point clouds.

Figure 3(a) illustrates the dynamic scanning process, and 3(b) shows 8 extracted frames from an 8 s video. The intensity images captured at each timestamp are displayed along the timeline. The frames are featured by a variety of objects, such as buildings and trees. The reconstruction results in the second row demonstrate the method's consistency and its ability to deliver precise 3D reconstructions across all frames. We randomly chose the image captured at t=3 s for further evaluation. Figure 3(c1) and 3(c2) illustrate that the reconstruction surface precisely captures the spatial arrangements and intricate textures of the scene's elements, such as trees and buildings. Figure 3(c3) and 3(c4), enlargement of the areas shown in Fig. 3(c2), expose distinct details of the wall's gaps and lattice fences. The method consistently reconstructs intricate texture details across various regions in each frame, underscoring its suitability for the dynamic 3D imaging of natural scenes.

Figure 4 showcases the enhancements of the proposed method over existing polarization 3D (F-C integration) and stereo vision[23]. Figure 4(a) displays a captured intensity image of a natural scene, featuring separate objects like a

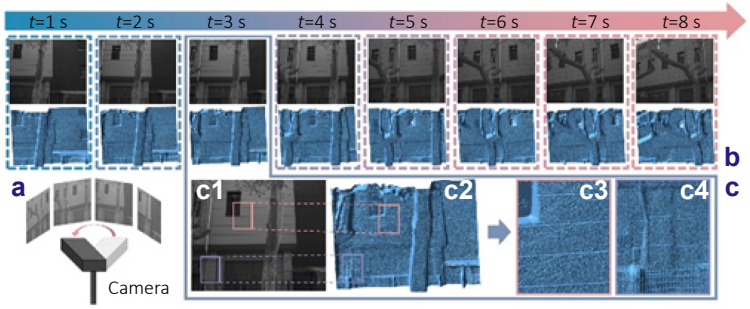

Fig. 3 | Natural scene dynamic reconstruction. (a) Video-rating camera. (b) Eight pairs of intensity maps and reconstruction results at 1st to 8th seconds. (c) Details demonstration taking the image from 3rd second as an example.

lamppost, trees, and buildings, which typify a discontinuous scene with first-kind discontinuity points. We have delineated four feature regions, denoted as ①~④. Area ① indicates discontinuity problem; area ② indicates intricate - texture problem and area ③ and ④ correspond to various targets. Figure 4(b1) and 4(b2) illustrate the depth pseudo-color maps of area ① reconstructed by the F-C integration and the proposed method. Regarding depth distribution, the reconstruction by the proposed method more closely aligns with the true depth distribution, while in contrast the F-C integration fails to accurately represent the spatial relationships between objects in the scene. Figure 4(b3) shows a transverse slice of the depth values taken at y=400 in Fig. 4(b1) and 4(b2). In Fig. 4(a), area ① provides two independent objects: the lamppost and the tree. The actual distance between the lamppost and the tree is 7.88 m. According to Fig. 4(b3), the F-C integration's reconstruction distance measures 2.04 meters, deviating significantly from the actual distance. In contrast, the proposed method's reconstruction distance is 7.79 meters, with a mere 1.1% deviation from the actual distance. This demonstrates that the proposed method effectively tackles the challenge of first-kind discontinuity points and verifies that the incorporation of the contour fidelity term in Eq. (1) efficiently mitigates the integration issues arising from discontinuities within the F-C integration algorithm.

For the texture fidelity term in the proposed method, we focused on the intricate textures in area ② in Fig. 4(a). Figure 4(c1) and 4(c3) are respectively reconstructed by stereo vision and the proposed method. Since the stereo vision depth constrains the contour fidelity term, Fig. 4(c1) and 4(c3) demonstrate a high degree of conformity in overall shape contour. However, upon magnifying the region with intricate textures, the stereo vision reconstruction reveals scarcely any discernible useful texture details. In contrast, our proposed method successfully captures the necessary detailed texture information, thereby highlighting the effectiveness of the texture fidelity term in Eq. (1). Figure 4(d1) to 4(d4) indicate its robustness for more various objects. Thus, the optimization function, which seamlessly integrates coarse depth and polarization information, not only addresses discontinuity challenges in natural scene reconstruction but also effectively preserves texture details.

For precision assessment, we performed in-depth analyses of the reconstructed scene's finer details. Taken as example, the natural scene depicted in Fig. 5(a) encompasses both large-scale structures and intricate textures. Figure 5(b1) and 5(b2) display the reconstructed 3D surfaces from stereo vision and the proposed method, respectively. Visually, Fig. 5(b2) not only precisely defines the spatial positions but also retains richer texture details compared to Fig. 5(b1), suggesting enhanced performance in details representation. Figure 5(d) and 5(e) depicted intensity statistics along x and y orientations as indicated by Fig. 5(a). Both methods enable overall depth indication. But only the proposed method

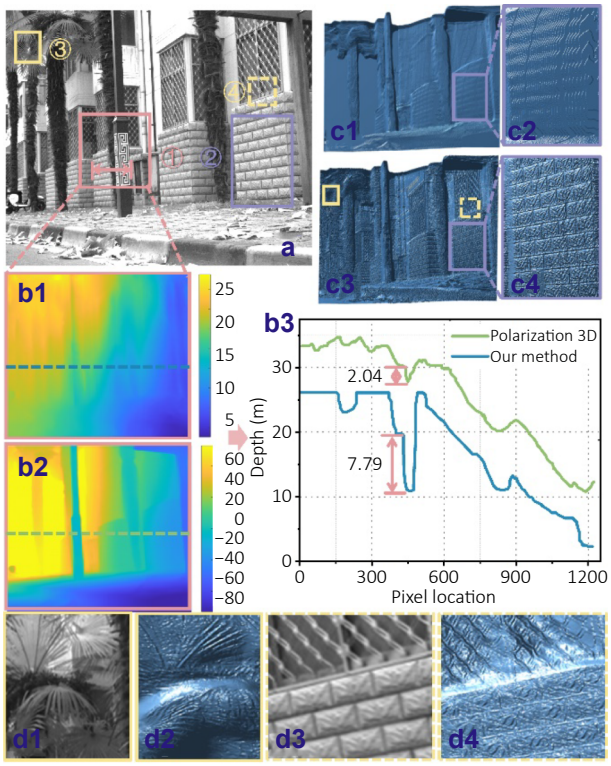

Fig. 4 | Detail analyses of a natural scene. (a) Intensity map. (b) Depth maps of area ① in scene (a) from polarization 3D (b1) and our method (b2, b3) depth variation of (b1, b2) along the indicated lines. (c) Reconstruction and detail illustration of (a) by stereo vision (c1, c2) and our method (c3, c4). (d) Details presentation of the proposed method.

effectively reconstructs details including the brick seams and surface textures. Due to noise and cumulative errors in the reconstruction process, minor irregularities are visible at the edges of the brick seams in Fig. 5(e). The proposed method's average reconstructed seam depth is about 2.0 cm, close to the actual measurements of 1.5 cm to 2.1 cm, averaging 1.65 cm. The three blue × in Fig. 5(a) are three randomly selected points. The corresponding reconstruction data as depicted by Fig. 5(c). The curves fit well with the highest discrepancy being only around 1.4%. These findings confirm that our method delivers precise 3D imaging for natural scenes, capturing both large-scale structures and small-scale details with a level of accuracy surpassing prior 3D imaging techniques.

In conclusion, we have introduced a passive scene-level non-integral polarization 3D imaging method. It offers two key advantages: (1) scene-level polarization measurement and further scene-level 3D reconstruction in a passive way and (2) 3D reconstruction at absolute depth level. Specially, we designed an integrated binocular polarization imaging system simultaneously capturing polarization and depth data. An energy function containing three energy terms controlling structures and details is developed for iterative optimization by taking the depth and normal direction work

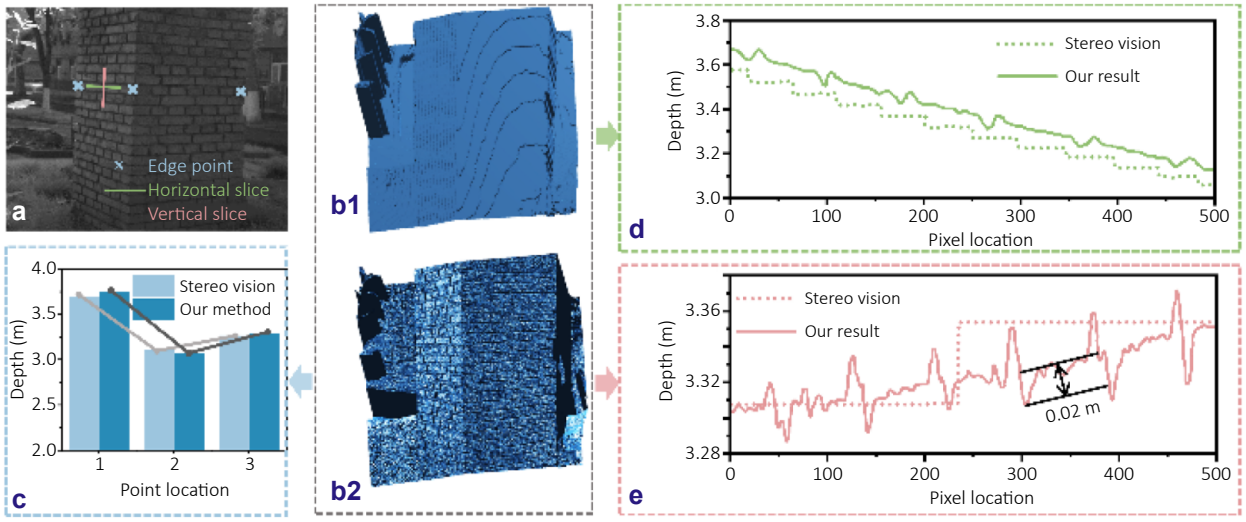

Fig. 5 | Precision analyses of a natural scene. (a) Intensity map. (b) Reconstruction of (a) by stereo vision (b1) and our method (b2). (c) Depth comparison of the two methods at marked locations by ×. (d) Depth variation of (b1, b2) along horizontal direction indicated by green in (a). (e) Depth variation of (b1, b2) along vertical direction indicated by pink in (a).

as mutual constraints, which realizes natural discontinuous scene reconstruction and true depth recovery. 3D imaging to various scenes demonstrates the proposed method enabled robust scene-level 3D reconstruction, and centimeter-level reconstruction precision. It holds promise for providing new technical means for 3D reconstruction applications requirement of natural scenes. Future work may pay attention to more effective optimization model development or more diverse targets complementation for potential applications.

References

1. Chowdhury S, Chen M, Eckert R et al. High-resolution 3D refractive index microscopy of multiple-scattering samples from intensity images. *Optica* **6**, 1211–1219 (2019).
2. Zhang C, Wang YP, Yin YK et al. High precision 3D imaging with timing corrected single photon LiDAR. *Opt Express* **31**, 24481–24491 (2023).
3. Tan CS, Kong W, Huang GH et al. Development of a near-infrared single-photon 3D imaging LiDAR based on 64×64 InGaAs/InP array detector and Risley-prism scanner. *Opt Express* **32**, 7426–7447 (2024).
4. Bi YW, Li CB, Tong XR et al. An application of stereo matching algorithm based on transfer learning on robots in multiple scenes. *Sci Rep* **13**, 12739 (2023).
5. Lee J, Usmani K, Javidi B. Polarimetric 3D integral imaging profilometry under degraded environmental conditions. *Opt Express* **32**, 43172–43183 (2024).
6. Usmani K, O'Connor T, Javidi B. Three-dimensional polarimetric image restoration in low light with deep residual learning and integral imaging. *Opt Express* **29**, 29505–29517 (2021).
7. Atkinson GA, Hancock ER. Recovery of surface orientation from diffuse polarization. *IEEE Trans Image Process* **15**, 1653–1664 (2006).
8. Wang JZ, Cossairt O, Willomitzer F. 3D imaging of complex specular surfaces by fusing polarimetric and deflectometric information. *Optica* **12**, 446–450 (2025).
9. Shen ZC, Zhao F, Ni YB et al. Extended monocular 3D imaging via the fusion of diffraction- and polarization-based depth cues. *Optica* **12**, 872–878 (2025).
10. Hu Y, Xu CH, Hao Q et al. High-precision polarization imaging for three-dimensional reconstruction aided by a separate coarse depth map. *Appl Opt* **63**, 4799–4808 (2024).
11. Kadambi A, Taamazyan V, Shi BX et al. Polarized 3D: high-quality depth sensing with polarization cues. In *IEEE International Conference on Computer Vision (ICCV)* 3370–3378 (IEEE, 2015). https://doi.org/10.1109/ICCV.2015.385.
12. Wen SH, Liu X, Zhang H et al. Dense point cloud map construction based on stereo VINS for mobile vehicles. *ISPRS J Photogramm Remote Sens* **178**, 328–344 (2021).
13. Zhu DZ, Smith WAP. Depth from a polarisation + RGB stereo pair. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 7578–7587 (IEEE, 2019). https://doi.org/10.1109/CVPR.2019.00777.
14. Han XF, Laga H, Bennamoun M. Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era. *IEEE Trans Pattern Anal Mach Intell* **43**, 1578–1604 (2021).
15. Usmani K, Krishnan G, O'Connor T et al. Deep learning polarimetric three-dimensional integral imaging object recognition in adverse environmental conditions. *Opt Express* **29**, 12215–12228 (2021).
16. Han PL, Cai YD, Liu F et al. Computational polarization 3D: new solution for monocular shape recovery in natural conditions. *Opt Lasers Eng* **151**, 106925 (2022).
17. Luo XY, Wang S, Liu JP et al. Revolutionizing optical imaging: computational imaging via deep learning. *Photonics Insights* **4**, R03 (2025).
18. Yang K, Liu F, Liang SY et al. Data-driven polarimetric imaging: a review. *Opto-Electron Sci* **3**, 230042 (2024).
19. Tian X, Liu R, Wang ZY et al. High quality 3D reconstruction based on fusion of polarization imaging and binocular stereo vision. *Inf Fusion* **77**, 19–28 (2022).
20. Cui ZP, Gu JW, Shi BX et al. Polarimetric multi-view stereo. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 369–378 (IEEE, 2017). https://doi.org/10.1109/CVPR.2017.47.
21. Almeida MSC, Figueiredo M. Deconvolving images with unknown boundaries using the alternating direction method of multipliers.

IEEE Trans Image Process **22**, 3074–3086 (2013).
22. Maier-Hein L, Franz AM, dos Santos TR et al. Convergent iterative closest-point algorithm to accomodate anisotropic and inhomogenous localization error. *IEEE Trans Pattern Anal Mach Intell* **34**, 1520–1532 (2012).
23. Frankot RT, Chellappa R. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans Pattern Anal Mach Intell* **10**, 439–451 (1988).

Acknowledgements

This research was funded by the National Natural Science Foundation of China (62375212, 62405231, 62405235).

Competing interests

The authors declare no competing financial interests.

Data availability

Data underlying the results presented in this Letter are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplementary information

Supplementary information for this paper is available at https://doi.org/10.29026/oea.2026.250267