OEA | Opto-Electronic Advances

Supplementary information

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*]

[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

This file includes:

Section 1: Polarization characteristics in natural scenes
Section 2: Surface normal correction
Section 3: Optimization by iteration
Section 4: Point cloud registration of large scenes
Section 5: ADMM parameters determination

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

Section 1: Polarization characteristics in natural scenes

Since our research focused on 3D imaging of natural scenes based on polarization. It's essential to first test the existence of available polarization information. The tests should cover key elements of natural scenes. Vegetation, roads, and buildings constitute fundamental elements of natural scenes, spanning typical categories from natural organic to artificial objects. Data are illustrated in Fig. S1. Overall, the polarimetric characteristics of the ground objects are relatively weak. But the inverted azimuth and zenith angles still contain structural information of targets. This information will be encoded into the surface normal during subsequent reconstruction.

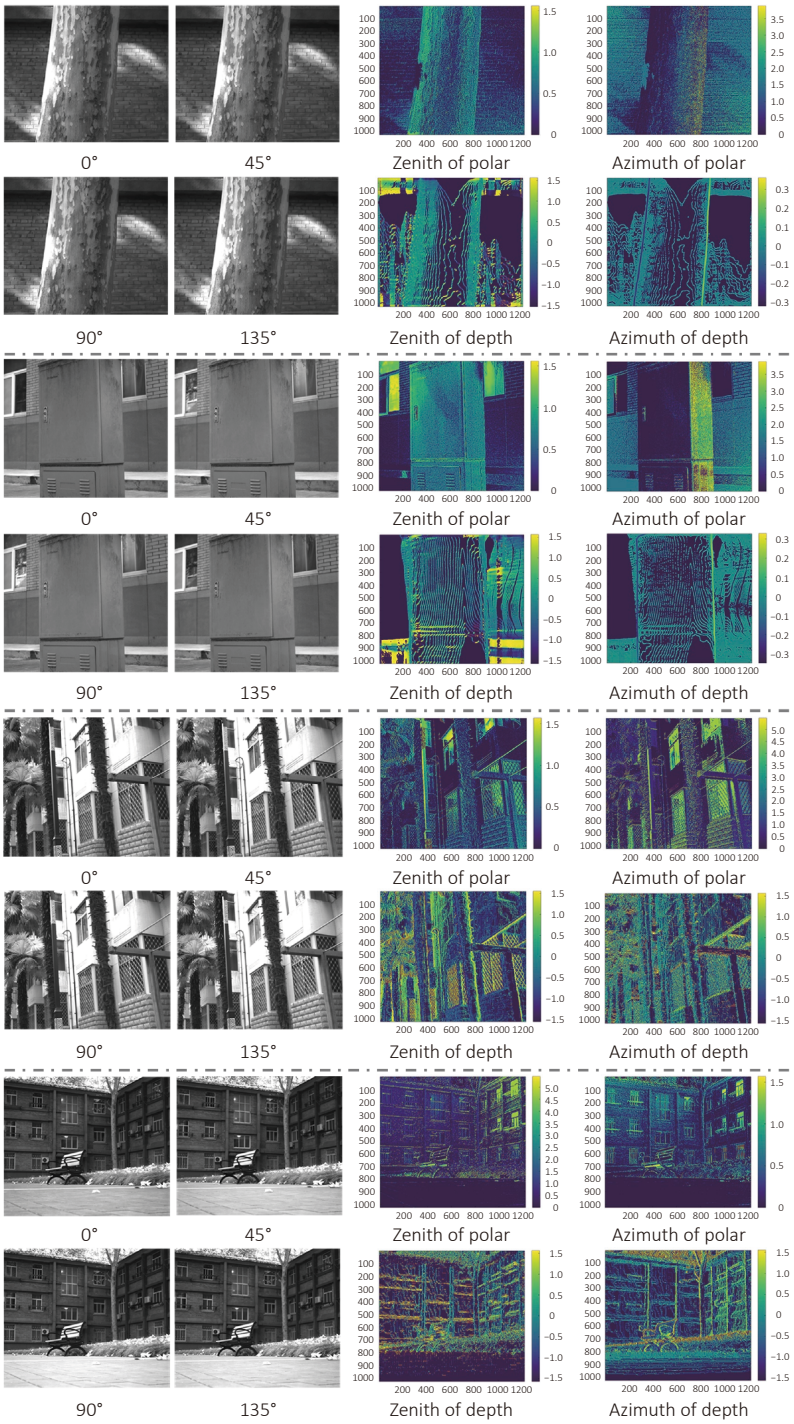


Fig. S1 | Polarimetric and depth information of typical objects in natural scenes.

Section 2: Surface normal correction

Polarization 3D establishes a mapping that correlates the polarization of the light field and the 3D morphology. The essence lies in the comprehensive exploitation of the degree of polarization (DoP) and the angle of polarization (AoP) of the light field detected by the detector, thereby facilitating the determination of the zenith θ and azimuth Φ of the surface normal field. We express surface normal by Eq. (S1),

$$\boldsymbol{N}_{\mathrm{P}} = \begin{Bmatrix} N_{\mathrm{P}x} \\ N_{\mathrm{P}y} \\ N_{\mathrm{P}z} \end{Bmatrix} = \begin{Bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{Bmatrix}, \tag{S1}$$

where N_{Px}, N_{Py} and N_{Pz} denote the components of the surface normal in x, y and z directions. The trigonometric functions in the inverse solution of the *AoP* lead to a possible ϕ phase ambiguity of π ($\phi = \varphi$ or $\phi = \varphi + \pi$) and further surface normal ambiguity. Moreover, given the unit norm length of the normal vector got from Eq. (S1), the integration to surface normal only provides a measure of relative height of the surface.

To realize polarization 3D from 2D images with $M\times N$ pixels, we need to accurately determine $2MN$ unknowns (θ, φ). The depth of the scene derived from stereo vision, typically aligns with the consistent variation trend observed in the polarization normal field information. Assume the gradients of this depth map along the X and Y axes are represented as $g = [\partial Z/\partial X, \partial Z/\partial Y]^{\mathrm{T}}$. To facilitate subsequent algorithmic processing, we convert this gradient vector field into a normal vector field perpendicular to the micro-facets and perform normalization, which can be expressed as:

$$\boldsymbol{N}_{\mathrm{B}} = \begin{bmatrix} N_{\mathrm{B}x} \\ N_{\mathrm{B}y} \\ N_{\mathrm{B}z} \end{bmatrix} = \begin{bmatrix} -g \\ 1 \end{bmatrix} = \frac{1}{\sqrt{(\partial Z/\partial x)^2 + (\partial Z/\partial y)^2 + 1^2}} \begin{bmatrix} -\partial Z/\partial x \\ -\partial Z/\partial y \\ 1 \end{bmatrix}, \tag{S2}$$

where N_{Bx} is the normal from binocular depth, N_{Bx}, N_{By} and N_{Bz} are the components of N_B along three directions.

We optimize the scene's normal distribution by aligning the stereo-vision-derived normal field with the polarization data, using a tailored optimization function in Eq. (S3),

$$\begin{cases} N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} = N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} & \text{if } N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} \times N_{\mathrm{B}_x\mathrm{or}\mathrm{B}_y} \geq 0 \\ N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} = -N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} & \text{else if } N_{\mathrm{P}_x\mathrm{or}\mathrm{P}_y} \times N_{\mathrm{B}_x\mathrm{or}\mathrm{B}_y} < 0 \end{cases}. \tag{S3}$$

Section 3: Optimization by iteration

As stated in the main text, the problem we seek to solve is expressed by Eq. (S4),

$$\begin{aligned} L(\boldsymbol{S}, \boldsymbol{J}_1, \boldsymbol{J}_2, \boldsymbol{L}_1, \boldsymbol{L}_2) = {} & \frac{1}{2}\left\|\boldsymbol{D} - \boldsymbol{S}\right\|_2^2 + \frac{\xi}{2}\left\|\boldsymbol{N}_P - \boldsymbol{J}_1\right\|_2^2 + \eta\left\|\boldsymbol{J}_2\right\|_* - \langle \boldsymbol{L}_1, \boldsymbol{J}_1 - \nabla \boldsymbol{S}\rangle - \langle \boldsymbol{L}_2, \boldsymbol{J}_2 - \boldsymbol{F}\boldsymbol{S}\rangle \\ & + \frac{\rho}{2}(\left\|\boldsymbol{J}_1 - \nabla \boldsymbol{S}\right\|_2^2 + \left\|\boldsymbol{J}_2 - \boldsymbol{F}\boldsymbol{S}\right\|_2^2) \end{aligned}. \tag{S4}$$

We reformulated the original problem into an augmented form by introducing Lagrange multipliers. This was done to facilitate variable separation for alternating iterations, simplifying the original problem. The problem is solved by ADMM where we first derive the optimality conditions for each primal variable, assuming the others are fixed:

$$\begin{cases} \boldsymbol{S}^{k+1} = \underset{S}{\mathrm{argmin}} \dfrac{1}{2}\left\|\boldsymbol{D} - \boldsymbol{S}\right\|_2^2 + \langle \boldsymbol{L}_1{}^k, \boldsymbol{J}_1{}^k - \nabla \boldsymbol{S}\rangle + \langle \boldsymbol{L}_2{}^k, \boldsymbol{J}_2{}^k - \boldsymbol{F}\boldsymbol{S}\rangle \\ \qquad\qquad + \dfrac{\rho}{2}(\left\|\boldsymbol{J}_1{}^k - \nabla \boldsymbol{S}\right\|_2^2 + \left\|\boldsymbol{J}_2{}^k - \boldsymbol{F}\boldsymbol{S}\right\|_2^2) \\ \boldsymbol{J}_1^{k+1} = \underset{J_1}{\mathrm{argmin}} \dfrac{\xi}{2}\left\|\boldsymbol{N}_P - J_1\right\|_2^2 + \langle \boldsymbol{L}_1{}^k, \boldsymbol{J}_1 - \nabla \boldsymbol{S}^{k+1}\rangle + \dfrac{\rho}{2}\left\|\boldsymbol{J}_1 - \nabla \boldsymbol{S}^{k+1}\right\|_2^2 \\ \boldsymbol{J}_2^{k+1} = \underset{J_2}{\mathrm{argmin}} \eta\left\|\boldsymbol{J}_2\right\|_* + \left\langle \boldsymbol{L}_2{}^k, \boldsymbol{J}_2 - \boldsymbol{F}\boldsymbol{S}^{k+1}\right\rangle + \dfrac{\rho}{2}\left\|\boldsymbol{J}_2 - \boldsymbol{F}\boldsymbol{S}^{k+1}\right\|_2^2 \\ \boldsymbol{L}_1^{k+1} = \boldsymbol{L}_1^k + \rho(\boldsymbol{J}_1^{k+1} - \nabla \boldsymbol{S}^{k+1}) \\ \boldsymbol{L}_2^{k+1} = \boldsymbol{L}_2^k + \rho(\boldsymbol{J}_2^{k+1} - \boldsymbol{F}\boldsymbol{S}^{k+1}) \end{cases}. \tag{S5}$$

By updating each variable, we get the solution at iteration k expressed by Eqs. (S6–S10),

$$\boldsymbol{S}^{k+1} = (1 + \rho\nabla^{\mathrm{T}}\nabla + \rho\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F})^{-1}(D + \nabla^{\mathrm{T}}\boldsymbol{L}_1^k + \boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}_2^k + \rho\nabla^{\mathrm{T}}\boldsymbol{J}_1^k + \rho\boldsymbol{F}^{\mathrm{T}}\boldsymbol{J}_2^k), \tag{S6}$$

$$\boldsymbol{J}_1^{k+1} = \frac{\xi \boldsymbol{N}_P + \rho \nabla \boldsymbol{S}^{k+1} - \boldsymbol{L}_1^k}{\xi + \rho} , \tag{S7}$$

$$\boldsymbol{J}_2^{k+1} = \boldsymbol{S}_{\eta/\rho}(\boldsymbol{F}\boldsymbol{S}^{k+1} - \boldsymbol{L}_2^k/\eta) , \tag{S8}$$

$$\boldsymbol{L}_1^{k+1} = \boldsymbol{L}_1^k + \rho(\boldsymbol{J}_1^{k+1} - \nabla \boldsymbol{S}^{k+1}) , \tag{S9}$$

$$\boldsymbol{L}_2^{k+1} = \boldsymbol{L}_2^k + \rho(\boldsymbol{J}_2^{k+1} - \boldsymbol{F}\boldsymbol{S}^{k+1}) . \tag{S10}$$

$\boldsymbol{S}$ is our target variable, i.e., the to-be-reconstructed 3D structure. We need to iterate the algorithm based on the analytical solution until the error falls below the set threshold. Then the algorithm is considered converged, and $\boldsymbol{S}$ is the final reconstruction.

Figure S2 provides an illustration of the optimization. We chose a simulated finely textured saddle surface and extended by adding a background to represent the discontinuous regions in Fig. S2(h). As shown in the Fig. S2(a), the initial surface is all-zero, i.e., a flat surface with no stereo depth information. It's expressed by a zero matrix and the normal gradients in both x and y directions are 0. Figure S2(b–d) show the depth dynamic estimation by iterations. Twenty iterations give a result visually approximate to the ground truth. Error evaluation in Fig. S2(e) also proves the algorithm has converged after 20 iterations, which means Fig. S2(d) is the final reconstruction.

Figure S2(f, g) plotted the depth value variation of Fig. S2(b–d) along the red and green directions in Fig. S2(a), where we can see the algorithm iterates rapidly at the beginning and then slowly approaches convergence. Final comparison is between the reconstruction Fig. S2(d) and ground truth Fig. S2(h), from which we each selected 400 random points and listed in Fig. S2(i). The reconstruction shows high reliability compared with the ground truth.

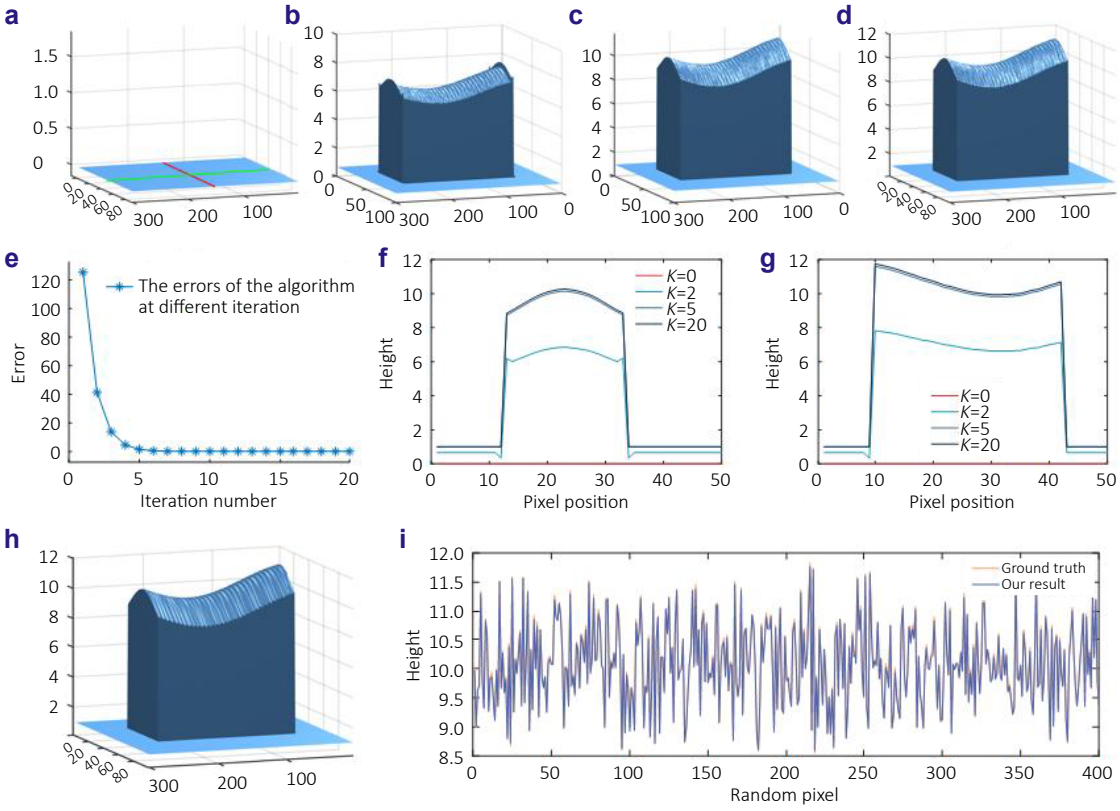


Fig. S2 | Iteration results analyses. (a) Initial. (b) Result at the iteration k=2. (c) Result at the iteration k=5. (d) Result at the iteration k=20. (e) The error after iterations. (f) The slice of (a–d) at along the red line. (g) The slice of (a–d) along the green line. (h) Ground truth. (i) comparison.

Section 4: Point cloud registration of large scenes

The point clouds corresponding to multiple frames are not in the same coordinate system, making point cloud registration necessary. Figure S3 give a set of matched feature points:

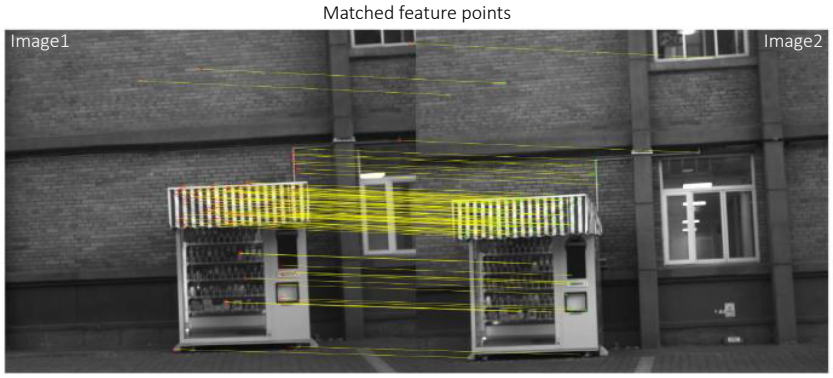


Fig. S3 | Matched feature points from two images.

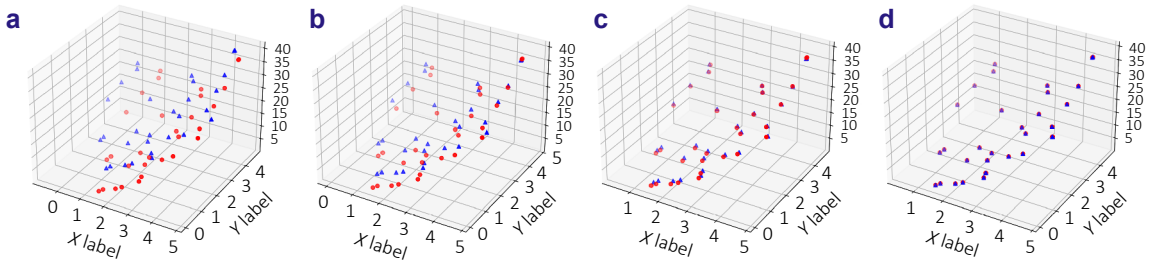
$$\mathrm{P} = \{p_1, \cdots, p_n\}, \quad P' = \{p'_1, \cdots, p'_n\} \quad . \tag{S11}$$

We need to find a Euclidean transformation R and t:

$$\forall i, p_i = Rp'_i + t \; . \tag{S12}$$

The transformation is completed by Iterative Closest Point (ICP) in Eq. (S5) in the main text. By continuously iterating, the algorithm gives the optimal solution at the minimum value.

In Fig. S4(a), the red and blue points represent feature points detected from two sets of point clouds. Figure S4(b) to S4(d) are the results of iterations by ICP. After several iterations, the red and blue point clouds gradually align from an initial misalignment, indicating accurate registration. Figure S4(e) depicts the ground truth of the pose transformation matrix we set, while Fig. S4(f) is the pose transformation matrix from the algorithm. The difference between the two matrices is minimal, with the Mean Squared Error of the error matrix $2.672e^{-10}$. It means the derived pose matrix is reliable for point cloud fusion, as illustrated by Fig. S5. Figure S5(a) is an image series when imaging natural large scene. Figure S5(b) shows the estimated camera poses and the final reconstruction is presented in the main text Fig. S5(c).



e

8.66000000e−01	5.00000000e−01	0.00000000e+00	−3.23210000e+00
−5.00000000e−01	8.66000000e−01	0.00000000e+00	−1.59810000e+00
0.00000000e+00	0.00000000e+00	1.00000000e+00	−1.00000000e+00
0.00000000e+00	0.00000000e+00	0.00000000e+00	−1.00000000e+00

f

8.66025404e−01	5.00000000e−01	−5.23791601e−17	−3.23205081e+00
−5.00000000e−01	8.66025404e−01	−6.06655244e−17	−1.59807621e+00
1.59869241e−15	−5.73623444e−16	1.00000000e+00	−1.00000000e+00
0.00000000e+00	0.00000000e+00	0.00000000e+00	−1.00000000e+00

Fig. S4 | ICP algorithm process. (a) Two sets of unaligned feature point clouds. (b, c) Iterative process. (d) Aligned two sets of feature point clouds. (e) The ground truth of the transformation matrix. (f) Iteration result of transformation matrix.

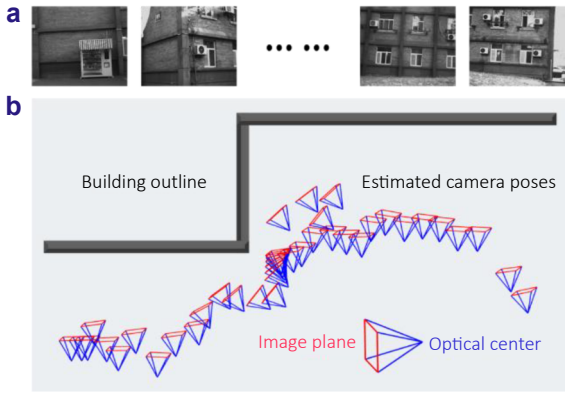


Fig. S5 | Schematic diagram of camera pose estimation. (a) Captured images. (b) The estimated camera poses corresponding to a series of captured images.

Section 5: ADMM parameters determination

The augmented Lagrange function involves three parameters: ξ, η, and ρ. We performed three groups of experiments, with only one parameter varied within each group. Figure S6(a) depicts the ground truth, while Fig. S6(b) and S6(c) show the simulated coarse depth and surface normals from polarization, respectively. By varying each parameter and calculating the SSIM between the simulated results and ground truth, we plotted the curves shown in Fig. S6(d–f). The curves indicate that the algorithm exhibits low sensitivity to weight parameter η but demonstrates considerable sensitivity to parameters ξ and ρ. Based on these tests, we set the parameter values as: ξ=0.2, η=0.1, ρ=0.2.

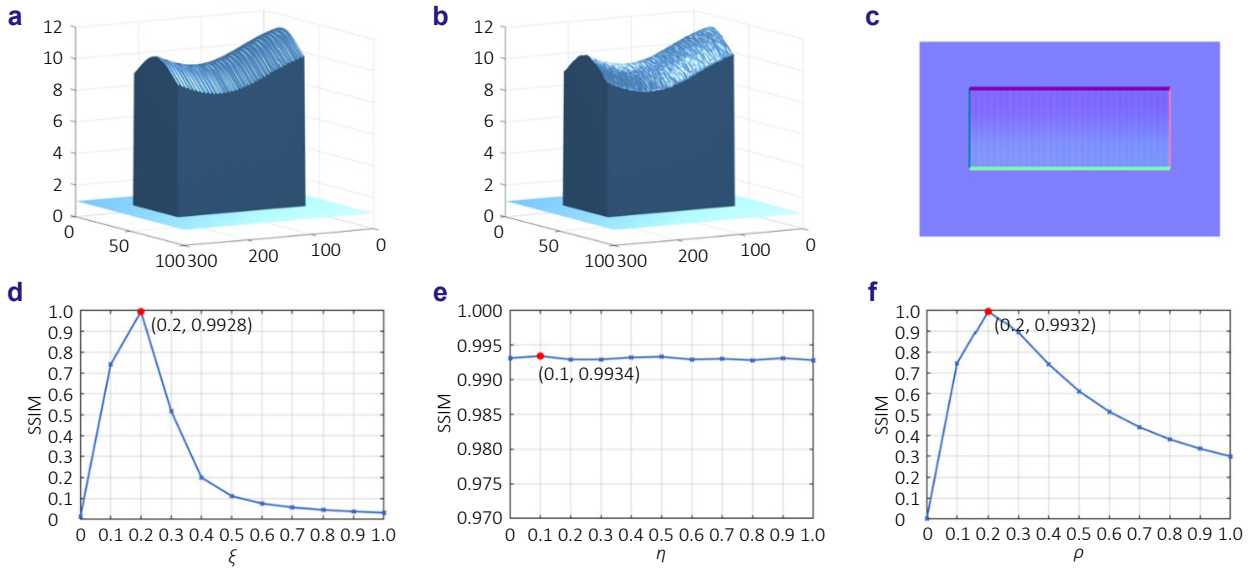


Fig. S6 | Parameters ξ, η, and ρ determination. (a) Ground truth; (b) Rough depth; (c) Surface normal; (d–f) SSIM sensitivity analysis of parameter ξ, η and ρ.