IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 10, 2021, accepted October 2, 2021, date of publication October 6, 2021, date of current version October 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3118345

Joint Dense 3D Reconstruction Method for Endoscopic Images of Weak Texture Scenes

YUNTING GU[1,2], JINGUANG LV[1], JIAN BO[1,2], BAIXUAN ZHAO[1,2], YUPENG CHEN[1,2], JIN TAO[1], YUXIN QIN[1], WEIBIAO WANG[1], AND JINGQIU LIANG[1]

[1]State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin 130033, China

[2]College of Materials Science and Optoelectronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding authors: Jingqiu Liang (liangjq@ciomp.ac.cn) and Jinguang Lv (jinguanglv@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61805239, Grant 61627819, Grant 61727818, and Grant 61905240; in part by Jilin Scientific and Technological Development Program under Grant 20190303063SF and Grant 20180201024GX; in part by the Youth Innovation Promotion Association Foundation of the Chinese Academy of Sciences under Grant 2018254; and in part by the National Key Research and Development Program of China under Grant 2019YFB2006000.

ABSTRACT Endoscopic inspection is an important non-destructive testing method. Traditional 3D endoscopic reconstruction methods, such as polarization reconstruction and shading reconstruction, have the drawbacks of not determining the object's actual size and positional information. The stereo vision method is limited by its operating principles and has the issue of sparse reconstructed point clouds. These drawbacks greatly restrict the applications of the endoscope. Therefore, this work proposes a joint dense 3D reconstruction method for endoscopic images of weak texture scenes. This method uses the shading reconstruction normal to correct the polarization reconstruction normal, then uses coordinate conversion and point cloud fusion to convert the polarization and shading 3D reconstruction results from the pixel coordinate system to the world coordinate system. Finally combines the reconstruction results from the world coordinate system's polarization, shading, and stereo vision. The fusion coefficients are obtained by solving the minimum error model, and then a complete and detailed 3D reconstruction surface was obtained in the world coordinate system. This method could avoid the difficulty of obtaining real coordinates for the 3D reconstruction of polarization and shading and prevent the issue of the sparse point cloud afforded by stereo vision reconstruction for weak texture scenes. The combined dense 3D reconstruction method had an average error of $<1\%$ for length measurement of a 3D curve, which is highly significance for industrial endoscopic inspection.

INDEX TERMS Endoscopic detection, dense point cloud, 3D reconstruction.

I. INTRODUCTION

Endoscopic inspection is a type of non-destructive inspection method that overcomes the visual limitations of the human eye. It can effectively detect damage and defects inside equipment through narrow channels that cannot be directly observed, and it avoids unnecessary damage resulting from disassembly and installation of the equipment [1]. This leads to significant benefits related to reduced testing time and costs. Endoscopic inspections are widely used in many fields, including for electronics, shipbuilding, and machinery manufacturing. Although traditional 2D endoscopic images contain a wealth of information, information on the actual size and depth of the damage and defects cannot be obtained because the image size is in pixels, and the planar image does not contain depth information. Such images are judged based on the experience of the operator. Hence, it is difficult to guarantee the reliability of the endoscopic detection results [2], [3]. Three-dimensional reconstruction is an important research direction in computer vision, which can create a 3D model of a scene based on 2D images. This affords endoscopic detection with more abundant image information, and it is more intuitive for the inspector to analyze a multi-dimensional image.

3D reconstruction methods mainly include polarization [4], shading [5], and stereo vision [6], [7]. The polarization 3D reconstruction method establishes the relationship between the surface normal of the object and the polarization

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan.

information by analyzing the polarization characteristics of the reflected light of the object and then restores the surface shape of the object. The ambiguity of surface normal is the key issue of polarization reconstruction. Huynh *et al.* [8] used wavelength as a constraint. For the first time, proposed a method to obtain the surface normal and refractive index of polarization reconstruction from a single viewpoint simultaneously and used synthetic and real images to prove the method's effectiveness. This method assumed that the surface is convex and used the property that the normal point to the outside to eliminate ambiguity, and calculated the optimal height of the singular point to eliminate the ambiguity of the concave-convex mixed surface. The process was relatively complicated [8]. Smith *et al.* [9] established a linear equation system using polarization images, which solved the problem of surface normal ambiguity while solving the polarization reconstruction surface height. This method was suitable for objects with uniform albedo. It can be used for depth estimation under single, uncalibrated illumination conditions, but it was not precise to estimate the refractive index of the object [9]. The shading 3D reconstruction method uses brightness changes of the measured object surface to obtain 3D shape information of the object. Zhang *et al.* [10] developed a shape restoration algorithm for weakly textured regions. They used the Landsat 7 ETM+ Band 4 image and SRTM 90m global DEM data as reference to evaluate the algorithm. This method required a Lambertian model of the variant albedo map to approximate the reflectivity of the satellite image [10]. Yang *et al.* [11] used synthetic images to train a deep network to achieve shading reconstruction. They used evolutionary algorithms to jointly evolve 3D shapes and train deep networks for the first time. They achieved state-of-the-art performance on authentic images without using an external shape dataset [11]. But whether synthetic image or actual image, lighting [12], [13] and surface albedo [14]–[16] conditions are required. The stereo vision method obtains 2D images of an object via two or more perspectives and uses the principle of triangulation to calculate the 3D shape of the object. Gai *et al.* [17] used the line constraint to reduce the search range, and an ant colony algorithm was used to optimize the stereo matching feature search function in the proposed search range. By establishing the stereo matching optimization model of the ant colony algorithm, the global optimization solution of stereo matching in 3D reconstruction based on binocular vision system was realized. This method could not obtain dense stereo matching results, although it improved the convergence speed and accuracy of the stereo matching process [17]. Hernandez-Beltran *et al.* [18] proposed a parallel implementation method of stereo vision algorithm for 3D scene reconstruction. The algorithm estimated the disparity map from a pair of rectified stereo images using an adaptive template matched filter. The estimated disparity was utilized to retrieve the 3D information by considering the stereo camera's intrinsic parameters. This algorithm improved the accuracy and speed of 3D reconstruction [18].

Polarization-based and shading-based 3D reconstruction methods can only obtain the relative depth in the pixel coordinate system, but they have good detail and dense reconstruction. The stereo vision method is limited by its matching principle [19]–[22], and the reconstruction point cloud of weak texture objects is sparse but can get accurate world coordinates from stereo vision. The complementarity between them has attracted many researchers to research and development. Haines and Wilson [23] proposed a method combining stereo and shading. They used the belief propagation method to obtain the stereo matching result and estimated the reflectance and surface normal vector of each point, and then used iteration to get a continuous estimate of the disparity [23]. Maurer *et al.* [24] proposed a joint variational method that combined SFS cues with minimisation framework. The Lambertian SFS method combined 3D model with the generated energy function, supplemented with the detail-preserving anisotropic second-order smoothing term. It extended the resulting model to estimate depth, albedo, and illuminance jointly. This method is suitable for objects with uneven albedo and scenes with unknown illumination [24].

In order to overcome the limitations of different reconstruction methods and obtain a more accurate dense 3D reconstruction surface in the world coordinate system, this paper proposes a joint dense 3D reconstruction method for endoscopic images of weak texture scenes, which combines SFP, SFS, and stereo vision. The method uses the surface normal of the shading reconstruction reliably and quickly estimates the polarization reconstruction's normal to solve the ambiguity. And use coordinate conversion and point cloud fusion to convert the polarization and shading reconstruction dense point cloud from the pixel coordinate system to the world coordinate system. Then combine the reconstruction results from the world coordinate system's polarization, shading, and stereo vision. The fusion coefficients are obtained by solving the minimum error model. Finally, bring the coefficients into the model to get the joint 3D reconstruction result. This method could realize dense 3D reconstruction of the weak texture target in the world coordinate system. By building an experimental optical platform, the effectiveness of the joint dense reconstruction method was verified.

II. JOINT DENSE 3D RECONSTRUCTION METHOD

The schematic diagram of the joint dense 3D reconstruction method in the world coordinate system is shown in Fig. 1. In the figure, $O_wX_wY_wZ_w$ represents the world coordinate system, which is the reference system for describing the true position of the target object. $O_0X_0Y_0Z_0$ is the object coordinate system, which can rotate with the rotating platform. $O_cX_cY_cZ_c$ is the camera coordinate system with origin at the position of the camera's optical center. *ouv* is the pixel coordinate system of the image, which uses pixels as the unit. o_p *xy* is the physical coordinate system of the image, in which the x axis and the y axis are parallel to the *u* axis and the *v* axis, respectively.

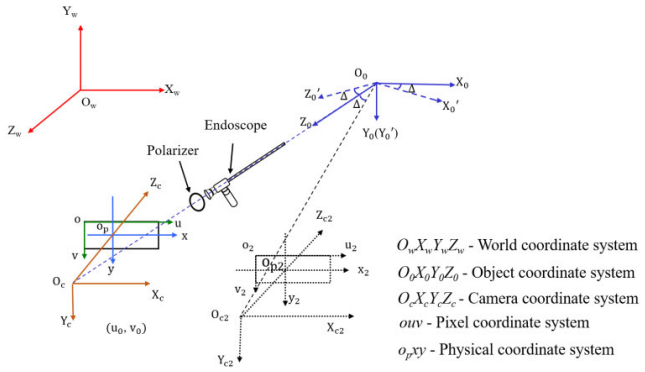

FIGURE 1. Schematic diagram of the joint dense 3D reconstruction method in the world coordinate system.

Step 1: Obtain the shading reconstruction dense point cloud in the pixel coordinate system. Take the horizontal direction as the reference direction, and use the polarizer in the directions 0°, 45°, 90°, and 135° to obtain polarization images I_0, I_{45}, I_{90}, and I_{135}, respectively. Then get the light intensity image $I = I_0 + I_{90}$. Assuming that the imaging geometry is an orthogonal projection, calculate the brightness equations E for the Lambertian surface:

$$E(u, v) = R(p, q) = \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}}, \quad (1)$$

where $E\ (u, v)$ is the brightness of the pixel (u, v), R is the reflectance, p and q are the gradients of $(u, v, z_s\ (u, v))$, $z_s\ (u, v)$ is the shading reconstruction surface, τ and σ are the tilt and slant of the lighting direction. There:

$$p = \frac{\partial z_s(u, v)}{\partial u}, \quad q = \frac{\partial z_s(u, v)}{\partial v}, \quad (2)$$

$$p_s = \frac{\cos\tau \sin\sigma}{\cos\sigma}, \quad q_s = \frac{\sin\tau \sin\sigma}{\cos\sigma}, \quad (3)$$

Tsai and Shah [25] discrete approximation of the surface gradient is:

$$\begin{aligned} p &= z_s(u, v) - z_s(u - 1, v) \\ q &= z_s(u, v) - z_s(u, v - 1), \end{aligned} \quad (4)$$

Equation (1) can be written as:

$$\begin{aligned} 0 &= f(E(u, v), z_s(u, v), z_s(u - 1, v), z_s(u, v - 1)) \\ &= E(u, v) - R(z_s(u, v) - z_s(u - 1, v), \\ &\quad z_s(u, v) - z_s(u, v - 1)), \end{aligned} \quad (5)$$

For a fixed point (u, v) and a given image E, the Jacobian iteration method is used to solve the linear system of the given depth map Z_s^{n-1}, and the Taylor series expansion of the first order f is simplified as:

$$\begin{aligned} 0 = f(z_s(u, v)) &\approx f(z_s^{n-1}(u, v)) \\ &+ (z_s(u, v) - z_s^{n-1}(u, v))\frac{d}{dz_s(u, v)} f(z_s^{n-1}(u, v)), \end{aligned} \quad (6)$$

Let $z_s\ (u, v) = z_s^n\ (u, v)$, the shading 3D reconstruction surface $Z_s\ (u, v)$ in the pixel coordinate system is obtained by iteration:

$$z_s^n(u, v) = z_s^{n-1}(u, v) - \frac{f(z_s^{n-1}(u, v))}{\frac{df(z_s^{n-1}(u,v))}{dz_s(u,v)}}, \quad (7)$$

Then obtain the shading reconstruction point cloud in the pixel coordinate system as (u^r, v^r, z_s^r), where $r = 1, 2, \ldots,$ m, and m is the number of all pixels on the detector.

Step 2: Use the shading reconstruction's surface normal to correct the polarization reconstruction's surface normal and obtain the dense point cloud of the polarization reconstruction in the pixel coordinate system. The degree of polarization ρ and the polarization phase angle ϕ are as follows:

$$\rho = \frac{\sqrt{(I_0 - I_{90})^2 + (I_{45} - I_{135})^2}}{\frac{1}{2}(I_0 + I_{45} + I_{90} + I_{135})}, \quad (8)$$

$$\phi = \frac{1}{2}\tan^{-1}\left(\frac{I_{45} - I_{135}}{I_0 - I_{90}}\right), \quad (9)$$

where I_0, I_{45}, I_{90}, and I_{135} are the intensity of linearly polarized light in the directions of 0°, 45°, 90° and 135°. According to the Fresnel principle, the law of refraction, and the definition of the degree of polarization, we can obtain the relationship between the zenith angle and the degree of polarization, and the azimuth angle and the phase angle of polarization:

$$\rho = \frac{(n - 1/n)^2 \sin^2\theta}{2 + 2n^2 - (n + 1/n)^2 \sin^2\theta + 4\cos\theta\sqrt{n^2 - \sin^2\theta}}, \quad (10)$$

$$\varphi = \phi \ \ or \ \ \varphi = \phi + 180°, \quad (11)$$

where n represents the refractive index of the object, ρ is the degree of polarization, θ represents the zenith angle, φ is the azimuth angle, and ϕ is the polarization phase angle.

The zenith angle θ and the azimuth angle φ jointly determine the surface normal of the target, and the expression of the surface normal is:

$$\begin{aligned} \vec{n} &= \begin{bmatrix} \tan\theta\cos\varphi \\ \tan\theta\sin\varphi \\ 1 \end{bmatrix} = \begin{bmatrix} \tan\theta\cos\phi \\ \tan\theta\sin\phi \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{\partial z_p(u,v)}{\partial u} \\ -\frac{\partial z_p(u,v)}{\partial v} \\ 1 \end{bmatrix}, \end{aligned} \quad (12)$$

Or:

$$\begin{aligned} \vec{n} &= \begin{bmatrix} \tan\theta\cos\varphi \\ \tan\theta\sin\varphi \\ 1 \end{bmatrix} = \begin{bmatrix} \tan\theta\cos(\phi + 180) \\ \tan\theta\sin(\phi + 180) \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial z_p(u,v)}{\partial u} \\ \frac{\partial z_p(u,v)}{\partial v} \\ 1 \end{bmatrix}, \end{aligned} \quad (13)$$

where z_p (u, v) is the polarization reconstruction surface, $\frac{\partial z_p(u,v)}{\partial u}$ and $\frac{\partial z_p(u,v)}{\partial v}$ are the gradients of $(u, v, Z_s(u, v))$.

The formula (12) and (13) shows that there is an ambiguity in the polarization reconstruction surface normal. The prior information of the shading reconstruction surface normal is used to correct the polarization reconstruction surface normal gradient, when:

$$\left|\frac{\partial z_s(u,v)}{\partial u} - \frac{\partial z_p(u,v)}{\partial u}\right| \geq \left|\frac{\partial z_s(u,v)}{\partial u} + \frac{\partial z_p(u,v)}{\partial u}\right|, \quad (14)$$

The polarization and shading reconstruction are in the opposite direction of the gradient of u, requiring a rotation of 180°. Otherwise, they're going in the same direction concerning the gradient of u, so we don't have to change it.

The same can be obtained as follows:

When

$$\left|\frac{\partial z_s(u,v)}{\partial v} - \frac{\partial z_p(u,v)}{\partial v}\right| \geq \left|\frac{\partial z_s(u,v)}{\partial v} + \frac{\partial z_p(u,v)}{\partial v}\right|, \quad (15)$$

The polarization and shading reconstruction are in the opposite direction of the gradient of v, requiring a rotation of 180°. Otherwise, they're going in the same direction concerning the gradient of v, so we don't have to change it.

After obtaining the corrected normal vector, the relative depth of each (u, v) point in the pixel coordinate system can be obtained by integrating, and the polarization reconstruction dense point cloud in the pixel coordinate system is (u^r, v^r, z_p^r) where $r = 1, 2, \ldots, m$, m is the number of all pixels on the detector.

Step 3: Obtain the rotating stereo vision reconstruction sparse point cloud in the world coordinate system. Rotating stereo vision is a special case of traditional stereo vision. A single-camera is used to obtain image from different directions when the motion parameters of the object are known. After a certain angle rotates the object, the polarizer is rotated to obtain the polarization images of 0°, 45°, 90°, and 135° in the second viewing angle, and the total intensity image before and after the object is rotated is used as the image pair for stereo vision reconstruction. The stereo matching algorithm of sift feature realizes the detection and matching of feature points [26]. Assuming the pixel coordinates (u_1^i, v_1^i) of the feature matching point i before rotation, the relationship between the pixel coordinates and the world coordinates is:

$$Z_{c1}\begin{bmatrix} u_1^i \\ v_1^i \\ 1 \end{bmatrix} = M^1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^1 & m_{12}^1 & m_{13}^1 & m_{14}^1 \\ m_{21}^1 & m_{22}^1 & m_{23}^1 & m_{24}^1 \\ m_{31}^1 & m_{32}^1 & m_{33}^1 & m_{34}^1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (16)$$

where M^1 is the projection matrix before rotation, which can be calculated through calibration of the endoscope system. Zc_1 is the Z coordinate in the camera coordinate system. The coordinate of the feature matching point i after rotating by Δ angle is (u_2^i, v_2^i), and its relationship with the world coordinates is:

$$\begin{aligned} Z_{c2}\begin{bmatrix} u_2^i \\ v_2^i \\ 1 \end{bmatrix} &= M^1 \begin{bmatrix} \cos\Delta & 0 & -\sin\Delta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\Delta & 0 & \cos\Delta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= M^2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} m_{11}^2 & m_{12}^2 & m_{13}^2 & m_{14}^2 \\ m_{21}^2 & m_{22}^2 & m_{23}^2 & m_{24}^2 \\ m_{31}^2 & m_{32}^2 & m_{33}^2 & m_{34}^2 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \end{aligned} \quad (17)$$

M_2 is the projection matrix after rotation, *and* Z_{c2} is the Z coordinate in the camera coordinate system. Four equations about X_w, Y_w, Z_w can be obtained by formulas (16) and (17). The least-square method is used to obtain the stereo vision point cloud in the world coordinate system (X^i, Y^i, Z^i) by combining the equations, where $i = 1, 2, \ldots, n$, n is the number of feature matching points.

Step 4: Extract the sparse point cloud corresponding to the feature matching point in the pixel coordinate system. The coordinates obtained by the stereo matching algorithm [26] are floating-point numbers to find the feature matching point. Therefore, the rounding method is used on (u_1^i, v_1^i) to get (u^i, v^i). Then take the z coordinate of the polarization and shading reconstruction in the pixel coordinate system corresponding to the position of (u^i, v^i) to obtain the polarization reconstruction sparse point cloud (u^i, v^i, z_p^i) and the shading reconstruction sparse point in the pixel coordinate system Cloud (u^i, v^i, z_s^i).

Step 5: Coordinate conversion and point cloud fusion. From the definition of pixel coordinate system, physical coordinate system, and camera coordinate system, we can see that their horizontal and vertical axis directions are parallel to each other, $u \,//\, x \,//\, Xc$, $v \,//\, y \,//\, Y_c$. When the world coordinate system is also parallel to the three coordinate axes of the camera coordinate system and the rotation axis $Y_0 \,//\, v \,//\, y \,//\, Y_c$ of the stereo vision, the conversion parameters between the coordinate systems are only translation parameters and scaling factors. Therefore, the least-squares method was used to obtain the conversion relationship of the polarization reconstruction point cloud of the feature matching point from the pixel coordinate system to the world coordinate system:

$$\begin{aligned} X^i &= k_{px}u^i + T_{px} \\ Y^i &= k_{py}v^i + T_{py} \\ Z^i &= k_{pz}z_p^i + T_{pz}, \end{aligned} \quad (18)$$

In the same way, the conversion relationship of the shading reconstruction point cloud of the feature matching point from the pixel coordinate system to the world coordinate system can be obtained:

$$\begin{aligned} X^i &= k_{sx}u^i + T_{sx} \\ Y^i &= k_{sy}v^i + T_{sy} \\ Z^i &= k_{sz}z_s^i + T_{sz}, \end{aligned} \tag{19}$$

where k_{px}, k_{py}, k_{pz} and T_{px}, T_{py}, T_{pz} are the scaling factors and translation parameters, respectively, of the polarization 3D reconstruction coordinate transformation in the three coordinate axis directions. k_{sx}, k_{sy}, k_{sz} and T_{sx}, T_{sy}, T_{sz} are the scaling factors and translation parameters of the shading 3D reconstruction coordinate transformation in the three coordinate axis directions. They are all obtained via least-squares fitting of the coordinate positions of the feature matching points.

Through coordinate conversion, the polarization reconstruction point cloud (X_p^i, Y_p^i, Z_p^i), shading reconstruction point cloud (X_s^i, Y_s^i, Z_s^i), scale scaling factor k and translation parameter T of the feature matching point in the world coordinate system can be obtained. The polarization reconstruction dense point cloud (X_p^r, Y_p^r, Z_p^r) and shading reconstruction dense point cloud (X_s^r, Y_s^r, Z_s^r) of the object in the world coordinate system can be obtained by using the obtained scale scaling factor k and translation parameter T for point cloud fusion. Since the result of polarization reconstruction sparse point cloud (u^i, v^i, z_p^i) and shading reconstruction sparse point cloud (u^i, v^i, z_s^i) are the same in the x and y directions: $k_{px} = k_{sx}$, $k_{py} = k_{sy}$, $T_{px} = T_{sx}$, $T_{py} = T_{sy}$, thus their X and Y coordinates of the matching point after coordinate conversion are the same: $X_p^i = X_s^i$, $Y_p^i = Y_s^i$, and their X and Y coordinates are the same after fusion: $X_p^r = X_s^r$, $Y_p^r = Y_s^r$. Figure 2 shows the coordinate conversion and point cloud fusion diagram.

Step 6: Combine the polarization, shading, and stereo vision in the world coordinate system. Although the two dense point clouds in the world coordinate systems of polarization and shading can be obtained after step 5, we can obtain higher reconstruction accuracy by fusing them, and transform the reconstruction problem into an optimization problem. We join the stereo vision sparse point cloud (X^i, Y^i, Z^i) and the polarization reconstruction sparse point cloud (X^i, Y^i, Z_p^i) and shading reconstruction sparse point cloud (X^i, Y^i, Z_s^i), and estimate the best scale conversion factor. The conversion error in X, Y, and Z direction is minimized. The minimum error function we constructed is:

$$\begin{aligned} E_{error\_X} &= \sum_{i=1}^{n} (X^i - \omega_{pX}X_p^i - \omega_{sX}X_s^i)^2 \\ E_{error\_Y} &= \sum_{i=1}^{n} (Y^i - \omega_{pY}Y_p^i - \omega_{sY}Y_s^i)^2 \\ E_{error\_Z} &= \sum_{i=1}^{n} (Z^i - \omega_{pZ}Z_p^i - \omega_{sZ}Z_s^i)^2, \end{aligned} \tag{20}$$

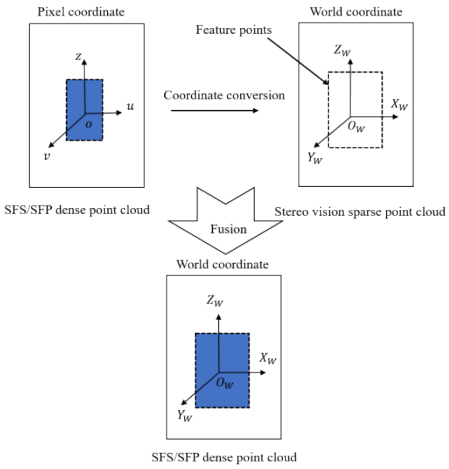

FIGURE 2. Coordinate conversion and point cloud fusion diagram.

where $i = 1, 2, \ldots, n$, n is the number of feature points. ω_{pX}, ω_{pY}, and ω_{pZ} are the scale factors in the X, Y, and Z directions for polarization reconstruction, and ω_{sX}, ω_{sY}, and ω_{sZ} are the scale factors in the X, Y, and Z directions for shading reconstruction. Since $X_p^i = X_s^i$, $Y_p^i = Y_s^i$, we can simplify the X and Y direction of the minimum error function:

$$\begin{aligned} E_{error\_X} &= \sum_{i=1}^{n} (X^i - \omega_{pX}X_p^i - \omega_{sX}X_s^i)^2 \\ &= \sum_{i=1}^{n} (X^i - 2 \cdot \omega_{pX}X_p^i)^2 \\ &= \sum_{i=1}^{n} (X^i - 2 \cdot \omega_{sX}X_s^i)^2 \\ E_{error\_Y} &= \sum_{i=1}^{n} (Y^i - \omega_{pY}Y_p^i - \omega_{sY}Y_s^i)^2 \\ &= \sum_{i=1}^{n} (Y^i - 2 \cdot \omega_{pY}Y_p^i)^2 \\ &= \sum_{i=1}^{n} (Y^i - 2 \cdot \omega_{sY}Y_s^i)^2, \end{aligned} \tag{21}$$

Compare the formula in coordinate conversion:

$$\begin{aligned} X^i &= k_{px}u^i + T_{px} = k_{sx}u^i + T_{sx} \\ Y^i &= k_{py}u^i + T_{py} = k_{sy}u^i + T_{sy}, \end{aligned} \tag{22}$$

The minimum error function and coordinate conversion both perform least-squares on all feature matching points. These formulas are the same, and the XY coordinates of the polarization or shading dense point cloud in the world coordinate system can be used instead of the fusion step.

Multiple linear regression analysis is a statistical analysis method that studies the relationship between a dependent variable and two or more independent variables. Therefore, we can transform the minimum error function in

the Z direction into a multiple linear regression model:

$$Z^i = \omega_0 + \omega_p Z_p^i + \omega_s Z_s^i + \varepsilon, \tag{23}$$

here, ω_0 is a constant term, ω_p and ω_s are regression coefficients in the linear regression model, and ε is a random error with a mean value of 0 and variance of σ^2. The least-squares method was used to obtain the values of ω_0, ω_p, and ω_s, and the linear regression model was obtained. We bring the polarization reconstruction dense point cloud (X_p^r, Y_p^r, Z_p^r) and the shading reconstruction dense point cloud (X_s^r, Y_s^r, Z_s^r) into this model and obtain the joint dense 3D reconstruction point cloud in the world coordinate system:

$$\begin{aligned} Z^r &= \omega_0 + \omega_p Z_p^r + \omega_s Z_s^r \\ X^r &= X_p^r = X_s^r \\ Y^r &= Y_p^r = Y_s^r, \end{aligned} \tag{24}$$

where $r = 1, 2, \ldots, m$, m is the number of all pixels on the detector, and (X^r, Y^r, Z^r) represents the joint dense point of the object in the world coordinate system.

III. 3D CURVE LENGTH MEASUREMENT

The size parameters of the object can be obtained by measuring the length of the 3D curve, and the size of internal defects in industrial equipment can be determined.

Traditional stereo vision endoscopic detection cannot provide complete point cloud information of the object, and there are only a few characteristic points that can be obtained via curve measurement. Hence, it is difficult to obtain accurate curve lengths. Current endoscopic length detection only measures the depth of a certain point of the object to the camera or the linear distance between two characteristic points on the object. The joint dense 3D reconstruction method proposed in this paper can be used to obtain the dense point cloud in the world coordinate system that corresponds to each pixel; additionally, the reconstructed object information is more accurate than traditional stereo vision, and the measurement results are more accurate. Figure 3 is a schematic diagram of the 3D curve length measurement.

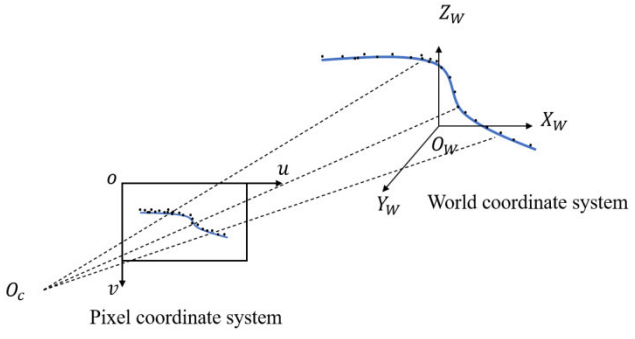

FIGURE 3. Schematic diagram of 3D curve length measurement.

The specific measurement steps are as follows. First, the light intensity image of the target object is obtained, and the Canny edge detection method [27] is used to extract the curve. Then, the joint dense 3D reconstruction method is used to map each pixel on this curve to the world coordinate system to obtain the point cloud of the curve, where $j = 1, 2, \ldots, l$, and l is the number of pixels in the detection curve. The polynomial fitting method is used to fit the 3D curve of the point cloud. During the fitting process, the segmented fitting method is used. The order of the polynomial is constantly adjusted to prevent under- or over-fitting and ensure the accuracy of the fitting results. Finally, the Euclidean distance calculation method is used to obtain the length of the curve L:

$$L = \sum_{j=1}^{l-1} \sqrt{(X_{j+1} - X_j)^2 + (Y_{j+1} - Y_j)^2 + (Z_{j+1} - Z_j)^2}, \tag{25}$$

IV. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL SYSTEM CONSTRUCTION AND ENDOSCOPE SYSTEM CALIBRATION

The experimental system was then built to verify the effectiveness of the joint dense 3D reconstruction method, as shown in Fig. 4. The system mainly includes: an endoscope system, cold light source, Thorlabs WP25M-VIS wire grid polarizer, and the object to be measured. The endoscope system was composed of a SONY XCG-C130 detector, an industrial endoscope, and an adapter lens.

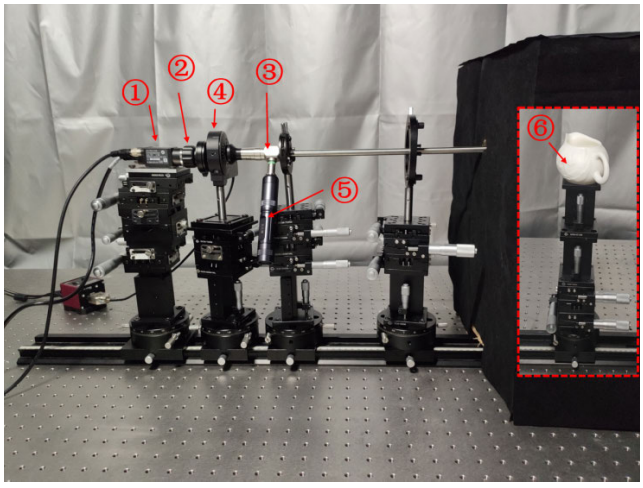

FIGURE 4. ① Detector; ② Adapter lens; ③ Endoscope; ④ Rotating mount with wire grid polarizer; ⑤ Cold light source; and ⑥ Object to be measured.

The experimental system was used to calibrate the monocular endoscope, and calibration was primarily carried out for the internal parameters and distortion coefficients. This paper adopts a calibration method based on a 2D checkerboard calibration board [28], which has the advantages of convenient implementation and high calibration accuracy. First, the checkerboard is located in the plane of the world coordinate system $Z = 0$, then, the world coordinates of any corner point on the checkerboard is $(X, Y, 0)$, and conversion

is carried out between the pixel coordinate system of this point and the world coordinate system:

$$\mu \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = A \begin{bmatrix} r_1 & r_2 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{26}$$

here, μ is the scale factor. $(u, v, 1)$ are the image pixel coordinates expressed in homogeneous coordinates. $(X, Y, Z, 1)$ are the world coordinate expressed in homogeneous coordinates. A represents the internal parameter matrix of the camera, R represents the rotation matrix, r_i is the ith column vector of R, and t represents the translation matrix. The camera internal parameter matrix A is:

$$A = \begin{bmatrix} f_u & 0 & u_0 \\ 1 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{27}$$

here, f_u and f_v are the pixel metric representations of the focal length of the endoscope system in the x and y directions, and (u_0, v_0) are the principal point coordinates of the camera. Formula (26) is described by the homography matrix H:

$$\mu \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{28}$$

H can be obtained using the points corresponding to the checkerboard plane and the imaging plane. Because the rotation matrix is an orthogonal matrix, according to the limited condition that the inner product of any two column vectors of H is 0 and the module is 1, the least-squares solution of the equation obtained from multiple calibration images was obtained via singular value decomposition (SVD), and each internal parameter of the endoscope system was obtained.

The endoscope system has a small size and a large field of view. Hence, each image must be undistorted after the endoscope is calibrated. The correction of radial distortion is as follows:

$$\begin{aligned} x_{dr} &= x(1 + s_1 r^2 + s_2 r^4) \\ y_{dr} &= y(1 + s_1 r^2 + s_2 r^4), \end{aligned} \tag{29}$$

here, (x, y) and (x_{dr}, y_{dr}) are the undistorted image coordinates and the distorted image coordinates, respectively, s_1 and s_2 are the radial distortion parameters, and r is the distance from the corresponding pixel to the image center, where $r^2 = x^2 + y^2$.

We constructed a checkerboard pattern calibration board with 9*6 corner points, as shown in Fig. 5.

The obtained 10 images of the calibration board were selected for calibration, and the obtained calibration parameters of the endoscope system are shown in Table 1.

The reprojection error method was used to judge the calibration accuracy of the endoscope system parameters, and the average reprojection error was calculated to be 0.21 pixels, which meets the calibration requirements. The calibration reprojection error of the endoscope system is shown in Fig. 6.

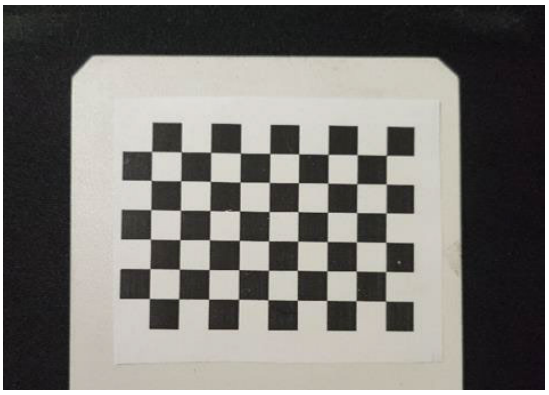

FIGURE 5. Calibration board image.

TABLE 1. Camera calibration parameters.

Parameter	Value
f_u/pixels	2145.9222
f_v/pixels	2131.7660
u_0/pixels	797.4471
v_0/pixels	434.8818
s_1/pixels	-0.4988
s_2/pixels	0.1947

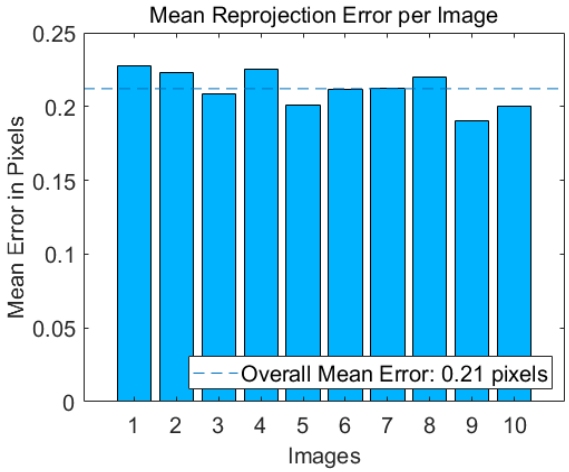

FIGURE 6. Reprojection error of the endoscope system.

B. JOINT DENSE 3D RECONSTRUCTION EXPERIMENT

The experimental system was used to investigate a white jade teacup, and the joint dense 3D reconstruction algorithm was verified. First rotate the polarizer to obtain polarization images of 0°, 45°, 90°, and 135°, and then calculate the intensity image, polarization degree image, and polarization phase angle image. Use the shading reconstruction algorithm for the light intensity image to obtain the shading 3D reconstruction surface in the pixel coordinate system. Use the shading reconstruction surface normal to correct the polarization

reconstruction normal and obtain the polarization 3D reconstruction surface in the pixel coordinate system, as shown in Figure 7.

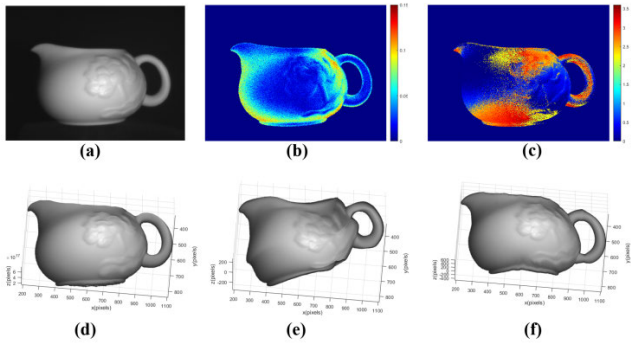

FIGURE 7. (a) Intensity image of the teacup. (b) Degree of polarization. (c) Polarization phase angle. (d) Shading reconstruction surface in pixel coordinate system. (e) Polarization reconstruction surface of uncorrected normal in pixel coordinate system. (f) Polarization reconstruction surface after correcting normal in pixel coordinate system.

The object is rotated at a certain angle, and the polarizer is also rotated to obtain polarization images at different angles. Before and after the object is rotated, the intensity image is used as the image pair for stereo vision. In the experiment, 1463 pairs of feature matching points are obtained by using the stereo matching algorithm based on sift features [26]. The stereo vision point cloud in the world coordinate system calculated using equations (16) and (17) is shown in Fig. 8.

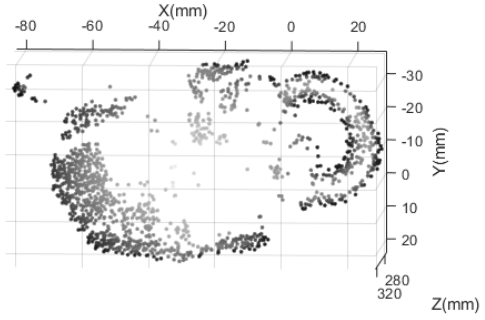

FIGURE 8. Stereo vision reconstruction point cloud.

We use the polarization reconstruction and shading reconstruction coordinates of the feature points in the pixel coordinate system to perform least-squares fitting with the coordinates in the world coordinate system and obtain the scaling factor and the translation parameter in the three directions of *XYZ*. The partial figure of feature matching points of stereo vision and polarization reconstruction in the Z direction is shown in Fig. 9:

Then, coordinate conversion and fusion are performed on the polarization reconstruction point cloud and the shading reconstruction point cloud. The obtained polarization reconstruction surface and the shading reconstruction surface in the world coordinate system are shown in Fig. 10.

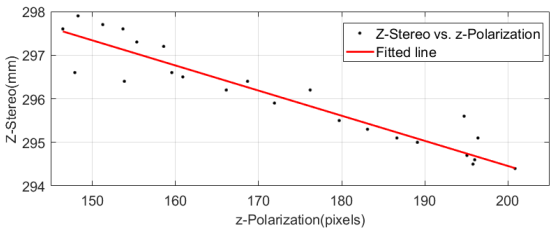

FIGURE 9. Figure of feature matching points of stereo vision and polarization reconstruction in the Z direction.

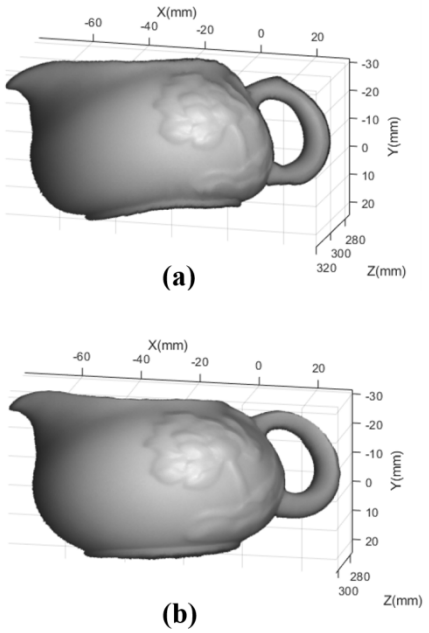

FIGURE 10. (a) Polarization 3D reconstruction surface in the world coordinate system after coordinate conversion. (b) Shading 3D reconstruction surface in the world coordinate system after coordinate conversion.

After coordinate conversion, the Z coordinates of the feature matching points of stereo vision are used as the dependent variable, and the Z coordinates of the polarization reconstruction and shading reconstruction as the independent variables. By inputting them into Eqn. (24) we can obtain the coefficients of the linear regression model. The coefficients obtained are $\omega_0 = 0$, $\omega_p = 0.5701$, and $\omega_s = 0.4388$, and the final joint dense 3D reconstruction surface of the object in the world coordinate system is shown in Fig. 11.

The established multiple linear regression equation in the Z direction is:

$$Z^i = 0.5701 \times Z_p^i + 0.4388 \times Z_s^i, \tag{30}$$

To verify the accuracy and rationality of the equation, the regression equation needs to be tested as follows:

(1) *F Inspection:* At a significance level α, calculate the F value. If $F > F_\alpha\ (k, N-k-1)$, then all the independent variables significantly impact the dependent variable as a whole, and the assumption of the linear relationship of the

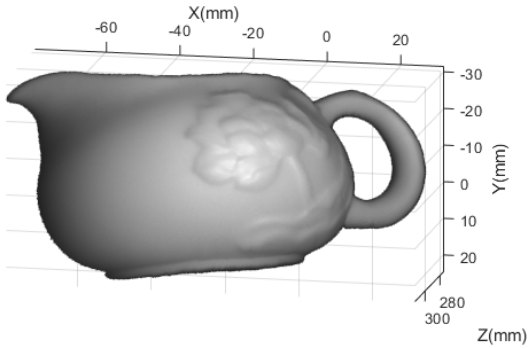

FIGURE 11. Joint dense 3D reconstruction surface.

regression equation is significant. Otherwise, the regression effect is considered to be insignificant, and the independent variable needs to be determined again. Here, N is the number of samples, and k is the number of independent variables.

(2) *t-Test:* At the significance level α, the t value is calculated. If $|t| > t_{\alpha/2}\ (N-k-1)$, It is considered that each independent variable has a significant impact on the dependent variable. Otherwise, it is considered insignificant, and the variable is excluded from the equation.

(3) *Goodness of Fit Test:* Calculate the coefficient of determination R^2 to evaluate the goodness of fit of the regression equation to the sample observations, $R^2 = RSS/TSS$, where TSS is the total sum of squares, and RSS is the regression sum of squares.

The numerical table of the analysis of variance of the regression model in this experiment is shown in Table 2:

TABLE 2. Numerical table of analysis of variance.

Source	Sum of Squares	Degree of Freedom	Mean Square	F-value	P-value
Model	93903.6302	2	46951.8151		
Error	99289.2977	1460	68.0064	690.4029	0
Corrected Total	193192.9279	1462	—		

At a significance level α, $P < 0.05$ is considered as ideal. The significance level is 0.05 in this paper, and the P value of the model is 0.

After calculation, the $F = 690.4029$, $t_1 = 7.6332$, $t_2 = 4.2352$, they are all greater than the critical value: $F > F_{0.05}(2, 1460) = 3$, $t1 > t2 > t_{0.05/2}(2, 1460) = 2.920$, the coefficient of determination R^2 is 0.4861. The residual scatter plot is evenly distributed around 0, and the residual histogram is normally distributed, so the regression model is reasonable overall. Part of the residual plot and the histogram of residual distribution are shown in Figure 12:

C. ANALYSIS OF 3D RECONSTRUCTION RESULTS

The accuracy and validity of the joint dense 3D reconstruction results were verified through qualitative analysis and quantitative analysis.

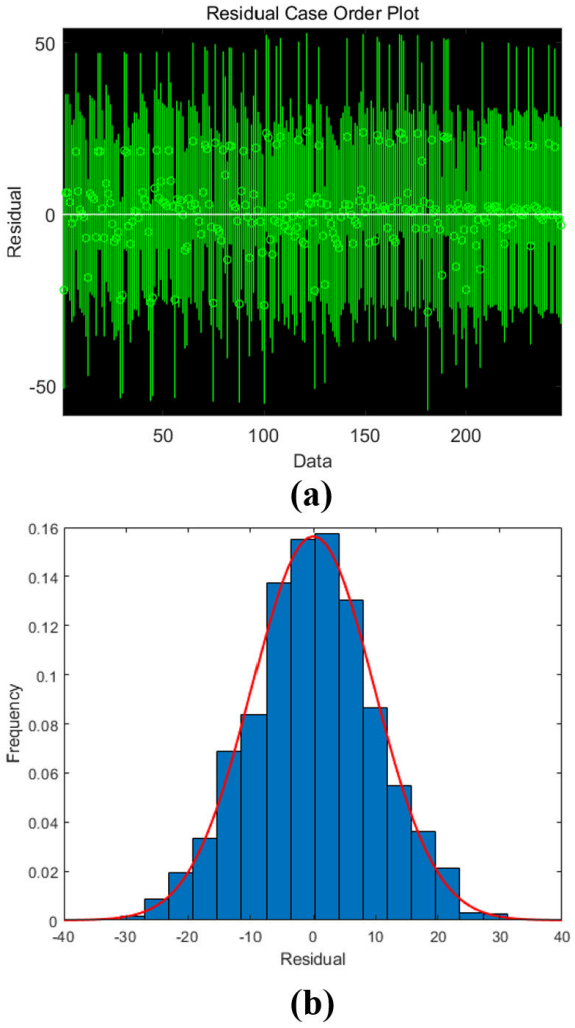

FIGURE 12. (a) Residual plot. (b) Histogram of residual distribution.

1) QUALITATIVE ANALYSIS

The 3D reconstruction surface in the world coordinate system was used to compare the combined dense 3D reconstruction method qualitatively in this paper. The local 3D surface shape was compared with the overall height in the Z direction, as shown in Fig. 13. Where H_1 and H_2 are the heights of the cup body and the handle in the Z direction of the shading 3D reconstruction, respectively, and H_3 and H_4 are the heights of the cup body and the handle in the Z direction of the joint dense 3D reconstruction, respectively.

Although both polarization-based reconstruction and shading-based reconstruction can restore the shape of the teacup surface, the polarization-based 3D reconstruction method is sensitive to noise at positions with a low polarization degree [4], [8] and a mutual reflection on the reconstructed surface [4], which results in some parts of the polarization reconstruction being distorted, as seen in Fig. 12(a). The shading-based 3D reconstruction method needs to make assumptions about the lighting conditions [12], [13] and surface albedo [14]–[16]. Because of

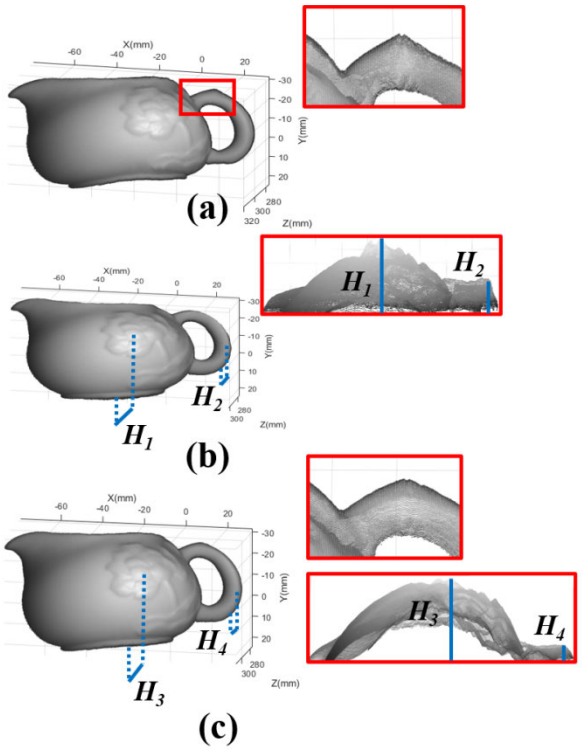

FIGURE 13. (a) Polarization reconstruction surface. (b) Shading reconstruction surface. (c) Joint dense 3D reconstruction surface.

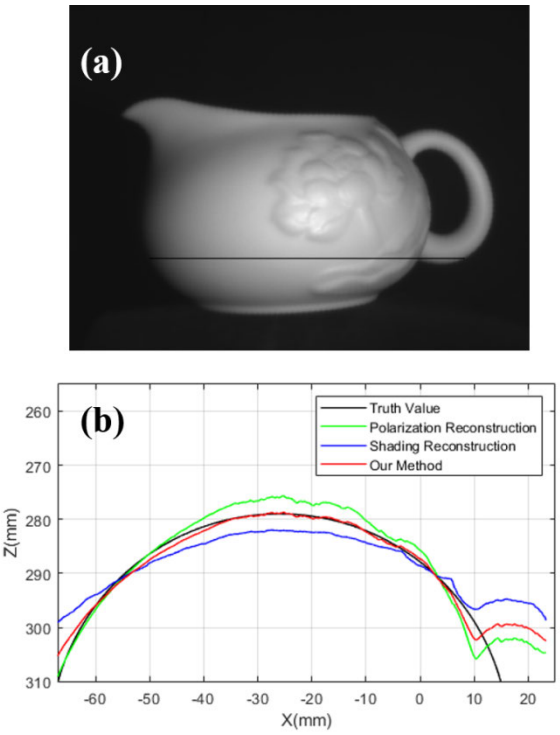

FIGURE 14. (a) Schematic diagram of cross-sectional position. (b) Comparison of cross-sections.

the diversity of the actual reflections and the difficulty of guaranteeing parallel light source conditions during endoscopic measurement [29], [30], a proportion of the shading reconstruction results are distorted in the Z direction, i.e., the height difference between H_1 and H_2 is small. In areas where the brightness is high and specular reflection and diffuse reflection coexist, the reconstructed size in the Z direction would be enlarged, as shown in Fig. 12(b). By combining stereo vision with polarization and shading reconstruction and constructing a minimum error model, the obtained 3D reconstruction surface is closer to the actual result, as shown in Fig. 12(c). It can be seen from the figure that the joint dense 3D reconstruction method has less distortion than the polarization reconstruction method, and the overall Z-direction ratio is more accurate than that determined via shading reconstruction.

2) QUANTITATIVE ANALYSIS

A cross-sectional view was used for comparison to quantitatively compare the joint dense 3D reconstruction surface and the actual surface, as shown in Fig. 14.

Since each horizontal section of the teacup is a standard circle except for the handle and pattern, find the position where the teacup pattern is obvious, and the horizontal circle pattern is as few as possible so that it is easy to take points and measure as shown in Fig. 14. Measure the center position and diameter of the teacup and make a standard circle as the true value. We take X at positions -60, -50, -40, -30, -20, and -10. The calculated average deviation between the

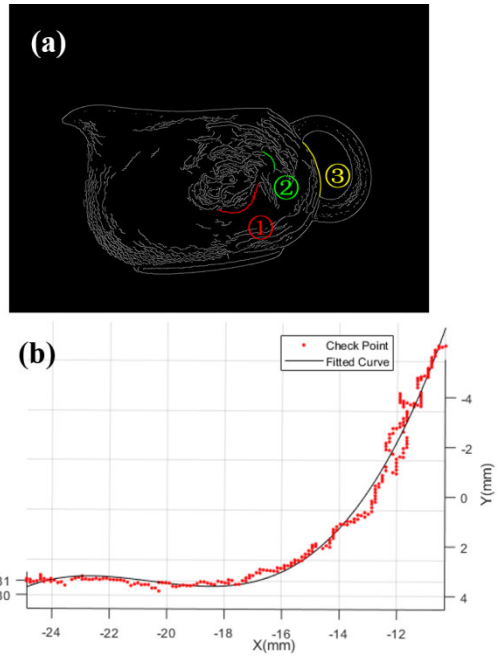

FIGURE 15. (a) Image of three edges to be measured. (b) 3D curve fitting image of one edge.

joint dense 3D reconstruction surface and the real surface is ∼0.38 mm, better than the reconstruction results for polarization and shading. It proves that the 3D shape is well restored by the joint 3D reconstruction method.

TABLE 3. Curve length analysis table of teacup.

Curve	Polarization /mm	Shading /mm	Joint Dense /mm	Actual Length /mm
1	20.5067	20.2181	20.3587	20.4
2	9.0106	8.2949	8.5758	8.7
3	17.9037	17.7455	18.3203	18.1
Mean Error	1.7259%	2.5022%	0.9490%	—

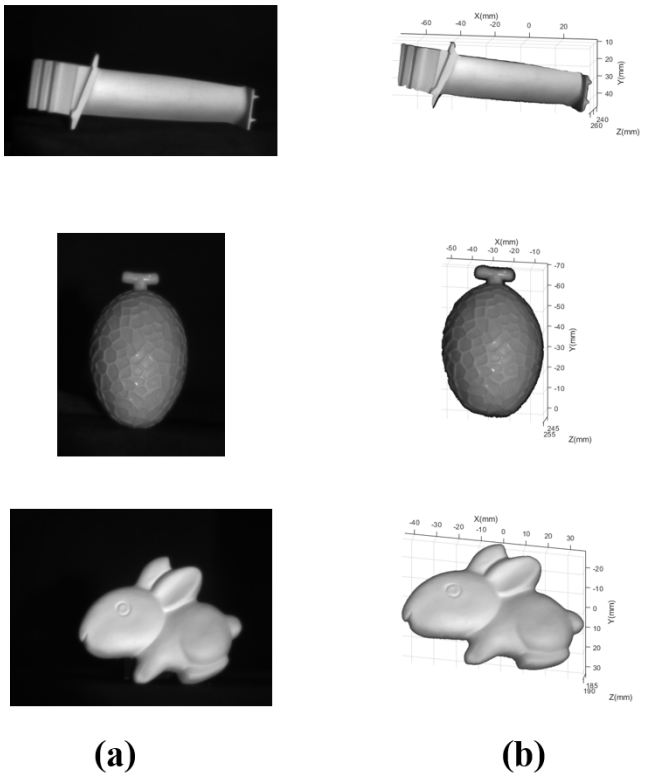

FIGURE 16. (a) Objects of various materials and shapes. (b) Reconstructed surface using the joint dense 3D reconstruction method.

The Canny edge detection method [27] with a low error rate, high localization, and minimum response characteristics was used to obtain the curve to be measured. The method was used to extract 3 edges at different positions and for different lengths of the teacup; then, the 3D polynomial fitting method was used to fit the curve and calculate the length. Figure 15 shows the image of the three edges of the teacup to be measured and a 3D curve fitting image of one edge.

We compare the curve length values obtained via the single-use polarization reconstruction method, the single-use shading reconstruction method in the world coordinate system after coordinate conversion, and the joint dense 3D reconstruction method with the actual lengths. The results of the curve length measurements are shown in Table 3.

It can be seen from Table 3 that the curve length obtained via the joint dense 3D reconstruction method is closer to the actual value. The error was the smallest compared with the polarization and shading reconstruction methods, and the mean error was $<1\%$.

Figure 16 shows the joint dense 3D reconstruction method applied to objects of various materials and shapes, and the

FIGURE 17. The intensity image of the scene.

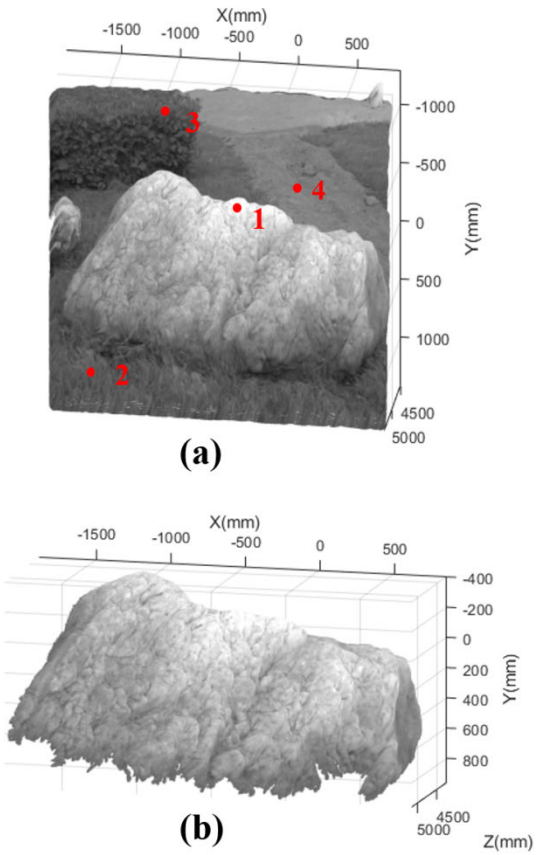

FIGURE 18. (a) The 3D reconstructed surface of the scene. (b) Part of the reconstructed surface extracted from the scene.

results show that the surface shape could be well restored in the world coordinate system.

To verify the joint dense 3D reconstruction method's effectiveness in complex environments, we carry out outdoor experiments with the sun as the light source, and the intensity image is shown in Fig. 17.

The reconstructed surface obtained by the joint dense 3D reconstruction method is shown in Fig. 18.

Fig. 17(a) is the 3D reconstruction surface of the scene. There are different substances in the scene, and the height of 3D reconstruction surface is different. In 17(a), four locations are marked as 1234, and their heights are about 750mm, 300mm, 500mm and 400mm. The material properties can be distinguished by using the characteristics of different reconstruction heights of different materials. The key research objects can also be distinguished and analyzed separately, as shown in Figure 17(b).

V. CONCLUSION

This paper proposed a joint dense 3D reconstruction method for endoscopic images of weak texture scenes. The method could handle weak texture targets well and obtain the 3D surface shape of the target in the world coordinate system. It addressed the inability to get real-world coordinates when using only polarization or shading 3D reconstruction methods and the issue of a sparse point cloud in the weak texture part when using the stereo vision method. The target surface obtained via the joint dense 3D reconstruction method in this paper was close to the actual object shape, and the mean error of the 3D length measurement of the curves was $\leq 1\%$. The joint dense 3D reconstruction method can be applied to complex scenes with different materials and different shapes.

Our method used multiple linear regression methods to combine three different 3D reconstruction results, and while the restored 3D reconstruction surface was closer to the actual situation. Still, it did not eliminate the influence of distortion because we're combining them all in proportion. In the future, we plan to divide the object into several parts with different characteristics and then combine polarization, shading, and stereo vision in those parts to realize a more accurate 3D reconstructed surface.

REFERENCES

[1] J. F. Guo, L. Peng, and J. Guohua, "Binocular optical system of industrial endoscope for three-dimensional measurement," *Opt. Precis. Eng.*, vol. 22, no. 9, pp. 2337–2344, 2014.

[2] H. M. Park and K. N. Joo, "Endoscopic precise 3D surface profiler based on continuously scanning structured illumination microscopy," *Current Opt. Photon.*, vol. 2, no. 2, pp. 172–178, 2018.

[3] S.-P. Yang, J.-J. Kim, K.-W. Jang, W.-K. Song, and K.-H. Jeong, "Compact stereo endoscopic camera using microprism arrays," *Opt. Lett.*, vol. 41, no. 6, pp. 1285–1288, 2016.

[4] G. A. Atkinson and E. R. Hancock, "Recovery of surface orientation from diffuse polarization," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1653–1664, Jun. 2006.

[5] N. Al-Najdawi, H. E. Bez, J. Singhai, and E. A. Edirisinghe, "A survey of cast shadow detection algorithms," *Pattern Recognit. Lett.*, vol. 33, no. 6, pp. 752–764, 2012.

[6] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 4104–4113.

[7] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski, "Building Rome in a day," *Commun. ACM*, vol. 55, no. 10, pp. 105–112,2011.

[8] C. P. Huynh, A. Robles-Kelly, and E. R. Hancock, "Shape and refractive index from single-view spectro-polarimetric images," *Int. J. Comput. Vis.*, vol. 101, no. 1, pp. 64–94, Jan. 2013.

[9] A. P. W. Smith, R. Ramamoorthi, and S. Tozza, "Linear depth estimation from an uncalibrated, monocular polarization image," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, 2016, pp. 109–125.

[10] Y. Zhang and J. Peng, "Surface shape estimation of textureless area using shape from shading for landsat imagery," *Proc. SPIE*, vol. 9259, Nov. 2014, Art. no. 92591S.

[11] D. Yang and J. Deng, "Shape from shading through shape evolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3781–3790.

[12] Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz, "The visual Turing test for scene reconstruction," in *Proc. Int. Conf. 3D Vis. (3DV)*, Seattle, WA, USA, Jun. 2013, pp. 25–32.

[13] M. Zollhöfer, A. Dai, M. Innmann, C. Wu, M. Stamminger, C. Theobalt, and M. Nießner, "Shading-based refinement on volumetric signed distance functions," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 1–14, Jul. 2015.

[14] D. A. Forsyth, "Variable-source shading analysis," *Int. J. Comput. Vis.*, vol. 91, no. 3, pp. 280–302, Feb. 2011.

[15] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *Proc. CVPR*, Providence, RI, USA, Jun. 2011, pp. 20–25.

[16] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt, "High-quality shape from multi-view stereo and shading under general illumination," in *Proc. CVPR*, Providence, RI, USA, Jun. 2011, pp. 969–976.

[17] Q. Y. Gai, "Optimization of stereo matching in 3D reconstruction based on binocular vision," in *Proc. 2nd Int. Conf. Commun., Image Signal Process. (CCISP)*, vol. 960, 2018, pp. 1–7.

[18] J. E. Hernandez-Beltran, M. O. Lugo-Aguilar, V. H. Diaz-Ramirez, and R. Juarez-Salazar, "Stereo vision parallelization for three-dimensional reconstruction," *Proc. SPIE*, vol. 11509, Aug. 2020, Art. no. 1150906.

[19] D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai, and T.-J. Cham, "Recovering surface details under general unknown illumination using shading and coarse multi-view stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1526–1533.

[20] J. Jiang, L. Liu, R. Fu, Y. Yan, and W. Shao, "Non-horizontal binocular vision ranging method based on pixels," *Opt. Quantum Electron.*, vol. 52, no. 4, p. 223, Apr. 2020.

[21] T. P. Nguyen and J. W. Jeon, "Wide context learning network for stereo matching," *Signal Process., Image Commun.*, vol. 78, pp. 263–273, Oct. 2019.

[22] J. Huo and X. Yu, "Three-dimensional mechanical parts reconstruction technology based on two-dimensional image," *Int. J. Adv. Robotic Syst.*, vol. 17, no. 2, Mar. 2020, Art. no. 172988142091000.

[23] T. S. F. Haines and R. C. Wilson, "Combining shape-from-shading and stereo using Gaussian-Markov random fields," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1064–1067.

[24] D. Maurer, Y. C. Ju, M. Breuß, and A. Bruhn, "Combining shape from shading and stereo: A joint variational method for estimating depth, illumination and albedo," *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1342–1366, Dec. 2018.

[25] T. Ping-Sing and M. Shah, "Shape from shading using linear approximation," *Image Vis. Comput.*, vol. 12, no. 8, pp. 487–498, Oct. 1994.

[26] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[27] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[29] E. Prados and O. Faugeras, "Shape from shading: A well-posed problem?" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern*, vol. 2, no. 1, Jun. 2005, pp. 870–877.

[30] H.-S. Ng, T.-P. Wu, and C.-K. Tang, "Surface-from-gradients without discrete integrability enforcement: A Gaussian kernel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2085–2099, Nov. 2010.

YUNTING GU was born in Shuangyashan, Heilongjiang, China, in 1993. She received the B.S. degree from Jilin University, China, in 2015. She is currently pursuing the Ph.D. degree with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. Her main research interests include image processing and 3D reconstruction.

JINGUANG LV received the B.S. degree from the School of Physics, Jilin University, in 2004, and the Ph.D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2013. He is currently an Associate Professor and a Master Supervisor with the State Key Laboratory of Applied Optics of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His research interests include micro-optical systems, optical information processing, and coherent spectral imaging.

JIAN BO was born in Gongzhuling, Jilin, China, in 1994. He received the B.S. degree from Harbin University of Technology, Weihai Campus, China, in 2016. He is currently pursuing the Ph.D. degree in optics with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His current research interest includes development of polarization spectral imager.

BAIXUAN ZHAO was born in Changchun, Jilin, China, in 1993. He received the B.S. degree in physics from Jilin University, China, in 2016. He is currently pursuing the Ph.D. degree in optics with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. His current research interest includes spectral data processing.

YUPENG CHEN was born in Huludao, Liaoning, China, in 1996. He received the B.S. degree from Dalian University of Technology, China, in 2017. He is currently pursuing the Ph.D. degree with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China. His research interests include infrared optics design and physical optical analysis.

JIN TAO received the B.S. and Ph.D. degrees in measurement technology and instrument from Tianjin University, in 2010 and 2016, respectively. He is currently a Research Associate with the State Key Laboratory of Applied Optics of Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. He is currently committed to the development and application of micro-LED chips and the research of uncooled infrared detection chips based on micro nano resonators.

YUXIN QIN was born in Mudanjiang, Heilongjiang, in 1985. He received the B.S. degree in optical information science and technology and the M.S. degree in optical engineering from Harbin Engineering University, in 2008 and 2011, respectively. Since 2011, he has been working with the State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. His main research interests include the experimental research of infrared spectroscopy instrument optical design and assembly adjustment.

WEIBIAO WANG was born in Yangzhou, Jiangsu. He received the doctor's degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 1999. He is currently a Researcher with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. He is also a Doctor, a Researcher, and a Doctoral Supervisor. His research interests include photonic crystal and micro-nano photonics, LED array chip integration and application, field emission materials, and electron emission characteristics.

JINGQIU LIANG received the B.S. degree in professional semiconductor physics and devices from the Department of Electronic Science, Jilin University, in 1984, and the Ph.D. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, in 2003. She is currently a Professor with Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences. Her research interests include micro-optical electro mechanical system (MOEMS), micro-structure optics, micro-Fourier transform spectrometer, Fourier transform imaging spectrometer, micro-LED arrays, and devices in visible light communication systems.