

Three-dimensional reconstruction based on binocular structured light with an error point filtering strategy

Xinyu Li, Kai Yang,* Yingying Wan, Zijian Bai^{ORCID}, Yunxuan Liu, Yong Wang, and Liming Xie

Southwest Jiaotong University, School of Physical Science and Technology, Chengdu, China

ABSTRACT. Gray code assisted phase shifting technology can achieve robust and noise-tolerated three-dimensional (3D) shape measurements. To solve the issues of unsynchronized brightness changes, local overexposure, and edge coding errors caused by inconsistent reflectivity of the surface in complex industrial scenes, as well as defocusing caused by noncontinuous surfaces and varying distances, we combine the advantages of a large imaging range in passive stereo vision and high precision in active structured light imaging. It uses a consumer-grade projector to project gray code and stripe patterns, whereas two precalibrated color industrial cameras capture raw images and obtain the original channel data. Gray code and reverse gray code images are projected to solve the problems of binarization and boundary blur. In addition, an error point filtering strategy is proposed to retain pixels with decoding errors of less than two bits. The use of softargmin for subpixel matching of absolute phase results in a high precision disparity map. We present a simple and high precision 3D measurement system for industrial objects. Experiments on 3D measurements in complex industrial scenes showed that the proposed method can achieve high precision and robust 3D shape measurements.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.OE.63.3.034102](https://doi.org/10.1117/1.OE.63.3.034102)]

Keywords: structured light; gray code; binocular; subpixel

Paper 20231108G received Nov. 22, 2023; revised Feb. 2, 2024; accepted Feb. 28, 2024; published Mar. 12, 2024.

1 Introduction

Optical 3D measurement technology has the advantages of being noncontact, having high precision and fast speed, and being applicable to static and dynamic measurements; it is widely used in fields such as machine vision, industrial inspection, biometrics, and reverse engineering.¹⁻⁴ Binocular vision and structured light are important methods of optical 3D measurement technology.

Binocular stereo vision⁵ uses the matching of corresponding points of scenes captured by two cameras at different angles to obtain the disparity and then converts it into the 3D information of the scene. Although the corresponding points must be on the epipolar line, the two images can be matched by defining the similarity of the window and sliding the window to find the corresponding points.⁶ The horizontal pixel difference between the two corresponding points is called the disparity. Binocular matching depends on the texture information and surface features of the corresponding points, which limits the dense and accurate reconstruction of the real 3D scene with less texture.

*Address all correspondence to Kai Yang, yangkai_swjtu@163.com

Structured light replaces one of the cameras with a projector and projects sinusoidal fringes or gray codes onto the tested object. The camera captures the deformed pattern modulated by the object's height, and the depth information is calculated based on the principle of triangulation.⁷ The gray code algorithm is simple and robust, but it requires the projection of multiple frames of coded patterns. The stripe projection has a high spatial resolution, but the phase obtained by the phase shifting algorithm is between $(-\pi, \pi)$, which needs phase unwrapping. The spatial unwrapping algorithm is only suitable for flat surfaces, whereas the temporal unwrapping algorithm requires the projection of more patterns. Both algorithms have limitations and difficulty meeting the real-time measurement requirements.^{8,9}

In addition, both the binocular vision and structured light methods are susceptible to the surface reflectivity of the tested object and ambient light. The dense reconstruction of objects in complex lighting conditions has gradually become a research hotspot.¹⁰⁻¹²

Sun et al.¹³ used gray code assisted phase shifting technology and the additional projection of a complementary gray code pattern to use the two decoding results to correct the error position and solve the problems of false edges and mismatched wrapped phase. However, this method is not suitable for scenes with inconsistent reflectivity. Wu et al. re-encoded traditional gray codes in the time and space domains and used cyclic complementary gray codes¹⁴ and moving gray codes,¹⁵ combined with binary defocusing phase shifting technology, to achieve high-speed dynamic measurement of a fan and falling blocks. However, the defocusing technology limits the application of noncontinuous surfaces in industrial scenes and reduces the precision in the Z direction. The robustness of this method in complex scenes also needs to be verified. Lohry et al.¹⁶ used a binocular structured light method, first using binocular stereo matching to obtain a rough disparity map and then using locally wrapped phase information to further refine the disparity map for higher precision. However, the signal-to-noise ratio is low when measuring steep surfaces, and there are large shadowed areas, which cannot meet the measurement requirements of industrial scenes. Lu et al.¹⁷ also proposed a method based on phase shifting profilometry and stereo vision measurement systems. They used constraints from matched raw images to obtain a rough disparity and used subpixel disparity optimization to reduce matching errors. However, the process of matching wrapped phases is easily affected by inconsistent reflection in the scene, and it is difficult to perform filtering, making it challenging to achieve 3D imaging in complex scene. Yu et al.¹⁸ added a set of low-frequency fringes on top of gray code assisted phase shifting technology, effectively correcting the period jump error of the stripes and allowing for the measurement of surfaces with drastic height changes. However, in scenes with a large depth of field, the imaging precision often decreases for objects that are not in the focal plane. Hu et al.¹⁹ used the high dynamic range imaging surface 3D measurement method based on adaptive stripe projection to dynamically adjust the brightness of the projected stripes by establishing coordinate mapping between the camera and the projector. The surface highlights of complex watch parts and mobile phone parts are avoided, and high-quality point clouds are obtained. Chen et al.²⁰ used the sampling moiré fringe method based on binocular vision to improve the speed of phase matching. The fixed-point iterative method solved the problem of large deformations caused by uneven grating distribution in 3D measurements, allowing for the measurement of Poisson's ratio during the deformation of a stretched cylinder. Yuan et al.²¹ established the optimal projection strategy and the coordinate mapping between the camera and the projection and combined it with the response function of the local camera. They used a binary search method to determine the optimal projection brightness in the overexposed area, effectively avoiding the phase error and enabling an accurate measurement of metal workpieces and metal plates. Engel²² summarized various 3D measurement methods, among which the binocular structured light method uses gray code assisted phase shifting technology and stereo matching based on the absolute phase of the left and right images. It is not affected by the color and texture information of the measured object's surface and is less affected by ambient light. The phase solved by the left and right cameras only contains the height information of the object, resulting in higher matching precision and a shorter processing time, making it suitable for 3D imaging in complex industrial scenes.

The methods proposed in the above literature often achieve 3D measurements of simple objects under laboratory settings with good lighting conditions, and there are many limitations in extending them to industrial scenes. Our approach focuses on monitoring and maintenance of

critical components of high-speed rails, such as measuring the wheel size, detecting the missing bolts, identifying damage and missing parts in pipelines, and assisting with the positioning of mechanical arms, as well as detecting pantographs and lifting arms. Each of these tasks has different requirements for the range, precision, and speed of 3D measurement. The imaging scenes in this paper are more diverse, with significant variations in surface reflectivity and more complex lighting conditions, which place higher demands on the imaging accuracy and robustness of our proposed method.

This paper optimizes multiple key steps of binocular structured light imaging. To address the issue of the lower image quality of color cameras compared with black and white cameras, raw images are captured to improve the image quality. To tackle the challenges of complex lighting conditions, large measurement range, and long depth of field in industrial scenes, gray codes and inverse gray codes are used for projection, and an error point filtering strategy is proposed to effectively select the masked region, thereby improving the system's robustness. The use of sub-pixel stereo matching results in refined disparity values, achieving 3D reconstruction of large scenes. Experimental results demonstrate the strong robustness and high precision of the proposed method in complex industrial scenes.

2 Principle

2.1 Phase Shifting Profilometry

Figure 1 shows the basic process of 3D reconstruction using stereo structured light. This section will provide a brief introduction to the key principles involved.

Phase shifting profilometry²³ projects $N(N \geq 3)$ sinusoidal stripe patterns with equal phase shifts within one cycle onto the surface of the measured object, and the camera captures the deformed fringes to accurately solve the phase information modulated by the height of the measured object's surface. The five-step phase-shifting algorithm is used in this paper, and the captured images are represented as

$$I_n(u, v) = A(u, v) + B(u, v) \cos[\varphi(u, v) + \delta_n], \quad (1)$$

$$\delta_n = 2\pi \cdot \frac{(n-1)}{N} \quad (n \leq N), \quad (2)$$

where $A(u, v)$ is the background light intensity, $B(u, v)$ is the modulation, $\varphi(u, v)$ is the phase modulated by the height information of the object, and δ_n is the phase shift amount. The corresponding wrapped phase expression is

$$\varphi(u, v) = \arctan \frac{\sum_{n=1}^N I_n(u, v) \sin \delta_n}{\sum_{n=1}^N I_n(u, v) \cos \delta_n}, \quad (3)$$

where the wrapped phase is calculated using the four-quadrant inverse tangent function and the phase $\varphi(u, v)$ is truncated between $(-\pi, \pi)$ and needs to be unwrapped to restore a continuous phase. Gray code assisted phase unwrapping is used in this paper because it is fast and simple and does not suffer from error propagation.

The gray code method uses a set of binary coded gratings to mark the sinusoidal patterns, with M gray codes marking 2^M fringe cycles. The distribution of 4-bit gray code words, for example, is shown in Fig. 2.

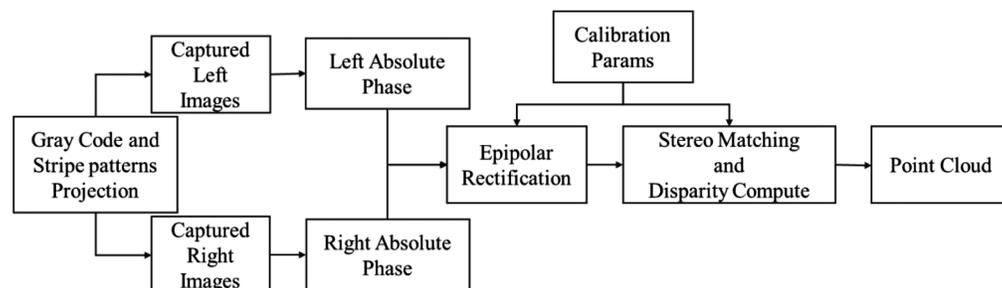


Fig. 1 Basic process of the binocular structured light.

GC1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
GC2	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
GC3	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0
GC4	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Fig. 2 4-bit Gray code word distribution diagram.

In Fig. 2, GC1, GC2, GC3, and GC4 represent the horizontal gray values of the four projected gray code patterns, with 0 indicating a gray value of 0 and 1 representing a gray value of 255. k represents the order of each gray code. Decoding gray codes first requires binarizing the captured images, with the binarization threshold determined by the average of the five captured stripe images, which is given as

$$H(n) = \frac{1}{5} \sum_{i=1}^5 I_i(n), \quad (4)$$

where $H(n)$ is the binarization threshold for a single pixel and $I_i(n)$ is the grayscale value when projecting sinusoidal fringes.

The conversion between gray code and binary code is given by

$$B(n) = -B(n+1) \text{ XOR } G(n), \quad (5)$$

where $B(n)$ is the n 'th bit of the gray code, $B(n+1)$ is the $(n+1)$ 'th bit of the binary code, and XOR is the exclusive OR operation.

The absolute phase of the left and right images is determined by

$$\Phi(u, v) = \varphi(u, v) + 2\pi k(u, v), \quad (6)$$

where $\Phi(u, v)$ is the absolute phase, $\varphi(u, v)$ is the wrapped phase calculated by Eq. (3), and k is the decimal gray code level.

2.2 Binocular Stereo Vision

After calibration of the binocular cameras, the diagonal camera is simplified to a parallel camera as shown in Fig. 3, where O_l and O_r represent the optical centers of the left and right cameras, respectively; x^l and x^r represent the pixel projections of point p in space onto the left and right cameras, respectively; T is the baseline distance between the left and right cameras; and f is the camera's focal length. The depth Z of the point p in space is obtained from the following equation:²⁴

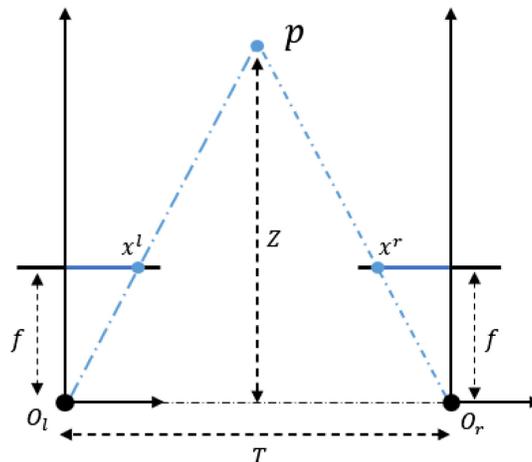


Fig. 3 Schematic diagram of the binocular stereo vision.

$$Z = \frac{T \cdot f}{x^l - x^r} = \frac{T \cdot f}{d}. \quad (7)$$

Equation (7) is derived from a simple similar triangle, where d is the disparity.

After performing epipolar line correction on the absolute phase of the left and right images, the disparity is determined by

$$e(u, v, d) = \min\{|I_l(u, v) - I_r(u + d, v)|, \text{maxdisp}\}, \quad (8)$$

where e represents the disparity of each point in the left image; I_l and I_r represent the pixel values of the left and right images, respectively; and maxdisp is the estimated disparity value.

The absolute phase calculated by Eq. (6) is a double-precision floating point value, which allows for subpixel matching. After finding the integer pixel of the disparity shift,²⁵ linear interpolation is usually used to calculate the exact disparity as shown in Fig. 4

$$\Delta\tau = \frac{\Phi^l(u, v) - \Phi^r(u + \tau, v)}{\Phi^r(u + \tau + 1, v) - \Phi^r(u + \tau, v)}, \quad (9)$$

where $\Phi(v)$ represents the absolute phase of the epipolar line v ; $\Delta\tau$ represents the subpixel disparity shift along the line; and the superscripts l and r indicate the left and right cameras, respectively.

Finally, the disparity map $d(u, v)$ is obtained as

$$d(u, v) = e(u, v, d) + \Delta\tau. \quad (10)$$

With the disparity map, the 3D information of the measured object is obtained using the calibrated camera parameters.

3 Improved Method

3.1 Error Point Filtering Strategy

The consumer-grade projector that we used has a lateral resolution of 1280 and requires at least 2^{10} pixels for encoding. To achieve a stripe projection with a period of 16 pixels, this paper uses 7-bit gray code and five-step phase-shifting stripe patterns, as shown in Fig. 5. Seven additional black and white reverse gray code patterns are also projected for an error point filtering strategy.

The usual approach to solving the problem of level edge code errors is to use complementary gray code. However, due to the ambient light, different surface reflectivity of objects, and interference between projected lights in complex industrial scenes, there may be some areas where the projected bright fringes are darker than the dark fringes. Meanwhile, the captured intensities seldom change in the areas far away in the scene. In addition, the defocusing caused by non-continuous surfaces also increases the difficulty of decoding the gray code and solving for the continuous phase. In this situation, the use of the complementary gray code is no longer applicable.

To solve these special issues, we project both gray code patterns and reverse gray code patterns onto the scene and then decode using the corresponding two sets of images.

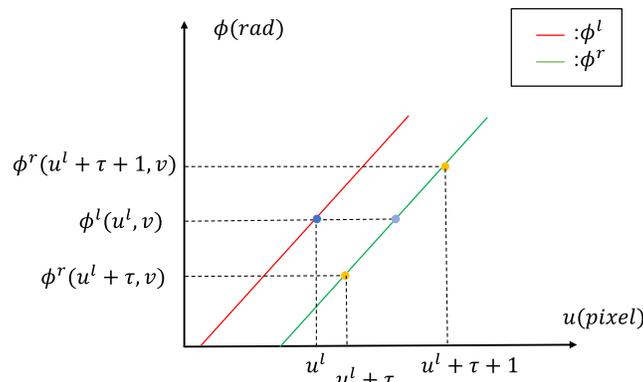


Fig. 4 Schematic diagram of the subpixel interpolation.

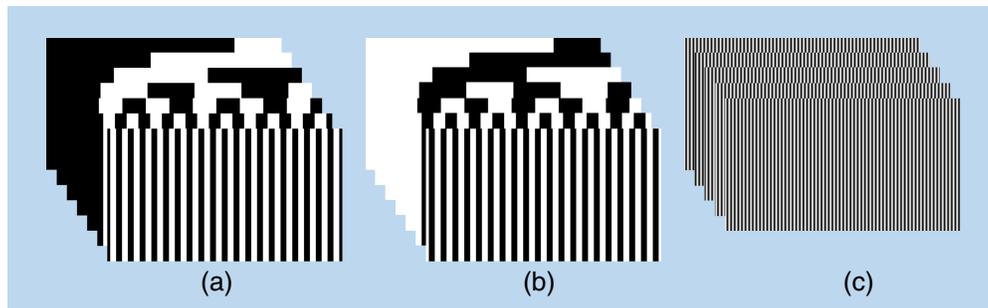


Fig. 5 Projected patterns: (a) gray code patterns, (b) reverse gray code patterns, and (c) stripe patterns.

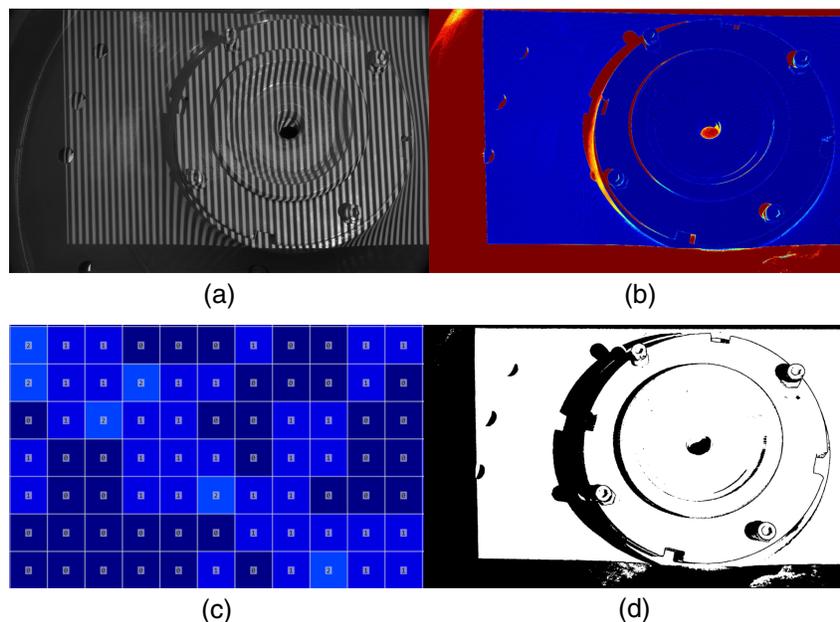


Fig. 6 Decoding of the image sequence on the side of the train axle: (a) raw image of the side of the train axle, (b) error code statistics of 7-bit gray code, (c) local enlarged view, and (d) mask.

The threshold in Eq. (4) is used for binarization. Only the gray code decoding values of a pixel that satisfy Eq. (9) are considered reliable:

$$|I(u, v)_+ - I(u, v)_-| > 15, \quad (11)$$

where $I(u, v)_+$ is the pixel grayscale value of the captured gray code pattern and $I(u, v)_-$ is the pixel grayscale value of the projected reverse gray code pattern. We need to perform seven checks for Eq. (9) on each pixel. This allows us to determine the number of incorrect bits for each pixel when solving the 7-bit gray code.

The decoding method using an error point filtering strategy is shown in Fig. 6. We first need to synchronously collect the gray code and stripe patterns modulated by the scene to be tested, as shown in Fig. 6(a). The number of incorrect bits during the decoding process is mapped, as shown in Fig. 6(b). The pixel values in this image represent the reliability of the current code value. Fig. 6(c) shows a partial enlarged view of Fig. 6(b), with most of the pixels having < 2 incorrect bits. We consider pixels with more than two decoding errors to be unreliable and remove them by adding a mask, as shown in Fig. 6(d).

3.2 Subpixel Stereo Matching

The traditional subpixel interpolation method relies on the monotonic distribution of adjacent pixel values, which is not applicable in high-noise scenes. The resolution of the projector is usually lower than that of the camera, which may result in three pixels corresponding to the

same gray code order during the decoding process. Although these pixels have different wrapped phases, the absolute phase difference between pixels with the same order is less than 2π . This can lead to incorrect matches during disparity calculation and cause clustering in the 3D point cloud. Therefore, we first need to calculate the absolute phase distribution of pixels with the same gray code order and then consider how to obtain the subpixel disparity.

The softargmin function was proposed by GC-Net²⁶ to solve the problem of discrete disparity in stereo matching, which cannot be used for subpixel estimation and differentiation with the argmin operation, making it impossible to backpropagate. The cost of calculating the absolute phase in this paper is a unimodal distribution, making it feasible to use the softargmin function to solve for subpixel disparity:

$$\text{softargmin} = \sum_{d=0}^{D_{\max}} d \times \sigma(-c_d), \quad (12)$$

where softargmin is the weighted sum of the disparity value d , c_d is the cost value for each disparity d , and σ is the softmax function, which is used to convert inputs into probability values.

4 Results and Discussion

The experimental setup includes a consumer-grade projector (XGIMI Z6X) and a color industrial camera (Basler ace acA1920-40gc). The used projector has a severe nonlinear response and poor contrast. Therefore, we employed the five-frame phase-shifting algorithm to reduce higher-order harmonic components and suppress the nonlinear response. The projector has a resolution of 1280 pixels \times 720 pixels, and the camera has a resolution of 1920 pixels \times 1200 pixels. The projected fringe period is 16, the baseline of the binocular camera is 165 mm, and the lens focal length is 12 mm. The measurement distance is 0.5 to 1.5 m.

4.1 Standard Sphere Precision Verification Experiment

To demonstrate the effectiveness of the proposed method in improving precision, a traditional binocular structured light method²⁷ and the proposed method are used to measure a standard sphere and a comparison and precision evaluation is performed.

The measurement results of standard sphere are shown in Fig. 7. Figure 7(a) represents the raw image of the standard sphere captured by the left camera, Fig. 7(b) represents the disparity

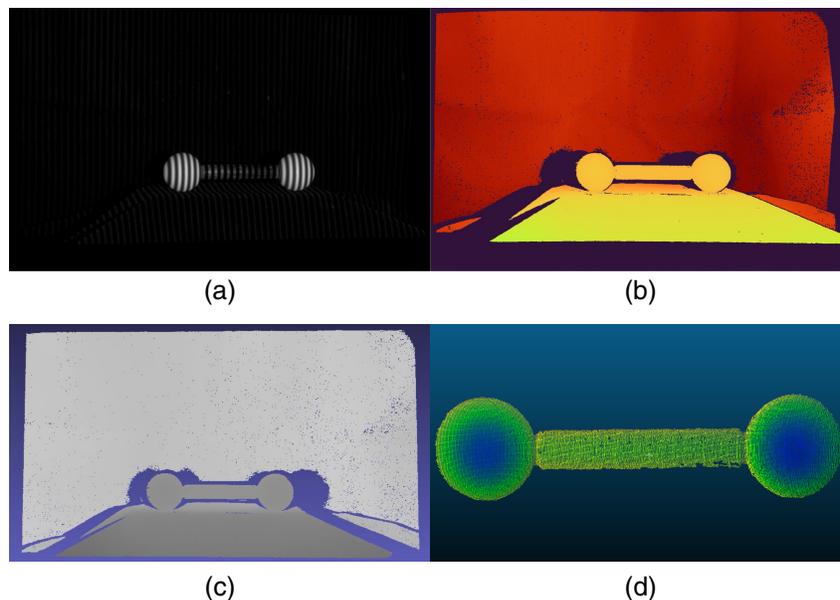


Fig. 7 Standard sphere experiment for the proposed method in this paper: (a) raw image of the standard sphere, (b) disparity map of the standard sphere, (c) point cloud of the standard sphere, and (d) fitted standard sphere.

Table 1 Comparison of precision in the standard sphere experiment.

	Left sphere radius (mm)	Left sphere RMS (mm)	Right sphere radius (mm)	Right sphere RMS (mm)	Sphere center distance (mm)
Standard sphere	25.4	0	25.4	0	150
Before improvement	25.57	0.2787	25.17	0.3279	151.69
Ours without Raw	25.51	0.1398	25.49	0.1504	150.52
Ours	25.43	0.1098	25.41	0.1359	150.48

map of the standard sphere, Fig. 7(c) represents the point cloud obtained by the improved method, and Fig. 7(d) represents the fitted standard sphere.

The scene is surrounded by a black curtain, but the proposed method can still recover the 3D information of the curtain from this dark environment, demonstrating its strong robustness. In Table 1, the precision of the proposed method in measuring the standard sphere is around 0.03 mm with a measurement error of 0.5 mm for the distance between the sphere centers, which is slightly lower than other experiments. This is due to using a low-cost consumer-grade projector, which has a much lower linearity and uniformity compared with industrial projectors. Additionally, the distance between the standard sphere and the camera is 500 mm, which means that one pixel on the camera represents ~ 0.3 mm in space. Achieving a measurement precision of 0.03 mm demonstrates the subpixel matching capability of our method. It is possible to significantly improve this metric using better equipment.

4.2 Comparison Experiment in Various Industrial Scenarios

4.2.1 Comparison experiment in a nonuniform reflectance scene

The measurement results of the train wheel are shown in Fig. 8. Figures 8(a) and 8(d) represent the Rgb image and raw image of the train wheel captured by the left camera, respectively; Figs. 8(b) and 8(e) represent the disparity map of the wheel obtained by the previous method and the proposed method, respectively; and Figs. 8(c) and 8(f) represent the point clouds obtained by the previous and proposed methods, respectively. The distance between the wheel tread and the axle is 0.6 m, and their reflectivity is different. Using the error point filtering strategy, the main area is preserved, and subpixel interpolation is used to achieve an accurate 3D

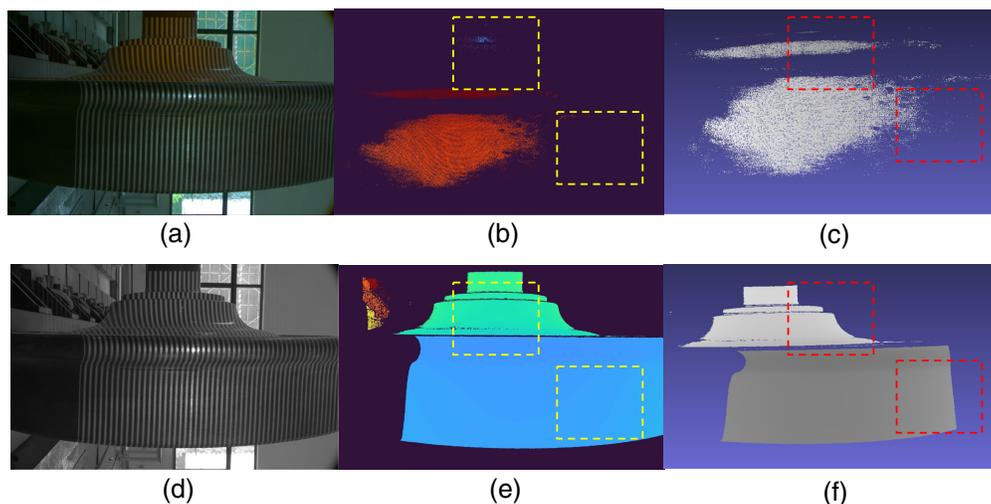


Fig. 8 Comparison experiment of train wheel imaging: (a) Rgb image of the wheel, (b) disparity map of the wheel before improvement, (c) point cloud of the wheel before improvement, (d) raw image of the wheel, (e) disparity map of the wheel after improvement, and (f) point cloud of the wheel after improvement.

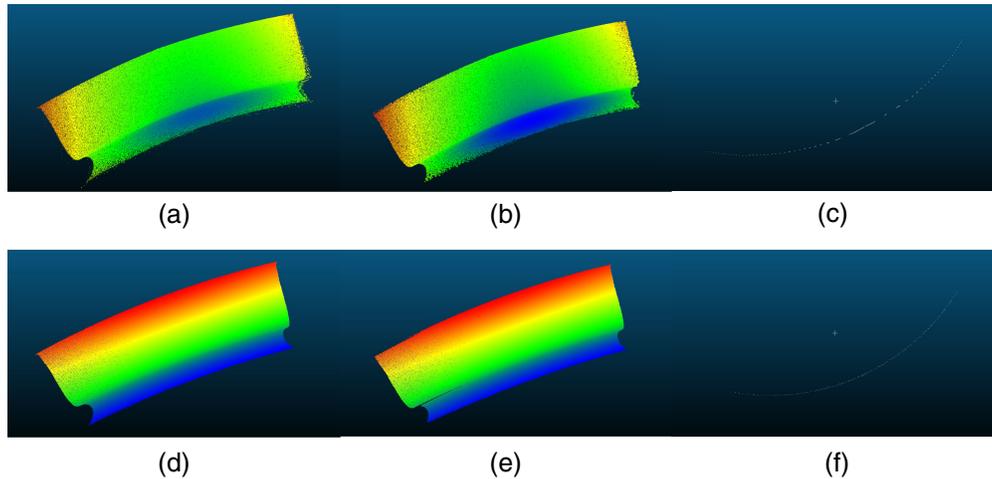


Fig. 9 Comparison experiment of the train wheel size measurement: (a) cropped point cloud obtained by the multifrequency heterodyne method, (b) filtered point cloud obtained by the multifrequency heterodyne method, (c) extracted base point set obtained by the multifrequency heterodyne method, (d) cropped point cloud obtained by our proposed method, (e) filtered point cloud obtained by our proposed method, and (f) extracted base point set obtained by our proposed method.

reconstruction of the train wheel. In comparison with Figs. 8(b) and 8(e), the previous method can only reconstruct the high-brightness parts of the tread and obtain sparse disparity, whereas the improved method can reconstruct the complete wheel and axle disparity model and obtain a dense point cloud with no obvious clustering.

The point clouds obtained by the multifrequency heterodyne method and our proposed method for measuring the same wheel tread under the same experimental environment are shown in Fig. 9. Figures 9(a) and 9(d) represent the point cloud after being cropped to retain only the main body of the wheel. Figs. 9(b) and 9(e) show the point cloud after the same filtering operation. Figures 9(c) and 9(f) represent the extracted base point set, which are points 70 mm away from the inner side of the wheel and can be used to determine the wheel's radius. Comparing Figs. 9(a) and 9(d), the point cloud obtained by the multifrequency heterodyne method has many noise points on both sides of the wheel tread, whereas the point cloud obtained by our proposed method has a higher quality. After the filtering operation, the point cloud shown in Fig. 9(b) becomes sparse, and many valid points are removed. Comparing Figs. 9(d) and 9(e), the number of points obtained by our proposed method is not significantly reduced after filtering, demonstrating the reliability of our point cloud data.

The radius of the standard wheel used in our experiments is 420 mm. As shown in Table 2, our proposed method has a more accurate measurement of the wheel radius and a lower error rate compared with the multifrequency heterodyne method. Additionally, our method obtains a higher number of (points) and a higher proportion of valid points, which is consistent with the conclusions in Fig. 9.

The measurement results of the robotic arm are shown in Fig. 10, which is a scene that contains objects with varying reflectivity. Figures 10(a) and 10(d) represent the Rgb image and

Table 2 Comparison of precision in the wheel size measurement experiment.

	Wheel radius (mm)	Radius measurement error (%)	Number of points before filtering	Number of points after filtering	Proportion of valid points (%)
Multifrequency heterodyne method	412.27	1.84	904,684	855,940	94.61
Ours	416.38	0.86	948,408	934,585	98.54

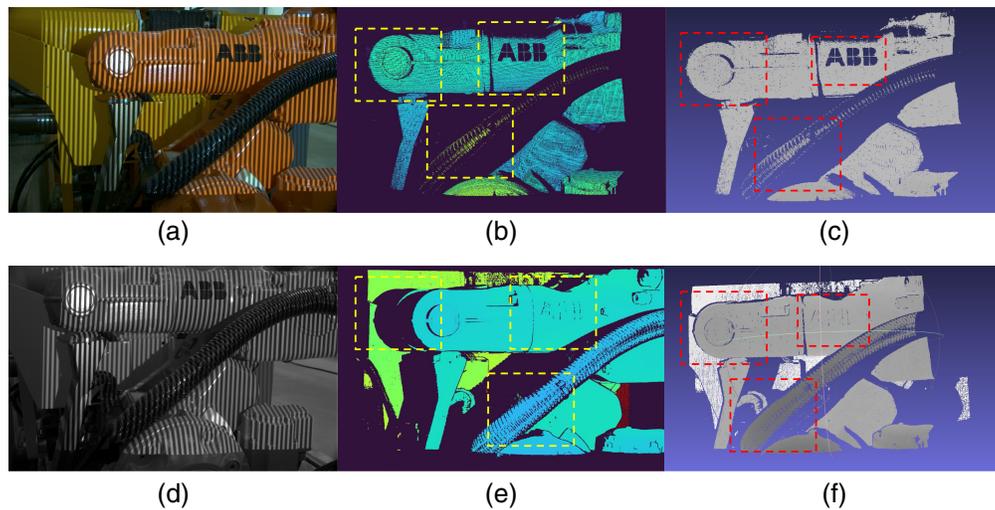


Fig. 10 Comparison experiment of robot arm imaging: (a) Rgb image of the robot arm, (b) disparity map of the robot arm before improvement, (c) point cloud of the robot arm before improvement, (d) raw image of the robot arm, (e) disparity map of the robot arm after improvement, and (f) point cloud of the robot arm after improvement.

raw image of the robotic arm captured by the left camera, respectively; Figs. 10(b) and 10(e) represent the disparity maps of the arm obtained by the previous method and the proposed method, respectively; and Figs. 10(c) and 10(f) represent the point clouds obtained by the previous and proposed methods, respectively. In comparison with Figs. 10(b) and 10(e), the previous method had reconstruction errors at the ABB symbol on the arm and only reconstructed the bright stripes on the corrugated tube. The proposed method can obtain the correct disparity at the ABB symbol and a complete dense disparity map of the corrugated tube.

4.2.2 Comparison experiment in a defocused scene

The measurement results of the roof pantographs and wire mesh are shown in Fig. 11. Figures 11(a) and 11(d) represent the Rgb image and raw image of the pantographs and wire mesh captured by the left camera, respectively; Figs. 11(b) and 11(e) represent the disparity map

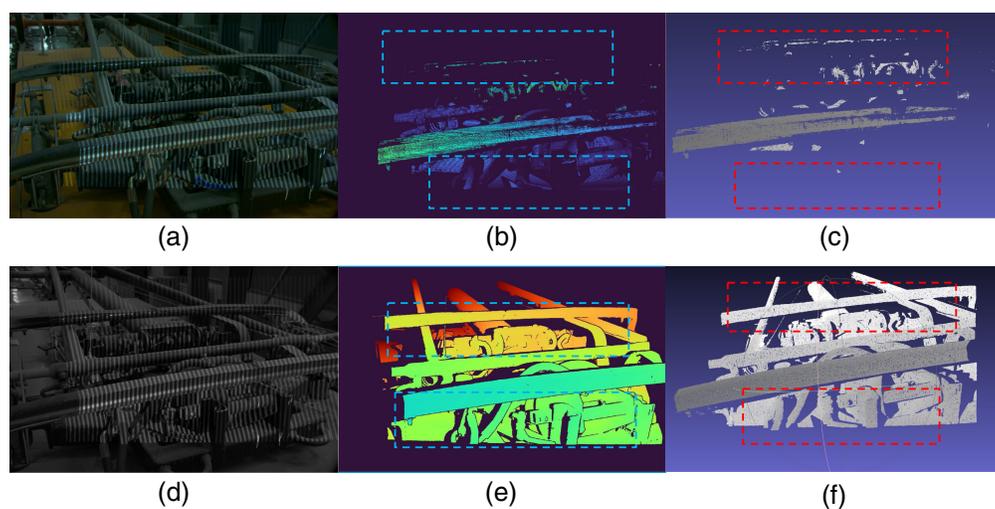


Fig. 11 Comparison experiment of train roof pantographs and wire mesh imaging: (a) Rgb image of the pantographs and wire mesh, (b) disparity map of the pantographs and wire mesh before improvement, (c) point cloud of the pantographs and wire mesh before improvement, (d) raw image of the pantographs and wire mesh, (e) disparity map of the pantographs and wire mesh after improvement, and (f) point cloud of the pantographs and wire mesh after improvement.

of the pantographs and wire mesh obtained by the previous method and the proposed method, respectively; and Figs. 11(c) and 11(f) represent the point clouds obtained by the previous and proposed methods, respectively. We keep the focal length of the projector and camera unchanged and change the distance of the scene from 0.5 to 1.5 m, causing the projector and camera to be out of focus. In comparison with Figs. 11(b) and 11(e), the traditional method can only recover partial disparity of the front pantographs and completely fail to reconstruct the bottom frame. Using the error correction strategy proposed in this paper, the complete 3D information of the entire roof pantographs and wire mesh can be reconstructed, and the impact of low reflectivity (e.g., pantograph carbon brushes or other black components) on the disparity calculation can be reduced. This experiment demonstrates the high precision and robustness of the proposed method in low reflectivity and out-of-focus scenarios.

4.2.3 Comparison experiment in a complex comprehensive scene

The measurement results of the ultrasonic probe and base are shown in Fig. 12. Figures 12(a) and 12(d) represent the Rgb image and raw image of the ultrasonic probe and base captured by the left camera, respectively; Figs. 12(b) and 12(e) represent the disparity map of the ultrasonic probe and base obtained by the previous method and the proposed method, respectively; and Figs. 12(c) and 12(f) represent the point cloud obtained by the previous and proposed methods, respectively. There is a significant difference in reflectivity across the scene. The ultrasonic probe is made of a semitransparent material. In the traditional method [Fig. 12(b)], only the base and the left part of the probe have recovered disparity, but the disparity of the component in the upper right corner cannot be calculated. In the improved method [Fig. 12(e)], a complete and dense model of the semitransparent probe can be reconstructed.

The experimental scene in Fig. 13 is even more complex, with various colored pipelines and highly reflective metal surfaces. Figures 13(a) and 13(d) represent the Rgb image and raw image of the complex pipeline scene captured by the left camera, respectively; Figs. 13(b) and 13(e) represent the disparity map of the scene obtained by the previous method and the proposed method, respectively; and Figs. 13(c) and 13(f) represent the point clouds obtained by the previous and proposed methods, respectively. The previous method [Fig. 13(b)] cannot calculate the correct disparity in overexposed areas of the pipeline surface or areas with insufficient stripe brightness. The proposed method [Fig. 13(e)] can accurately reconstruct the scene without removing ambient light interference during the day, demonstrating the strong robustness and imaging precision of the proposed method in complex industrial scenes.

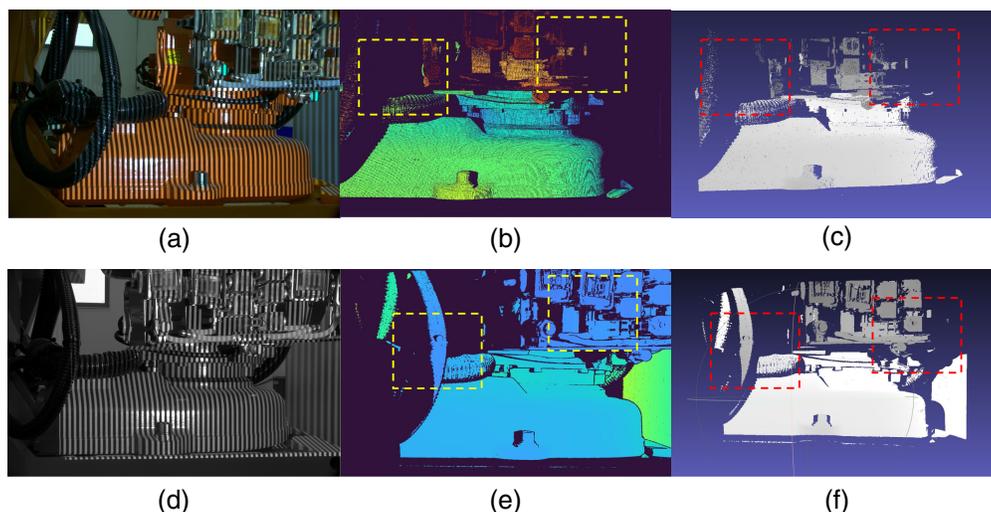


Fig. 12 Comparison experiment of ultrasound probe and base imaging: (a) Rgb image of the ultrasound probe and base, (b) disparity map before improvement, (c) point cloud before improvement, (d) raw image of the ultrasound probe and base, (e) disparity map after improvement, and (f) point cloud after improvement.

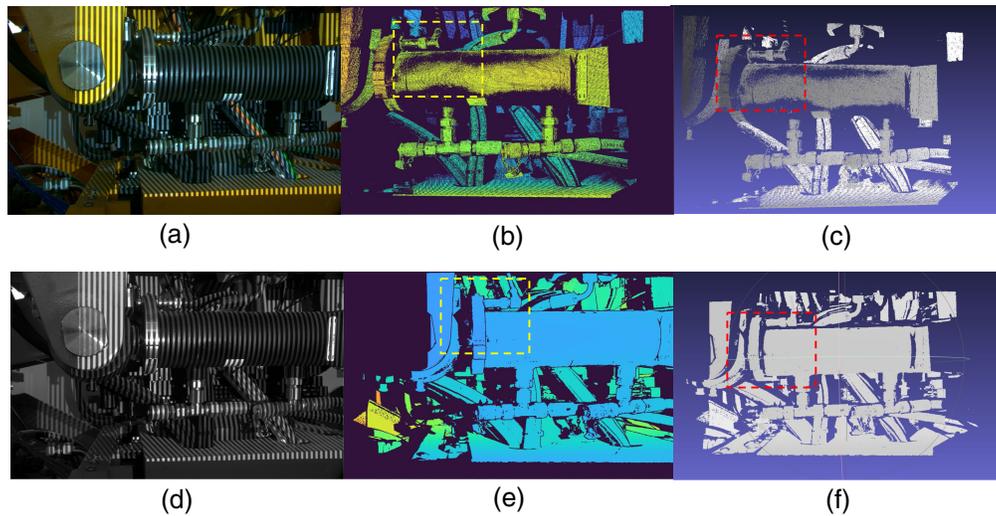


Fig. 13 Comparison experiment of complex pipeline imaging: (a) Rgb image of the pipeline, (b) disparity map of the pipeline before improvement, (c) point cloud of the pipeline before improvement, (d) raw image of the pipeline, (e) disparity map of the pipeline after improvement, and (f) point cloud of the pipeline after improvement.

4.2.4 Comparison experiment in a big scene

The measurement results of the 1.5-m long and 1-m wide train bogie are shown in Fig. 14. Figures 14(a) and 14(d) represent the Rgb and raw image captured by the left camera, respectively; Figs. 14(b) and 14(e) show the disparity map obtained using the unimproved method and the method proposed in this paper, respectively; and Figs. 14(c) and 14(f) show the point cloud generated by the unimproved and proposed methods, respectively. The robustness of our proposed method is tested through complete large-scale 3D reconstruction. Comparing Figs. 14(c) and 14(f), the unimproved method can only reconstruct a small portion near the focal plane, and due to the high level of occlusion in the images captured by the left and right camera, it can only produce a sparse and discrete point cloud. Our proposed method, on the other hand, benefits from the error point filtering strategy, which enables accurate reconstruction of the complete point cloud of the object. Additionally, the softargmin method has stronger anti-interference capabilities and better stability on noncontinuous surfaces.

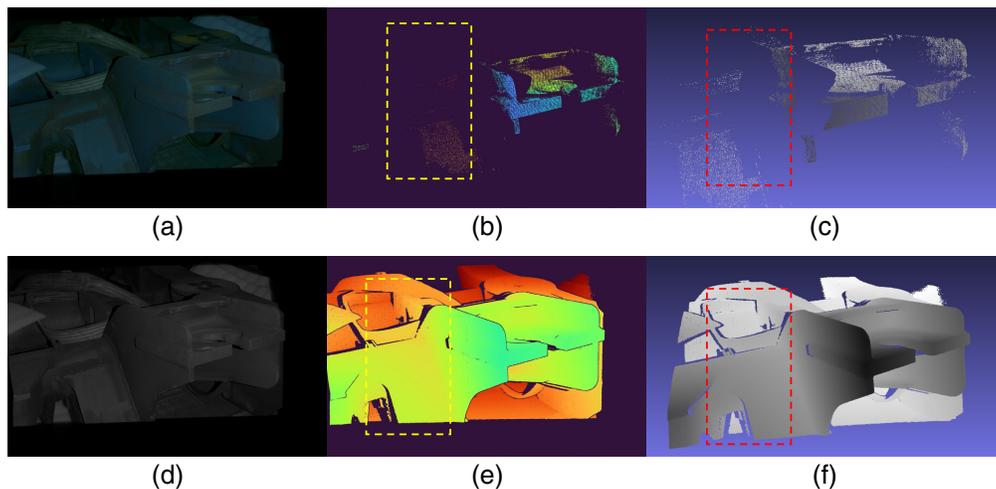


Fig. 14 Comparison experiment of complex train bogie imaging: (a) Rgb image of the train bogie, (b) disparity map of the train bogie before improvement, (c) point cloud of the train bogie before improvement, (d) raw image of the train bogie, (e) disparity map of the train bogie after improvement, and (f) point cloud of the train bogie after improvement.

5 Conclusions

The proposed method is simple and low cost, using a consumer-grade projector and color industrial cameras. Based on traditional binocular structured light, key steps were optimized for 3D imaging problems in complex industrial scenes. The use of raw images improved imaging quality, and the error point filtering strategy solved the problems of sublevel edge errors, defocusing, and reliable regions. The subpixel matching of disparity resulted in a higher precision disparity map, meeting the requirements of industrial measurement. Comparison experiments in industrial scenes showed that the proposed method has high precision and strong robustness in 3D reconstruction of complex scenes. However, the number of gray code and stripe patterns projected in this paper was relatively high, and it was difficult to quantitatively analyze the imaging accuracy in some scenarios. Future research can focus on reducing the number of projected patterns to improve the speed and efficiency of 3D measurement. Additionally, quantitative analysis can be conducted on various scenes using the experimental system that we built.

Disclosures

The authors declare that they have no competing interests.

Code and Data Availability

The data that support the findings of this study are available from the author, Xinyu Li, via email at 19935685@qq.com upon reasonable request.

Acknowledgments

This research was supported by the National Nature Science Foundation of China (Grant No. 61960206010).

References

1. Z. Song et al., "A high dynamic range structured light means for the 3D measurement of specular surface," *Opt. Lasers Eng.* **95**, 8–16 (2017).
2. J. Zhang et al., "Three-dimensional shape measurement based on speckle-embedded fringe patterns and wrapped phase-to-height lookup table," *Opt. Rev.* **28**(2), 227–238 (2021).
3. T. Shi et al., "Three-dimensional microscopic image reconstruction based on structured light illumination," *Sensors* **21**(18), 6097 (2021).
4. C. Sun and X. Zhang, "Real-time subtraction-based calibration methods for deformation measurement using structured light techniques," *Appl. Opt.* **58**(28), 7727–7732 (2019).
5. D. Jiang et al., "Gesture recognition based on binocular vision," *Cluster Comput.* **22**(S6), 13261–13271 (2019).
6. M. Bleyer, C. Rhemann, and C. Rother, "PatchMatch stereo–stereo matching with slanted support windows," in *Proc. Br. Mach. Vis. Conf.*, British Machine Vision Association, Dundee, pp. 14.1–14.11 (2011).
7. S. Van Der Jeught and J. J. J. Dirckx, "Real-time structured light profilometry: a review," *Opt. Lasers Eng.* **87**, 18–31 (2016).
8. Y. Yu et al., "Dual-projector structured light 3D shape measurement," *Appl. Opt.* **59**(4), 964–974 (2020).
9. L. Su et al., "Application of modulation measurement profilometry to objects with surface holes," *Appl. Opt.* **38**(7), 1153–1158 (1999).
10. S. Jung et al., "3D reconstruction using 3D registration-based ToF-stereo fusion," *Sensors* **22**(21), 8369 (2022).
11. Y. Ou et al., "Binocular structured light 3-D reconstruction system for low-light underwater environments: design, modeling, and laser-based calibration," *IEEE Trans. Instrum. Meas.* **72**, 1–14 (2023).
12. J. Jeong et al., "High-quality stereo depth map generation using infrared pattern projection," *ETRI J.* **35**(6), 1011–1020 (2013).
13. Q. Zhang et al., "3-D shape measurement based on complementary Gray-code light," *Opt. Lasers Eng.* **50**(4), 574–579 (2012).
14. Z. Wu et al., "High-speed three-dimensional shape measurement based on cyclic complementary Gray-code light," *Opt. Express* **27**(2), 1283–1297 (2019).

15. L. Lu et al., "High-efficiency dynamic three-dimensional shape measurement based on misaligned gray-code light," *Opt. Lasers Eng.* **150**, 106873 (2022).
16. W. Lohry, V. Chen, and S. Zhang, "Absolute three-dimensional shape measurement using coded fringe patterns without phase unwrapping or projector calibration," *Opt. Express* **22**(2), 1287–1301 (2014).
17. F. Lu, C. Wu, and J. Yang, "High-speed three-dimensional shape measurement using phase measurement profilometry without phase unwrapping," *Opt. Eng.* **57**(8), 085101 (2018).
18. S. Yu et al., "3D shape measurement based on the unequal-period combination of shifting Gray code and dual-frequency phase-shifting fringes," *Opt. Commun.* **516**, 128236 (2022).
19. J. Hu, J. Zhu, and P. Zhou, "Efficient 3D measurement of a HDR surface based on adaptive fringe projection," *Appl. Opt.* **61**(30), 9028 (2022).
20. R. Chen et al., "3D sampling moiré measurement for shape and deformation based on the binocular vision," *Opt. Laser Technol.* **167**, 109666 (2023).
21. H. Yuan et al., "An adaptive fringe projection method for 3D measurement with high-reflective surfaces," *Opt. Laser Technol.* **170**, 110062 (2024).
22. T. Engel, "3D optical measurement techniques," *Meas. Sci. Technol.* **34**(3), 032002 (2023).
23. H. Wu et al., "A novel phase-shifting profilometry to realize temporal phase unwrapping simultaneously with the least fringe patterns," *Opt. Lasers Eng.* **153**, 107004 (2022).
24. H. Li, "Application of integrated binocular stereo vision measurement and wireless sensor system in athlete displacement test," *Alexandria Eng. J.* **60**, 4325–4335 (2021).
25. J. Hyun and S. Zhang, "Phase-based stereo matching for high-accuracy three-dimensional optical sensing," in *Front. in Opt. + Laser Sci. APS/DLS, OSA Tech. Digest*, p. FW6F.2 (2019).
26. A. Kendall et al., "End-to-end learning of geometry and context for deep stereo regression," arXiv:1703.04309 (2017).
27. T. Tao et al., "Real-time 3-D shape measurement with composite phase-shifting fringes and multi-view system," *Opt. Express* **24**(18), 020253 (2016).

Xinyu Li is a third-year graduate student at the School of Physical Science and Technology of Southwest Jiaotong University. His main research interests are structured light, binocular vision, and three-dimensional reconstruction.

Kai Yang is an associate professor at Southwest Jiaotong University. He received his bachelor's, master's, and PhD degrees from Southwest Jiaotong University in 2003, 2006, and 2015, respectively. He primarily teaches simulation program design and practice. His research fields are photoelectric detection and information processing, sensor technology, MATLAB simulation, digital image, and signal processing.

Yingying Wan received her BS degree and PhD from Sichuan University, Chengdu, China, in 2012 and 2021, respectively. She was a visiting scholar at the University of Waterloo, Waterloo, Canada, from 2018 to 2020. She is currently a lecturer at Southwest Jiaotong University, Chengdu, China, and her research interests include 3D imaging and optical measurement.

Zijian Bai received his master's degree in physics (optics) from the School of Physics Science and Technology of Southwest Jiaotong University. He works at MEGVII Company, specializing in 3D imaging, deep learning, and computer vision.

YunXuan Liu received his bachelor's degree in electronic information science and technology from the School of Physics Science and Technology, Southwest Jiaotong University, Sichuan, China, in 2021. He is currently pursuing a master's degree in physics (optics) at the School of Physics Science and Technology of the Southwest Jiaotong University. His current research interests are in stereo matching, 3D imaging, deep learning, and computer vision.

Yong Wang is an associate professor of Southwest Jiaotong University and a master's supervisor. He mainly engages in the research of optical image 3D registration, measurement, machine vision, and artificial intelligence.

LiMing Xie completed his master's degree in physics from the School of Physics Science and Technology of Southwest Jiaotong University. He obtained his PhD from the same institution, specializing in 3D measurement and optical imaging, computer vision, and deep learning.