



Micrometer-level 3D measurement techniques in complex scenes based on stripe-structured light and photometric stereo

ZHAO SONG,^{1,2}  ZHAN SONG,^{1,3,*} JUAN ZHAO,^{1,2} AND FEIFEI GU^{1,2}

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, 1068 Xueyuan Avenue, Shenzhen University Town, Shenzhen, China

²University of Chinese Academy of Sciences, No. 19(A) Yuquan Road, Shijingshan District, Beijing 100049, China

³The Chinese University of Hong Kong, Shatin, NT, Hong Kong

*zhan.song@siat.ac.cn

Abstract: Compared with existing depth cameras, such as RGB-D, RealSense and Kinect, stripe-based structured light (SL) has the potential for micrometer-level 3D measurement; this can be attributed to its higher coding capacity. While surface texture, high-reflective region, and occlusion remain some of the main sources leading to degraded reconstruction quality in complex objects, methods that are only based on SL cannot completely solve such problems in complex object reconstruction. In this paper, we developed an advanced fusion strategy for the reconstruction of complex objects in micrometer-level 3D measurement. This includes solving the above-mentioned inherent problems of a stripe-based SL system with the aid of photometric stereo (PS). Firstly, to improve the robustness of decoding and eliminate the effects of noise and occlusion on stripe detection, a novel scene-adaptive decoding algorithm based on a binary tree was proposed. Further, a robust and practical calibration method for area light sources in the PS system, which utilizes the absolute depth information from SL system, was introduced. A piecewise integration algorithm, which is based on a subregion divided by Gray code, was proposed by combining the depth values from SL with the normal information from PS. Remarkably, this method eliminates the effects of surface texture and high-reflective region on the reconstruction quality and improves the resolution to camera-level resolution. In experimental parts, a regular cylinder was reconstructed to demonstrate micrometer-level measurement accuracy and resolution enhancement by the proposed method. Then, improvement of the reconstruction accuracy for objects with surface texture was validated with a regular pyramid that had textures on it and a white paper with characters printed on it. Lastly, a complex object containing multiple phenomena was reconstructed with the newly proposed method to show its effectiveness for micrometer-level 3D measurement in complex objects. Evaluation of our proposed method shows the improvement of the proposed method on the existing methods being used for micrometer-level 3D measurement in complex objects.

© 2020 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

There are many structured light (SL) based 3D acquisition techniques [1–3] applicable to various scenarios. Differences of application scenes lie mainly on scanning speed and depth resolution. For real-time, single-shot spatially multiplexed methods [4–7] which projects a single pattern with code words, is usually used for human interaction, attitude estimation with their limited resolution. For micrometer-level 3D measurement, time multiplexing methods, which encodes code words along the time axis and require the projection of several patterns, is widely applied to industrial detection without real-time demanding.

For time multiplexing methods, binary code and Gray code are usual coding strategies with advantages and drawbacks. Given that similar pattern frequencies result in similar global components affecting all pixels, an alternative method to reduce these errors is to design projection sequences with patterns having all similar spatial frequencies [8]. To maximize the minimum stripe width, MinSW8 pattern was proposed in [9].

For stripe-based coding strategy, there are two types of decoding methods, i.e. intensity-based and edge-based. Intensity-based method binarizes each pixel by taking half of the maximum and minimum values as threshold, or comparing the intensity values from normal and inverse camera images with normal and inverse patterns projected, respectively. The pixels in camera image corresponding to the same stripe in patterns are assigned with the same Gray code value and reconstruction results with pixel accuracy can be acquired [8,10]. Edge-based method [11] firstly detects the stripe edge with subpixel accuracy and divides image region into subregion. Then, for the minimum width subregion corresponding to the same Gray code value, both the type of stripe and pixel intensity within the minimum subregion are used to binarize pixels in the subregion. Subpixel accuracy can be acquired by linear interpolation with the subpixel values of stripe detection.

Apparently, the second method is more accurate and continuous compared with the first. Thus, in this paper, we select arbitrary bit Gray code and line shifting as our coding strategy, and focus on improving accuracy and robustness of the edge-based decoding method. For well-behaved scenes, micrometer-level 3D measurement can be acquired. Whereas, for complex scenes, there are still three main error sources, i.e. occlusion, surface texture and high-reflective region, which influence stripe detection and degrade reconstruction quality. For surface texture, as shown in Fig. 1, (a) binary stripe pattern is projected onto a checkerboard plane. With a close observation, it is found that the stripe width has changed due to the boundary of surface texture in the checkerboard pattern. For the areas where surface intensity suddenly changes, ridge or valley is usually produced in the reconstructed 3D model. For high-reflective region and occlusion, as shown in Fig. 2, the encoding information is missing or confused, and stripe detection fails. This leads to noise and hole in the reconstruction results.

Although stripe-based SL has great potentials for micrometer-level 3D measurement, the noise and errors caused by the above error sources cannot be eliminated completely based on SL only. This limits the use of SL in complex scenarios.

In [12], photometric stereo (PS) was proposed to acquire normal vector and albedo of the reconstructed surface. Based on intensity deviation of different images illuminated by at least three direction light sources, the normal vector and albedo of the object can be acquired. Then relative heights of the object can be calculated by Frankot-Chellappa (*FC*) algorithm [13]. In this paper, with the aid of PS [12,14], we proposed a fusion method to eliminate the effects of the above error sources on reconstruction quality, and build a micrometer-level 3D measurement system for complex scenes. In addition, with the normal information achieved from PS, we further improve system resolution from projector-level resolution to camera-level resolution for preserving more details.

Apart from stripe-based SL, phase shifting SL is another main 3D reconstruction technique in the field of time multiplexing coding strategy, which combines Gray code or multiple-frequency patterns with sinusoidal fringe patterns. For complex scenes, compared with the phase-shifting SL method, the main advantages of stripe-based SL used in this paper are as follows:

Robustness. Compared with sinusoidal fringe patterns, stripe edges, rather than raw image intensities are encoded in the illuminated patterns. As stripe edges can be generally better preserved than individual image intensity in the presence of complex reflection characteristic, binary stripe coding strategy combined with subpixel detection of stripe edge is more robust [11].

High projection speed. For binary stripe patterns, only two grey values, i.e. 0 and 255, need to be generated and projected by projector whereas for phase-shifting patterns, more grey values

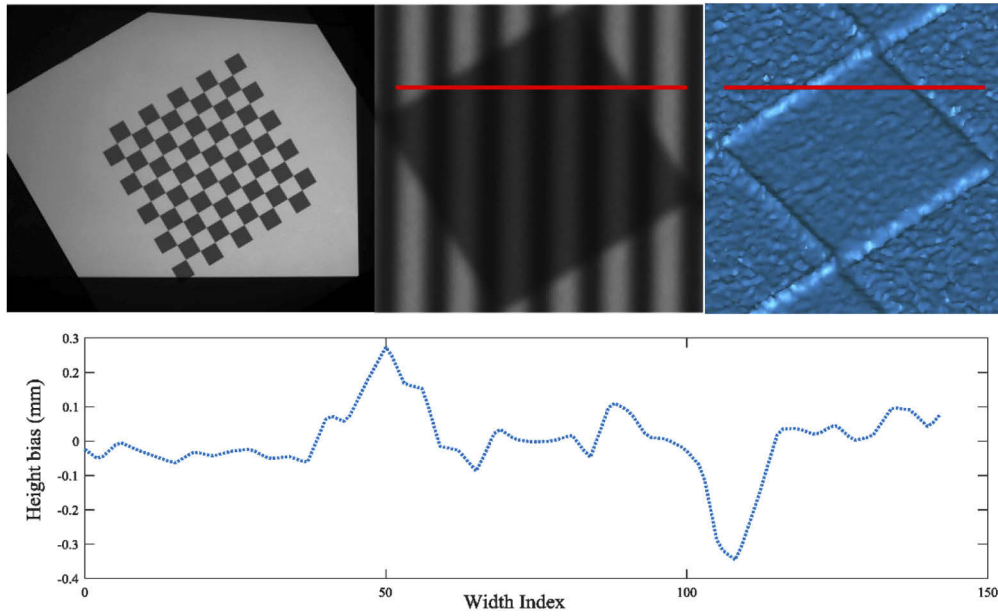


Fig. 1. Effect of texture on reconstruction results of checkerboard via Gray code and line shift pattern. The stripe edge is polluted near the boundary of surface texture. Ridge or valley exists in the reconstructed 3D model.

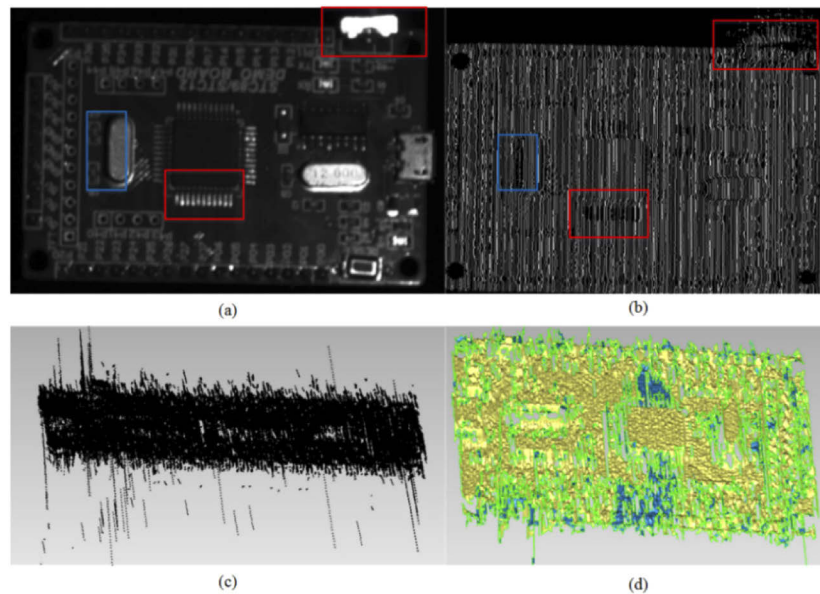


Fig. 2. Effect of texture, high-reflective region and occlusion on stripe detection. (a) A printed circuit board (PCB) contains multiple phenomena to scan. (b) The results of stripe detection. The coding information is confused or missed due to high-reflective region or occlusion. (c) Noisy point clouds due to high-reflective region and occlusion. (d) Poor reconstruction results by the edge-based decoding algorithm.

ranging from 0 to 255 need to be projected, which takes more time for scanning. Taking TI-DLP 4500 as an example, the maximum of external input pattern rate for binary pattern is 2880 Hz while only 120 Hz can be acquired for the projection of 8-bit phase-shifting patterns.

This paper is organized as follows. Section 2 gives a brief review of the previous works on stripe-based SL and the hybrid systems combining normal and depth values. A novel scene-adaptive decoding algorithm based on binary tree is introduced in Section 3. For close-range PS, a crucial and practical calibration method for area light source is proposed in Section 4. To combine depth values with normal information for complete scanning and acquire point cloud with camera-level resolution, a piecewise integration method is introduced in Section 5. Experiments on accuracy evaluation, micrometer-level measurement of complex objects, and comparisons with some commonly-used existing methods are presented in Section 6. A conclusion and possible future work are provided in Section 7.

2. Related work

2.1. Time multiplexing structured light technique

Stripe-based SL has been widely used for 3D measurement thanks to its high coding-capacity and high depth resolution. It has great potentials for micrometer-level measurement associated with the edge-based decoding method. Several methods have been proposed for general and specific scenes; however, some challenges, which limit the use of the edge-based decoding method in complex scenes, remain to be solved.

The difference of binary-stripe-based coding strategy lies in the number of patterns needed to be projected, as well as in the maximum and minimum stripe width for the same capacity. Binary code was firstly used to encode the column or row of pixel in projector image in [15]. To reduce the number of transitions and improve the robustness of decoding, Gray code was proposed in [16]. Gray code associated with line shifting [11,17,18] is another commonly used coding strategy with the advantage of minimum stripe width. Compared with 10 bit Gray code for 1024 indexes, 8 bit Gray code with 4 line-shifting patterns can acquire the same coding capacity with 4 pixels, rather than 2 pixels for 10 bit Gray code, as the minimum stripe width. To maximize the minimum stripe width, MinSW8 was proposed in [9]. For the same coding capacity, the minimum stripe width increases to 8 pixels with the maximum stripe width declining to 32 pixels. In addition, considering the different effects of global illumination on different frequency patterns, an alternative binary structured light patterns [8] by simple logical operations and tools from combinatorial mathematics was designed to improve accuracy and robustness for complex scenes. The specific data of several coding strategy are listed in Table 1.

Table 1. Comparison on coding strategy

Symbol	Number of patterns	Max. Stripe width	Min. Stripe width
Binary code	10	512	2
Gray code	10	512	2
Gray code(8) + line shifting in [11]	12	512	4
MinSW8 in [9]	10	32	8
XOR-02 in [8]	10	2	2

Intensity-based method and edge-based method [19] are the usual decoding algorithms for stripe-based SL. The former binarizes each pixel in camera image directly by taking half of the minimum and maximum values as threshold. The intensity value greater than threshold is set to 1 or 0. Depth values with pixel accuracy can be acquired. Whereas the latter divides the whole image into subregions by stripe detection firstly, and then the minimum width subregion corresponding to the same Gray code value is decoded as whole and each pixel in the same

subregion is assigned to different phase values based on stripe width and position. Thus, a subpixel accuracy can be acquired. Compared with the former, the latter can acquire more accurate and more continuous reconstruction results.

Several methods have been proposed to reduce errors for the both decoding algorithms. For intensity-based method, the intensity near the blurred stripe and system noise lead to misclassification. Some correction and fusion methods [8,20] based on the order of Gray code or Binary code were proposed to correct the decoding error. A scene-adaptive SL method was proposed in [21]. Based on a crude estimate of the scene geometry and reflectance characteristics, the local intensity ranges in the projected patterns are adapted, in order to avoid over and under exposure in the image. For the edge-based method, stripe detection is crucial. However, occlusion, surface texture and high-reflective region are three main error sources, which makes the stripe information confused, biased or even missing. In [22], taking the blurring effect of camera system into account, Gaussian model was used to present a blurred edge and the stripe detection at subpixel level can be acquired by a least squared error based solution. Jens Gühring [17] proposed the normalization method to reduce the effects of surface texture on stripe detection and for a white paper with characters printed, an average accuracy of 0.12 mm was obtained. In [11,23], an improved zero-crossing feature detector was proposed for stripe localization in high-reflective region. In addition, polygon segmentation technique [24] was used to extract and optimize light stripe centerline in line-structured laser 3D scanner.

N -step phase shifting algorithm [15] is another commonly-used technique for 3D measurements. In recent work, for textured surface, the method in [25] corrected the recovered phases by template convolution in 3×3 or 5×5 pixel windows. Apart from texture, shiny surface is another factor influencing the reconstruction accuracy as the intensity of camera image reaches the maximum intensity limitation of camera sensor for shiny surface. A high dynamic range (HDR) 3D measurement technique were proposed in [26–28]. By either changing the exposure time of camera or generating adaptive fringe patterns, multiple projection is needed to reconstruct the shiny surface. Recently, an adaptive fringe projection technique was proposed in [29]. With all-white and lower-intensity patterns projected, the adaptive sinusoidal patterns were generated based on the initial depth values and further projected to eliminate the saturation region. An improved reconstruction results with RMSE 9.23 μm can be acquired by projecting 38 patterns. In [30], the three-step phase-shifting fringe patterns with a digital speckle image for shiny surface was proposed. To avoid the camera saturation, two cameras were used to measure shiny objects from different directions. The erroneous phase obtained from saturated pixel can be corrected by that in the other camera.

2.2. Hybrid system consisting of SL and PS

It has been a common sense to improve the reconstruction accuracy and preserve the detail of the reconstructed objects by combining depth values with normal information. In [31], the corresponding literature were divided into three types of approaches, i.e. fusion approaches [32,33], subsequent approaches [34,35] and joint approaches [31,36]. The majority of previous fusion algorithms depend on a low-resolution depth camera, i.e. RGB-D, Kinect and Real Sense, etc. and improve the poor reconstruction results with normal information from PS or shape from shading. Different from the above-mentioned methods, our proposed strategy depends on stripe-based SL system to acquire initial point clouds and focus on eliminating the effects of the main error sources on the reconstruction accuracy for complex scenes. Compared with previous literature, our proposed method has a great potential to build a more accurate and effective micrometer-level 3D measurement system instead of only preserving the details of objects and improving the poor point cloud from depth cameras.

3. Scene-adaptive decoding algorithm based on a binary tree

In conventional decoding algorithms of stripe-based SL, stripe edge was generally searched individually, which leads to erroneous detection and decoding errors due to occlusion, high-reflective region and system noise. To cope with this problem, a scene-adaptive decoding method based on binary tree was developed. Taking sequence property of Gray code patterns and the pre-detection results of stripe edges into consideration, the minimum searching interval is defined and calculated firstly. Then the stripe edge to be detected corresponding to the node in binary tree is searched in the minimum searching interval and the noise and stripe edge outside the minimum searching interval will not affect the detection result of the stripe edge. By traversing the binary tree, a scene-adaptive and sequential decoding algorithm was implemented to improve the robustness of decoding and eliminate the interference of occlusion as well as noise on stripe detection.

To start with, the four key inherent attributes of Gray code are observed and analyzed as given in the following list. For convenience, two types of stripe edge in patterns are defined, i.e. the rising edge from 0 to 1 and the falling edge from 1 to 0.

- 1) With n bit Gray code, the projector region can be divided into 2^n subregions;
- 2) For each subregion with same Gray code value, the phase value corresponding to each pixel in camera image can be calculated by line interpolation;
- 3) In pattern sequences, there is a fixed position relationship between two adjacent Gray code patterns. As shown in Fig. 3, the first pattern, Pattern 1, contains a falling stripe dividing the whole region into 2 subregions, i.e. all-white and all-black subregion. In the second pattern, Pattern 2, a falling stripe divides the all-white subregion in Pattern 1 into two subregions further. Thus, the falling edge in Pattern 2 is to the left of the falling edge in Pattern 1. At the same time, the rising edge, which divides the all-black subregion in Pattern 1 into 2 subregions, is to the right of the falling edge in Pattern 1;
- 4) With normal and inverse pattern projected, a zero-crossing point L_P is detected with pixel accuracy, and the intersection l_p of the fitted lines from normal and inverse images is defined as the subpixel location of the corresponding stripe to detect. The linear least-squares problem can be solved analytically. The fitted line can be represented by the coefficients a^* and b^* as follows:

$$b^* = \frac{\sum_{i=1}^{2n+1} i \cdot I_i - (2n+1) \cdot \text{avg} \cdot \bar{I}}{\sum_{i=1}^{2n+1} i^2 - (2n+1) \cdot \text{avg}^2}, \quad (1)$$

$$a^* = \bar{I} - b^* \cdot \text{avg}, \quad (2)$$

where $\text{avg} = (\sum_{i=1}^{2n+1} i)/(2n+1)$ and $\bar{I} = (\sum_{i=1}^{2n+1} I_i)/(2n+1)$ with $\mathbf{I} = [I_1, I_2, \dots, I_{2n+1}]$ as the intensity vector centered on L_P .

With the fitted coefficients a_0^* , b_0^* from the normal image and a_1^* , b_1^* from the inverse image, (3) can be acquired:

$$a_0^* \cdot x + b_0^* = a_1^* \cdot x + b_1^*. \quad (3)$$

Thus, the subpixel value l_p of stripe edge is

$$l_p = L_P + \frac{b_1^* - b_0^*}{a_0^* - a_1^*} - (n+1). \quad (4)$$

In practice, if searching stripe edge within the whole row directly, occlusion and high-reflective region will influence stripe detection in a way that some geometric edges or edge caused by

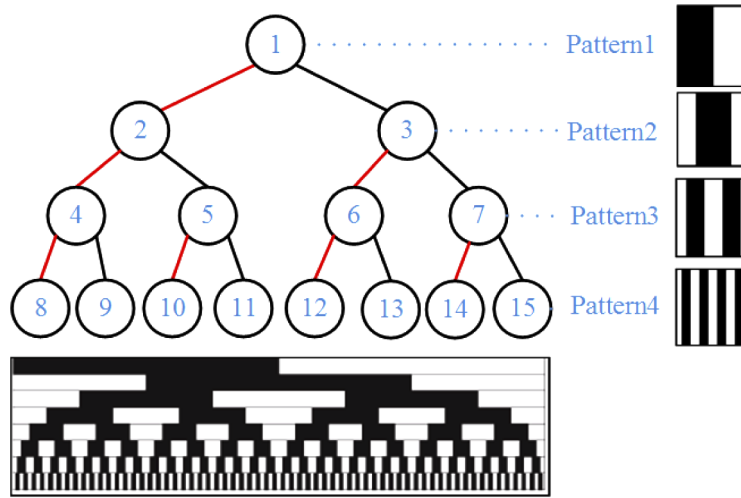


Fig. 3. Binary tree and Gray code patterns. Each layer corresponds to a Gray code pattern and each node corresponds to the stripe edge contained in pattern. A structure of node is defined to present the stripe property to detect.

occlusion could be erroneously defined as stripe edge. Based on the third attribute given above, we introduce a binary tree to characterize the structural relationships of stripe edges in patterns. As shown in Fig. 3, each layer of the binary tree corresponds to a pattern image and nodes of the binary tree in each layer represent the corresponding rising or falling edge in the corresponding pattern. All binary tree nodes are numbered in level order, from left to right. Thus for 4 Gray code patterns, a binary tree with 15 nodes on four layers can represent all stripe edges in Gray code patterns as shown in Fig. 3.

Four valuable observations are concluded as follows:

- 1) For all nodes of binary tree, even-numbered node corresponds to a falling edge, whereas the odd-numbered ones correspond to a rising edge;
- 2) In order of traversal, from left to right in each layer, the relative location relationships between nodes can characterize the location relationship between stripe edges in Gray code patterns;
- 3) For a root node, the left child of the node is numbered twice as much as the root node and the right child node is numbered twice plus one;
- 4) As shown in Fig. 3, the left child of a root node corresponds to a falling stripe marked in red whereas the right child node corresponds to a rising stripe marked in black.

Thus, a structure that represents tree node's property corresponding to stripe edge is defined as follows:

P_N : Node number in order of traversal;

P_S : Sign of the stripe existing or not;

P_t : Type of stripe, rising stripe or falling stripe;

P_{loc} : Subpixel value of stripe;

P_{Max} : The maximum value;

P_{Min} : The minimum value.

For n Gray code patterns, $P_N \in \{1, 2, \dots, 2n-1\}$. If the corresponding stripe edge is detected in a limited searching interval, P_S is set to 1, otherwise it is set to 0. When stripe edge is a rising edge, P_t is set to 1 whereas with a falling edge, P_t is 0. For the first several patterns corresponding to low-frequency Gray code patterns, P_{Max} and P_{Min} can be acquired based on the relative position of projector and camera to guarantee a limited searching interval and eliminate the effect of occlusion on decoding.

Our decoding algorithm is implemented by traversing the binary tree in level order from top to bottom and from the middle to the ends for each layer. Based on node's property defined in binary tree, a minimum searching interval is first calculated. With n bit Gray code and n camera images, we start with the first node within a fixed interval based on P_{Min} and P_{Max} . Then for subsequent nodes, taking the location information of previous nodes into consideration, a minimum searching interval is then calculated. With the minimum searching interval and the type of stripe to detect, subpixel location of the only rising or falling stripe edge can be acquired within the minimum searching interval by (1-4). This is achieved with the code snippet in.

Code Listing 1:

```

for each  $i=1 \dots n$  // all nodes
    Initializing start and end positons  $x_s = \text{img Width}$ ;  $x_e = 0$ ;
    for  $m = i \dots 1$  //each layer
        1. segment point  $P_{im}$  //find the segment point in each layer
        2.  $x_s \leftarrow \max(P_{im}, x_s)$ 
        3.  $x_e \leftarrow \min(P_{im}, x_e)$ 
    end
     $x_s \leftarrow \max(P_{Mini}, x_s)$ 
     $x_e \leftarrow \min(P_{Maxi}, x_e)$ 
    Stripe detection in  $[x_s, x_e]$ ;
end

```

Several advantages of the proposed decoding algorithm based on binary tree are concluded as follows:

- 1) Visiting nodes by level order traversal, from top to bottom and from the middle to the ends in each layer, all stripe edges in camera images corresponding to all nodes can be found.
- 2) Taking structural relationships into account, the stripe edge is searched within the minimum searching interval instead of the whole row and only the data within the interval needs to be accessed, which effectively reduces computation time.
- 3) Based on the number of nodes corresponding to the stripe edge, only a rising or falling stripe edge is detected within the interval.

Taking the node number 25 in the fifth layer of the binary tree as an example, as shown in Fig. 4(a), segment-point node corresponding to each layer is found based on the number of node, i.e. 12, 5, 2, 1. Then starting position can be calculated by getting the maximum value of all

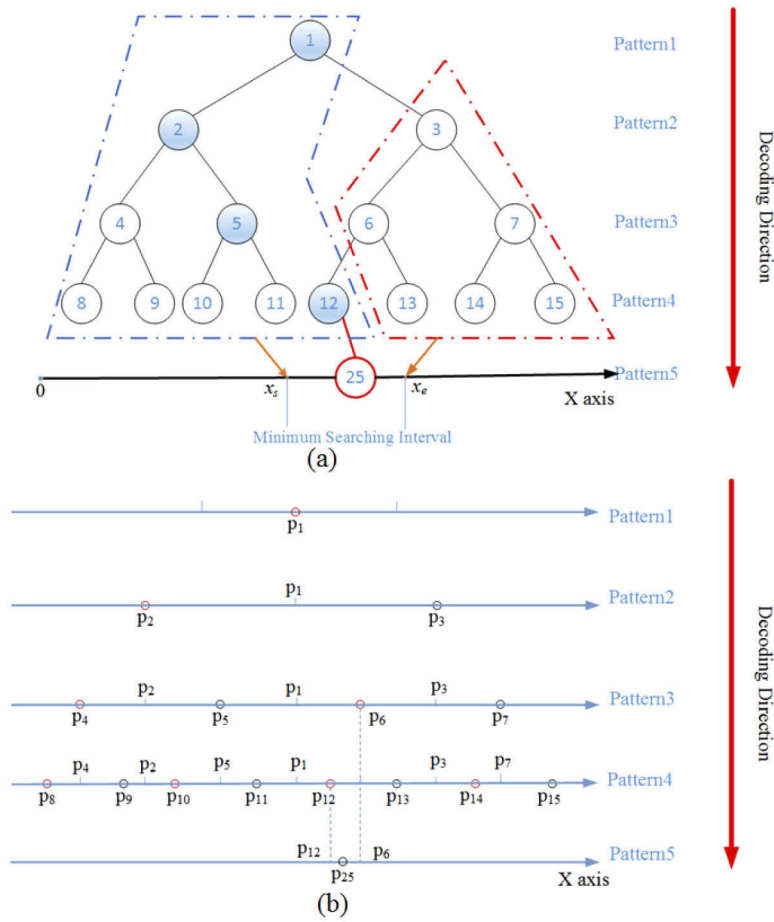


Fig. 4. (a) Illustration of the minimum searching interval based on previous location results. For node 25, segment point corresponding to each layer is found firstly, i.e. 12, 5, 2, 1. The starting position x_s is the maximum value of all nodes' positions in blue box and the ending position x_e is the minimum value of all nodes' positions in red box. The minimum searching interval for node 25 is defined as $[x_s, x_e]$, thus the noise and stripe edge outside the minimum searching interval in the same row will be removed during searching the stripe edge corresponding to node 25. (b) The detection results and minimum searching interval of stripe edges. $p_1 \sim p_{15}$ are the detection results of the stripe edges corresponding to node 1~15. The detection results of stripe edge contained in the corresponding pattern are marked below the X axis and the corresponding minimum searching interval is marked above the X axis.

nodes' positions in the blue box. Ending position can be acquired by getting the minimum value of all nodes' positions in the red box. The subpixel value of stripe edge corresponding to node 25 is acquired between the starting and ending positions instead of the whole row. Thus, the error caused by noise or occlusion can be eliminated by searching stripe edge in the minimum searching interval since the noise and occlusion outside the searching interval will not influence stripe detection. The detection results and the minimum searching interval of stripe edges are shown in Fig. 4(b). $p_1 \sim p_{15}$ are the detection results of the stripe edges corresponding to node 1~15. The detection results of stripe edge contained in the corresponding pattern are marked below the X axis and the corresponding minimum searching interval is marked above the X axis.

4. Using SL for the calibration of area light sources

For close-range PS system with area light source, the calibration of area light source is crucial to reduce deformation caused by the non-uniform illumination. We start with the observation that for a small Lambertian patch of known position relative to a rectangular illuminant, the illuminant can be replaced by an equivalent point light source at infinity [37]. Unlike the preceding calibration method which is based on several distance assumptions, our calibration method takes the absolute depth information into account.

As shown in Fig. 5, (a) Lambertian plane with several markers for coordinate transformation is placed parallel to the area light source to calibrate. Firstly, several notations are predefined for convenience. The camera coordinate system is defined as $o\text{-}xyx$ and the world coordinate system with $u\text{-}v$ plane lying in the calibration plane is defined as $m\text{-}uvh$. R and T are the rotation and transition matrix from the world coordinate system to the camera coordinate system, which can be calculated by inverse operation. The corners of area light source to calibrate are located at (u_1, v_1, D) , (u_2, v_1, D) , (u_1, v_2, D) , (u_2, v_2, D) in world coordinate system. (x_p, y_p, z_p) and $(u_p, v_p, 0)$ are coordinates of point P of the calibration plane in camera coordinate system and world coordinate system, respectively. (u_l, v_l, D) is the coordinate of surface point L in plane of area light source to calibrate. For point L, the corresponding coordinate (x_l, y_l, z_l) in camera coordinate system can be calculated as:

$$(x_l, y_l, z_l)^T = R \cdot (u_l, v_l, D)^T + T. \quad (5)$$

The intensity of surface point P receive from point L is:

$$I_{pl} = \frac{\rho \cdot ((x_l - x_p) \cdot n_x + (y_l - y_p) \cdot n_y + (z_l - z_p) \cdot n_z)}{\sqrt{((x_l - x_p)^2 + (y_l - y_p)^2 + (z_l - z_p)^2)^3}}, \quad (6)$$

where ρ is the albedo of surface point P and (n_x, n_y, n_z) is the unit normal vector of the calibration plane, which can be calculated based on the results of the reconstruction result from SL.

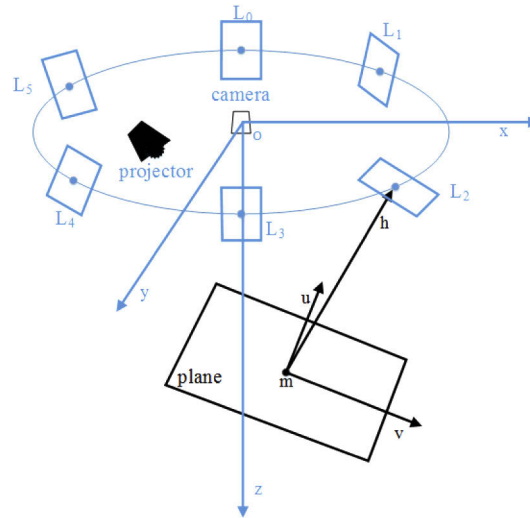


Fig. 5. Illustration of our system. The camera coordinate system $o\text{-}xyx$ and the world coordinate system $m\text{-}uvh$ with $u\text{-}v$ plane lying in the calibration plane are defined. A plane placed parallel to the corresponding area light source is used to calibrate the area light source, i.e. direction vector and illuminant intensity.

Thus, for all points lying in area light source's plane, the intensity of point P can be calculated as:

$$\hat{I}(u_1, v_1, u_2, v_2, D) = \int_{u_1}^{u_2} \int_{v_1}^{v_2} I_{pl} du dv \quad (7)$$

Since the coordinates of all surface points in calibration plane are acquired from SL and the intensity I_p of all surface points can be acquired from image, we select N surface points evenly in order to calculate parameters u_1, v_1, u_2, v_2, D in the following optimization formulation.

$$\min_{u_1, v_1, u_2, v_2, D} \left(\sum_{p=1}^N \left(\hat{I}_p(u_1, v_1, u_2, v_2, D) - I_p \right)^2 \right) \quad (8)$$

With known parameters $u_1^*, v_1^*, u_2^*, v_2^*, D^*$, the light source direction vector $\hat{l}$ and illuminant intensity E can be calculated as:

$$\hat{l} = \frac{(a_3, a_1, -a_2)}{\sqrt{a_1^2 + a_2^2 + a_3^2}} \quad (9)$$

and

$$E = \sqrt{a_1^2 + a_2^2 + a_3^2}, \quad (10)$$

where

$$a_1 = \log \left(\frac{\left(u_1 + \sqrt{D^2 + v_2^2 + u_1^2} \right) \left(u_2 + \sqrt{D^2 + v_1^2 + u_2^2} \right)}{\left(u_1 + \sqrt{D^2 + v_1^2 + u_1^2} \right) \left(u_2 + \sqrt{D^2 + v_2^2 + u_2^2} \right)} \right)$$

$$a_2 = \tan^{-1} \left(\frac{u_1 v_2}{D \sqrt{D^2 + v_2^2 + u_1^2}} - \frac{u_1 v_1}{D \sqrt{D^2 + v_1^2 + u_1^2}} \right) - \tan^{-1} \left(\frac{u_2 v_2}{D \sqrt{D^2 + v_2^2 + u_2^2}} - \frac{u_2 v_1}{D \sqrt{D^2 + v_1^2 + u_2^2}} \right)$$

$$a_3 = \log \left(\frac{\left(v_1 + \sqrt{D^2 + v_1^2 + u_2^2} \right) \left(v_2 + \sqrt{D^2 + v_2^2 + u_1^2} \right)}{\left(v_1 + \sqrt{D^2 + v_1^2 + u_1^2} \right) \left(v_2 + \sqrt{D^2 + v_2^2 + u_2^2} \right)} \right)$$

To acquire robust normal information, the reflectance model and the method to reduce the effect of non-Lambertian reflection on reconstruction in [38] was adopted associated with the proposed calibration method. Decomposing surface appearance into a diffuse component and a non-diffuse component, the method can recover complex scenes by PS such that accurate normal information can be acquired and used for normal integration.

5. Piecewise integration for resolution enhancement

For stripe-based SL, the upper limit of depth resolution is projector-level resolution. While PS can acquire normal vectors of object with camera-level resolution. Thus, to improve reconstruction resolution from projector-level to camera-level, a piecewise integration method is proposed in this section. Unlike previous fusion strategy combining normal and depth values directly in an optimization formulation [32,33], to eliminate the low-frequency deformation in PS system, not in the whole foreground region but in each subregion is normal integration implemented to enhance resolution.

After stripe detection in stripe-based SL system, the foreground region can be acquired and divided into several subregions by Gray code. As shown in Fig. 6(a), for the subregion

corresponding to the same Gary code value, the left and right position values v_l and v_r in camera image can be acquired with subpixel accuracy. X-axis coordinates of pixels within the subregion can be acquired as follows:

$$\begin{aligned} X_{c1} &= [v_l] + 1 \\ X_{cn} &= [v_r] \end{aligned}, \quad (11)$$

where $[]$ is the floor operator.

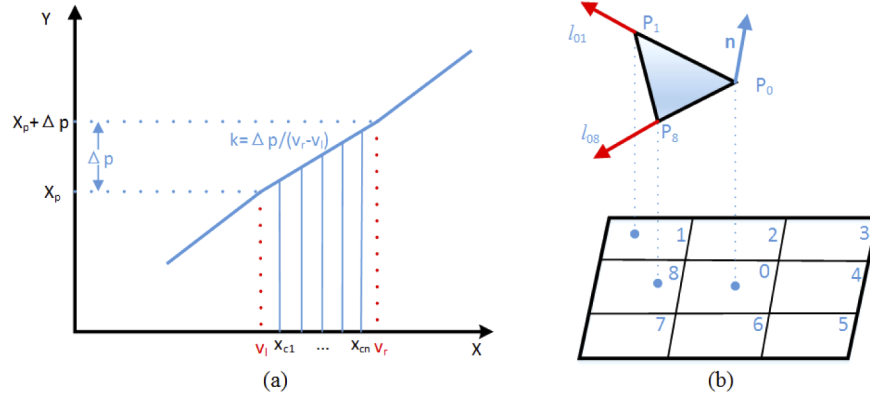


Fig. 6. (a) Linear Interpolation. The left and right positions of subregion are calculated by stripe detection. The phase values within the subregion are calculated by line interpolation. (b) An eight-neighborhood normal operator. The central points with another two pixels were used to estimate the normal vector corresponding to central points in clockwise order.

Thus, for arbitrary pixel $x_{ci} \in \{x_{c1}, \dots, x_{cn}\}$ within the subregion, the phase value x_{pij} is

$$x_{pij} = \frac{\Delta p}{(v_r - v_l)} \cdot (x_{ci} - v_l), \quad (12)$$

where Δp is the minimum stripe width in patterns.

In addition, by comparing the two masks from SLS and PS, respectively, the decoding-failed pixels are assigned to a value based on the left and right decoding values.

Thus, with the phase value x_{pii} corresponding to (x_i, y_i) in camera image, based on triangulation principle, the 3D coordinates of the point corresponding to (x_i, y_i) can be acquired. Then a normal vector operator was proposed over a 3×3 pixels window. As shown in Fig. 6(b), for the central points P_0 , another two pixels were used to estimate the normal vector of triangular patch in clock-wise order. Taking P_0 , P_8 and P_1 as an example, the normal vector $\mathbf{n}_8$ is calculated as follows:

$$\mathbf{n}_8 = \frac{\vec{l}_{08}}{|\vec{l}_{08}|} \times \frac{\vec{l}_{01}}{|\vec{l}_{01}|}. \quad (13)$$

The estimated normal vector $\tilde{\mathbf{n}}$ of the central point P_0 is

$$\tilde{\mathbf{n}} = \sum_{i=1}^8 \mathbf{n}_i. \quad (14)$$

We formulate our fusion strategy by considering the following criteria:

$$\varepsilon(x_{pij}) = \psi(x_{pij}) + \zeta(x_{pij}), \quad (15)$$

where

- $\psi(x_{pij})$ is a regularization term penalizing the normal difference between the estimated normal and the normal vector based on PS:

$$\psi(x_{pij}) = \|\tilde{\mathbf{n}}(i,j) \cdot \mathbf{n}_{ps}(i,j)\|_2^2. \quad (16)$$

- $\zeta(x_{pij})$ is a fidelity term penalizing the difference between an initial depth value and the estimated depth value:

$$\zeta(x_{pij}) = \lambda(i,j) \|\tilde{z}(i,j) - z^0(i,j)\|_2^2. \quad (17)$$

In (16), $\tilde{\mathbf{n}}(i,j)$ and $\mathbf{n}_{ps}(i,j)$ is the estimated normal by the normal vector operator based on the phase values x_{pij} from SL and the normal vector from PS, respectively. In (17), $z_0(i,j)$ is the initial depth value based on the phase value x_{pij} and $\lambda(i,j)$, $\{\lambda \in (0, 1)\}$ is the weight value which controls the respective influence of normal and absolute depth.

Thus, point clouds with camera-level resolution can be acquired by optimizing the following formulation (18) for all pixels in the same subregion Ω_i , $\Omega_i \in \Omega$.

$$\min_{z(i,j)} \int \int_{(i,j) \in \Omega_i} \|\tilde{\mathbf{n}}(i,j) \cdot \mathbf{n}_{ps}(i,j)\|_2^2 + \lambda(i,j) \|\tilde{z}(i,j) - z^0(i,j)\|_2^2 dx dy. \quad (18)$$

For specific solution of normal integration, reader is referred to the literature [33]. In fact, the proposed method was used for normal integration. To preserve details around boundary, the weight value λ is set for all pixels in foreground region and normal integration was implemented in the whole foreground region. In our proposed system, based on division by Gray code, to preserve details and eliminate low-frequency deformation from PS system, integration was implemented in the subregion with the initial depth values from structured light system. In experimental parts, we will demonstrate the advantage of the piecewise integration on detail-preserving and deformation-eliminating compared with normal integration in the whole foreground region.

6. Experiment and discussion

6.1. Hardware and calibration

This section presents reconstruction results of several objects by our proposed method and comparisons with existing methods. A regular cylinder was used firstly to show measurement accuracy and the improvement on reconstruction resolution. For surface texture, a pyramid with rectangular pattern and white paper with printed characters were projected and imaged to acquire 3D point cloud, respectively. At last, a complex object, printed circuit board (PCB) containing surface texture, high-reflective region and occlusion, was reconstructed to show the effectiveness of our fusion algorithm in general scenes.

As shown in Fig. 7(a), our hybrid system consists of a monochrome camera (Point Grey-Blackfly S, with a resolution of 2448×2048), an industrial projector (TI-DLP4500, with a resolution of 912×1140) and six area light sources (KM-FL150150). Six area light sources are placed on a circular plane centered on the camera. The camera and projector are triggered synchronously by the trigger wire. The camera and area light source are triggered by single-chip system. The working distance of the system is 35cm to 45cm and the working range of the equipment is 40cm×30cm. The system takes about 2s to complete a full scan. The PS algorithm is implemented parallel on GPU platform, which takes less than 1s to acquire the normal and albedo information. Five million points can be processed for SL system in less than 2s on a standard PC platform (Inter Xeon 3.3 GHz, with 16 GB of RAM).

The calibration method in [31–33] was used to calibrate SL system. The proposed calibration method in Section 4 was used to calibrate area light source with a calibration plane. The total time to complete reconstruction is less than 5 s with at most 5 million points acquired. All

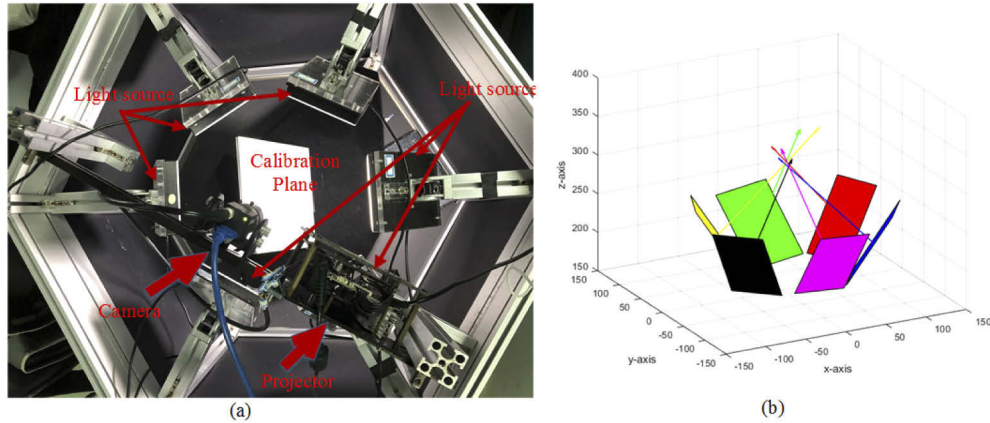


Fig. 7. (a) Our hybrid system consists of a monochrome camera, industrial projector and six area light sources (KM-FL150150). Six area light sources are placed on a circular plane centered on the camera. A calibration plane is placed parallel to the corresponding area light source to be calibrated, the location information of which can be acquired from the structured light system. (b) Illustration of calibration results via our calibration method. The area light source location can be calculated based on the estimated parameters u_1, v_1, u_2, v_2, D . The direction and length of the colored arrow represent the direction vector and intensity of the corresponding area light source, respectively.

optimization problems in this paper are solved by toolbox [29]. As shown in Fig. 7(a), the calibration plane is placed parallel to area light source to calibrate and point clouds of the calibration plane is acquired from SL system. With the area light source illuminating the plane, grey image is captured to implement the calibration of area light source. Then the relative depth of the object is acquired via *FC* algorithm. Figure 7(b) illustrates the calibration results of system in camera coordinate system. The location of area light source can be calculated based on the estimated parameters $u_1^*, v_1^*, u_2^*, v_2^*, D^*$ in (8). The direction and length of the colored arrow represent the direction vector and intensity of the corresponding area light source, respectively, with specific calibration data listed in Table 2.

Table 2. Calibration results of our PS system

Symbol	Direction vector	Intensity
L_0	-0.6235, 0.2675, 0.7346	154.5992
L_1	-0.6276, -0.1791, 0.7576	164.9824
L_2	-0.0010, -0.5169, 0.8560	183.9809
L_3	0.6111, -0.2147, 0.7619	177.0122
L_4	0.5988, 0.3937, 0.6974	144.0768
L_5	-0.0534, 0.6748, 0.7361	154.1084

Previous calibration methods were implemented and compared with ours. In [37], with the assumption that the camera plane aligns with area light source, a proper value of D is calculated by doing a search in a limited search space by using an optimization criteria based on consistency between solutions. With absolute depth values from SL system, our calibration do not need the above assumption and all location information can be acquired as shown in Fig. 7(b). In [39], initializing light source direction based on the distribution of area light source, a binary quadratic function was fitted to correct low-frequency deformation caused by non-uniform illuminant. A calibration plane was reconstructed and fitted to correct the deviation. In our PS system, a binary

quadratic function can be acquired as follows:

$$f(x, y) = 0.000066 x^2 + 0.00022 y^2 - 0.066x - 0.22 y + 48.74. \quad (19)$$

Three out of the 6 images of the plaster model with three area light sources illuminating in different direction are shown in Figs. 8(a)-(c). Though an improved result was acquired for the plane as ours, it is not feasible for free-form object due to overfitting. Compared with ours in Fig. 8(f), the result in Fig. 8(e) looks flat and is not suitable for improvement of accuracy.

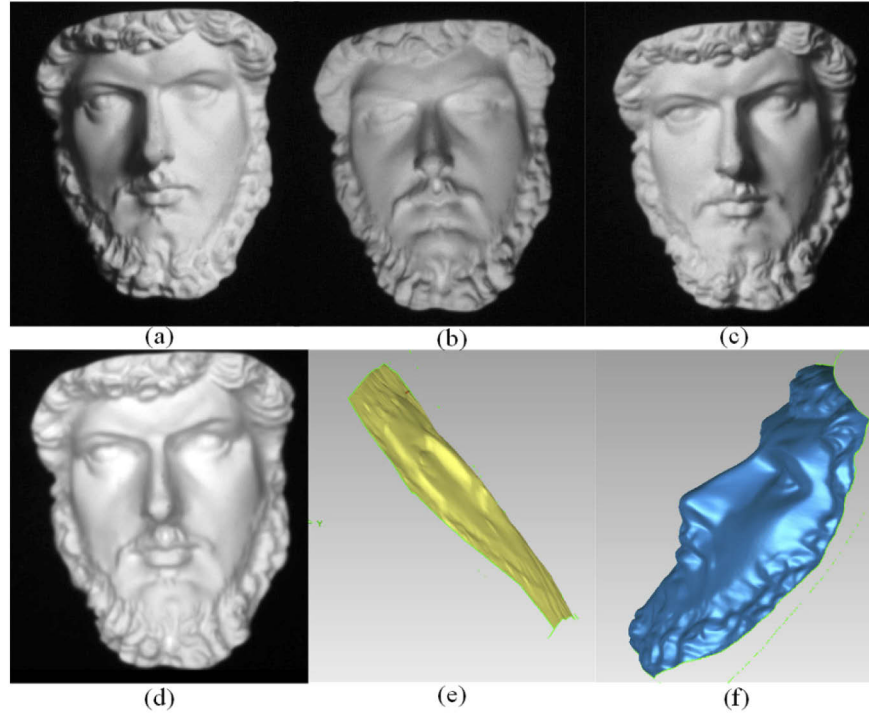


Fig. 8. Comparisons of reconstruction results in PS system. (a-b-c) Three out of the 6 images of the plaster model obtained with three area light sources illuminating in different directions. (d) The visual albedo map. (e) Results based on binary quadratic function in [39]. (f) Results based on our calibration results.

6.2. Improvement on resolution

For quantitative comparison, the point cloud was fitted into a cylinder and standard deviation, maximum and minimum error were used to estimate the performance listed in Table 3 and plotted in Fig. 9, respectively. To show the improvement of the proposed fusion method on reconstruction resolution, a regular cylinder was reconstructed firstly and the reconstruction results of 5~10 bit Gray code and the corresponding fusion results with PS were shown in Fig. 10. Parentheses contain the number of Gray code bit.

From visual perspective, compared with the results of Gray code only, the fusion method acquired smooth point cloud and improved depth resolution effectively. From statistical perspective, compared with results based on Gray code only, our proposed method acquired the minimum standard deviation, 0.0357 mm with 9 bit Gray code, which demonstrates effectiveness on resolution enhancement and micro-level measurement can be acquired.

To demonstrate the robustness of our proposed method to noise, zero mean Gaussian noise of strength σ ranging from 1% to 8% of 255 is appended to the original images acquired in PS

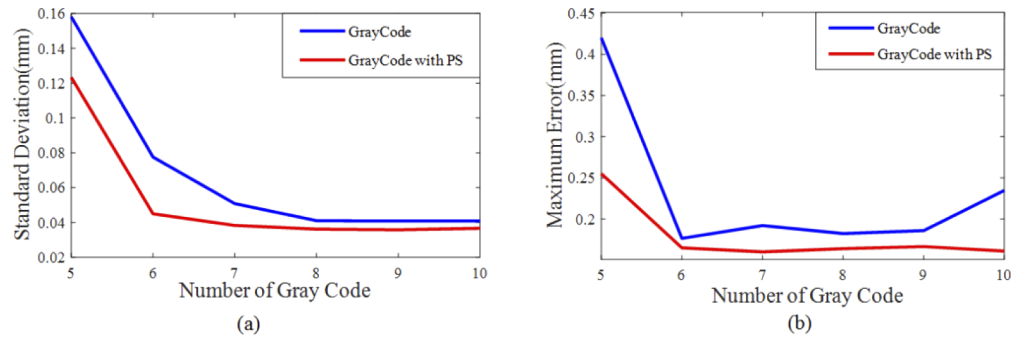


Fig. 9. Standard deviation and maximum error of point clouds corresponding to 5 ~ 10 bit Gray code with and without PS. The reconstruction accuracy and resolution was improved apparently with PS.

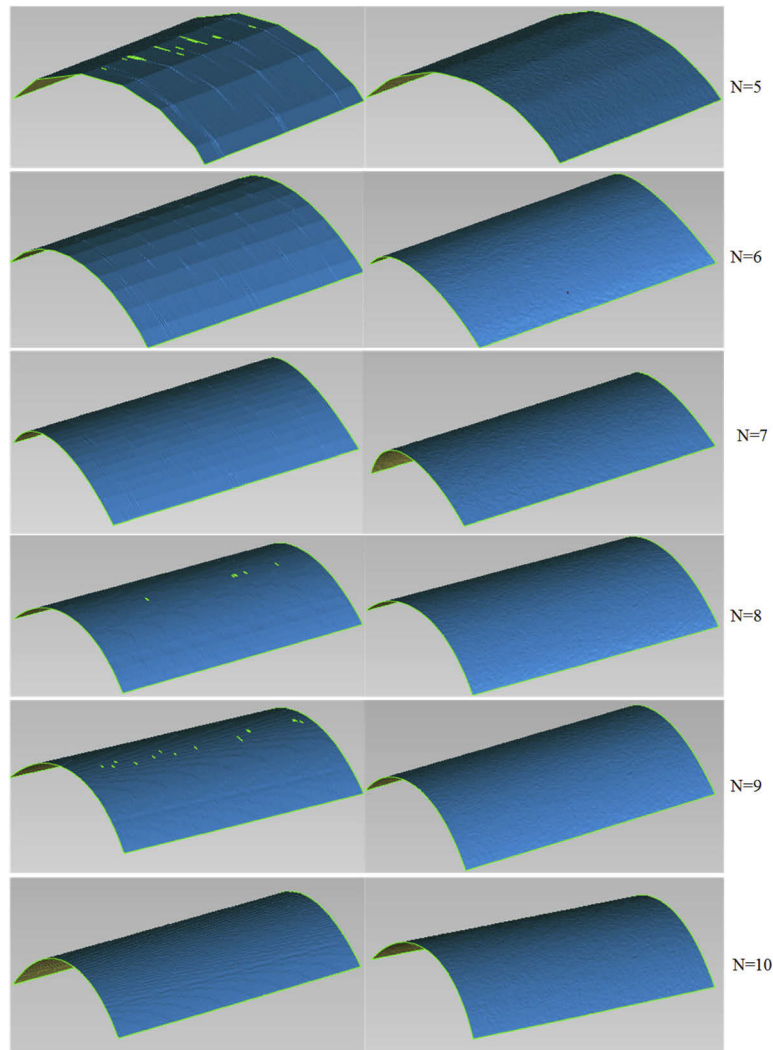


Fig. 10. The reconstruction results corresponding to 5 ~ 10 bit Gray code and the fusion results by our proposed method.

Table 3. Error comparison on cylinder (mm)

Symbol	Std	Max. residual	Min. residual
Gray code(5)	0.1581	0.4201	-0.1171
	0.1233	0.2551	-0.0968
Gray code(6)	0.0775	0.1765	-0.1352
	0.0449	0.1651	-0.1374
Gray code(7)	0.0507	0.1921	-0.1132
	0.0382	0.1601	-0.0781
Gray code(8)	0.0409	0.1823	-0.1107
	0.0361	0.1641	-0.0876
Gray code(9)	0.0407	0.1859	-0.1103
	0.0357	0.1665	-0.0909
Gray code(10)	0.0407	0.2347	-0.1118
	0.0366	0.1611	-0.0978

system, and finally the images are converted to unsigned 8 bit gray scale. Figure 11 shows the results of one of the original image with 0%, 3%, 5% and 8% additive Gaussian noise and the corresponding reconstruction results. The quantitative comparisons were listed in Table 4. We can observe that with an additive Gaussian less than 8% of 255, our proposed method can acquire stable and smooth reconstruction results, which validate the robustness of the proposed method to noise.

Table 4. Error comparison on cylinder (mm)

Symbol	Std	Max. residual	Min. residual
Gray code(9, +5cm)	0.0418	0.2052	-0.1782
Gray code(9, -5cm)	0.0406	0.2235	-0.2021
Gray code(9, $\sigma=0.01$)	0.0393	0.1986	-0.1080
Gray code(9, $\sigma=0.03$)	0.0401	0.2113	-0.1091
Gray code(9, $\sigma=0.05$)	0.0414	0.2160	-0.1220
Gray code(9, $\sigma=0.08$)	0.0464	0.2291	-0.1607
Gray code(9)	0.0392	0.2008	-0.1033

Finally, to illustrate the effectiveness of our proposed method, the cylinder was moved up and down by 5 cm. The standard deviation, maximum and minimum error of the reconstruction results were listed in Table 4 as well. It can be seen that within the focal length range, our proposed method can obtain stable reconstruction accuracy as well.

6.3. Improvement on object with texture

In this section, we focus on reduce the effect of surface texture on reconstruction results acquired by stripe-based SL method. As concluded in Section 1, surface texture changes the fine profile of stripe and a biased location leads to reconstruction error. We start with a pyramid with rectangular patterns as shown in Fig. 12(a). For quantitative comparison, reconstruction result of a textured surface of the pyramid was fitted by a plane, and the fitting residual distribution was plotted in Figs. 12(b)–12(f), respectively. The method in [11], which used 8 bit Gray code and line shifting as coding strategy was implemented to compare with ours. As shown in Fig. 12(b), the error increases apparently along the boundary of rectangular and the maximum error is 0.2344 mm. Previous fusion methods [32,33] which combines the normal and depth value in

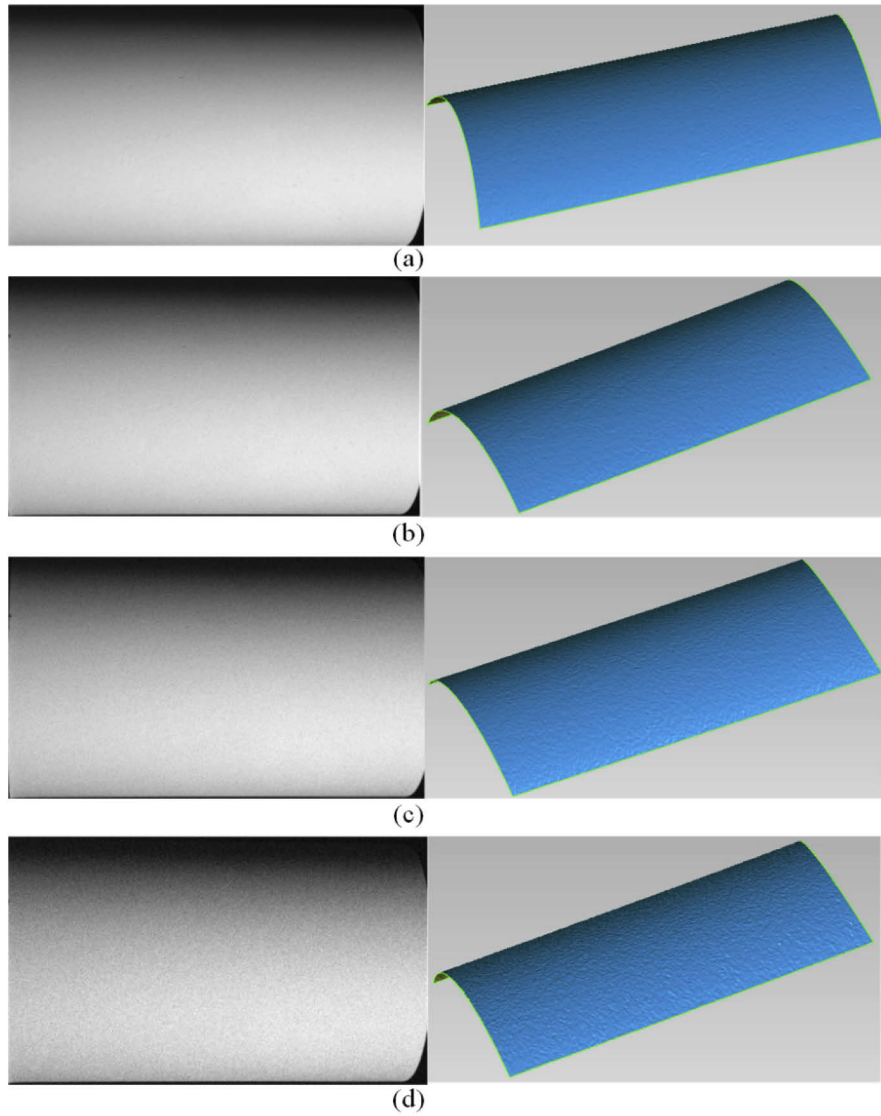


Fig. 11. The reconstruction results with image noise. (a) One of the images without noise and reconstruction result; (b) One of the images with 3% additive Gaussian noise and reconstruction result; (c) One of the images with 5% additive Gaussian noise and reconstruction result; (d) One of the images with 8% additive Gaussian noise and reconstruction result.

the whole foreground region were implemented as well. In Fig. 12(d), the maximum error declines to 0.2094 mm. Due to integration in the whole foreground region, the improvement on results is limited. Compared with reconstruction results by the methods in [32,33] and [11], the reconstructed shape by the proposed fusion method is more homogeneous as shown in Fig. 12(f). The maximum error declines to 0.0841 mm. In addition, three-step phase-shifting with adaptive albedo compensation algorithm in [25] was implemented and the reconstruction result was given in Fig. 12(e). The phase-shifting pattern is set to have a period of 32 pixels and shifted twice. For each pixel (x, y) in camera image, the phase value $\phi(x, y)$ can be calculated as:

$$\phi(x, y) = \tan^{-1} \left(\frac{\sqrt{3}(I_1 - I_2)}{2I_2 - I_1 - I_3} \right), \quad (20)$$

where I_1 , I_2 and I_3 are the grey values for pixel (x, y) in camera images with three phase-shifting patterns projected. And the ambiguity was solved by acquiring the absolute depth values with 5-bit Gray code patterns projected. The maximum error declines to 0.1253 mm and the obvious uneven along the boundary can be observed as well. At last, we try to eliminate the uneven by filtering the obtained surface from SL by Geometric Studio. The result is shown in Fig. 12(c). Although the maximum error declines to 0.0967 mm, the sharp edge of the pyramid have smoothed out as well.

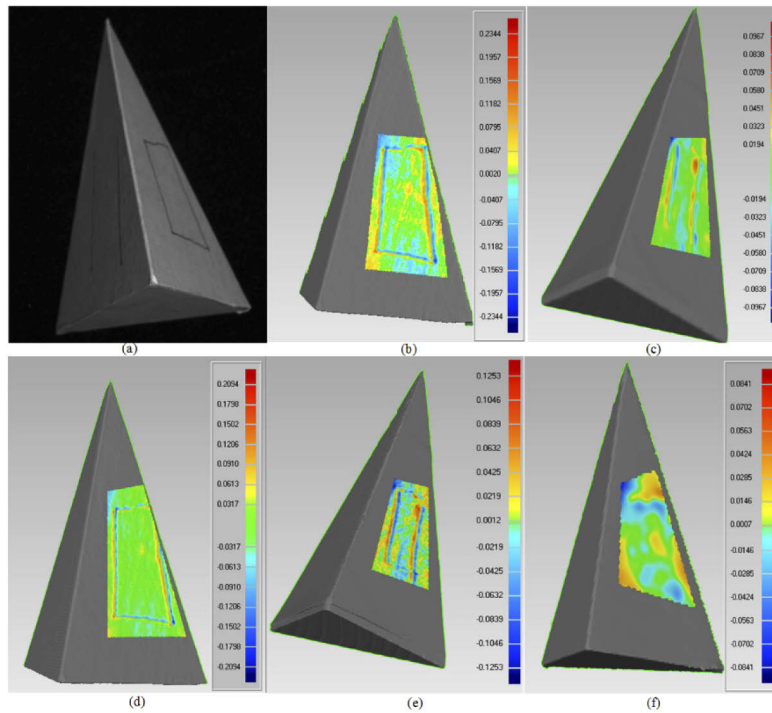


Fig. 12. Comparisons of reconstruction results of a pyramid with rectangular patterns. (a) A pyramid with rectangular patterns was used for reconstruction. (b) Planar fitting residual distribution by [11]. (c) The results by point cloud filtering algorithm. (d) Planar fitting residual distribution by [32]. (e) Planar fitting residual distribution by three-step phase-shifting pattern and adaptive albedo compensation in [25]. (f) Planar fitting residual distribution by our fusion method.

Maximum, minimum error and standard deviation of results by the above methods were listed in Table 5. By comparison, both maximal and minimum errors decline apparently by the proposed approach.

Table 5. Error comparison on pyramid with rectangular patterns (mm)

Symbol	Std	Max. residual	Min. residual
Gray code(9)	0.0402	0.3044	-0.3052
Results by [11]	0.0251	0.2344	-0.2034
Results by [32]	0.0180	0.2094	-0.1925
Results by [25]	0.0112	0.1253	-0.1035
Ours	0.0086	0.0841	-0.0720

In addition, a white paper with characters printed was reconstructed as well. Texture boundary of the printed text is not directional line segments but free curve, which is seen as high-frequency noise originating from surface texture. The target was reconstructed by the method in [11] as well. The reconstruction result based on 9 bit Gray code and the one by the proposed fusion method were shown in Fig. 13(b) and Fig. 13(d), respectively. Ridge and valley were eliminated and a smooth plane was acquired, which validates the effectiveness of the proposed fusion method against surface texture.

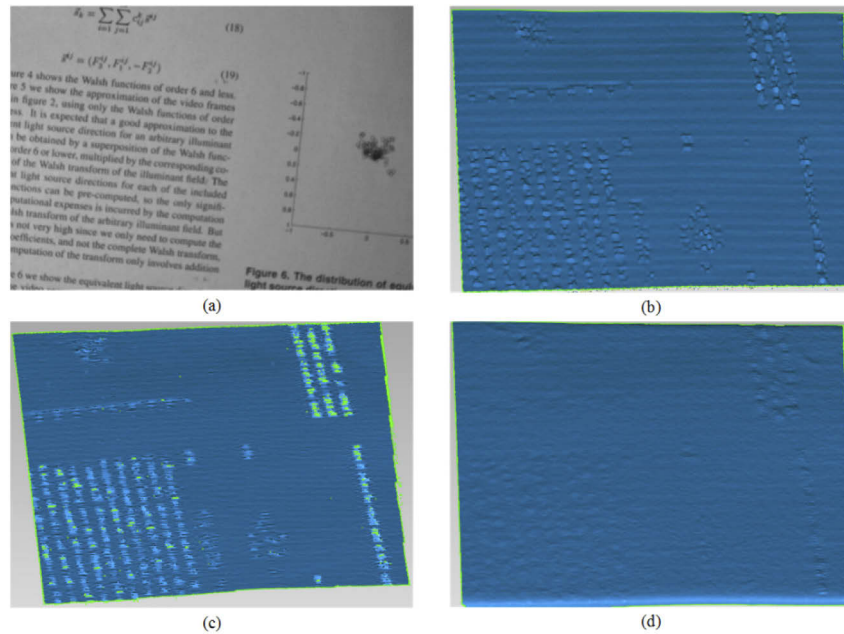


Fig. 13. Comparisons of reconstruction results. (a) A white paper with characters printed was used for reconstruction. (b) Results by Gray Code(9) only. (c) Results by Gray code(8) with line shifting in [11]. (d) Results by our fusion method, Gray code(9) with PS. Compared with results by Gray code(9) only and Gray code(8)+line shifting, the effect of surface texture on reconstruction is eliminated and a smooth surface can be acquired.

6.4. Improvement on complex scene

In this section, a porcelain bowl was reconstructed firstly as shown in Fig. 14(a). Due to the enamel of surface, distinct concave-convex reconstruction surface can be seen by stripe-based

SL in Fig. 14(c). The reconstruction results by PS and *FC* algorithm was shown in Fig. 14(b) without absolute depth. In addition, the reconstruction result by the three-step phase-shifting algorithm was given in Fig. 14(f) as well. Due to the non-diffuse property, apparent wavy uneven can be observed. We eliminated the uneven by filtering the reconstructed surface from SL based on Geometric Studio. As shown in Fig. 14(e), although the uneven can be eliminated by filtering, sharp edges of the object have smoothed out as well [highlighted in red in Fig. 14(e)]. Thus, point cloud filtering is an effective method for planar object, whereas for free-form object with sharp edges, it will smooth out sharp edge. A smooth results with sharp edges by our proposed fusion method was shown in Fig. 14(d), which demonstrated the effectiveness of our fusion method on non-diffuse surface.

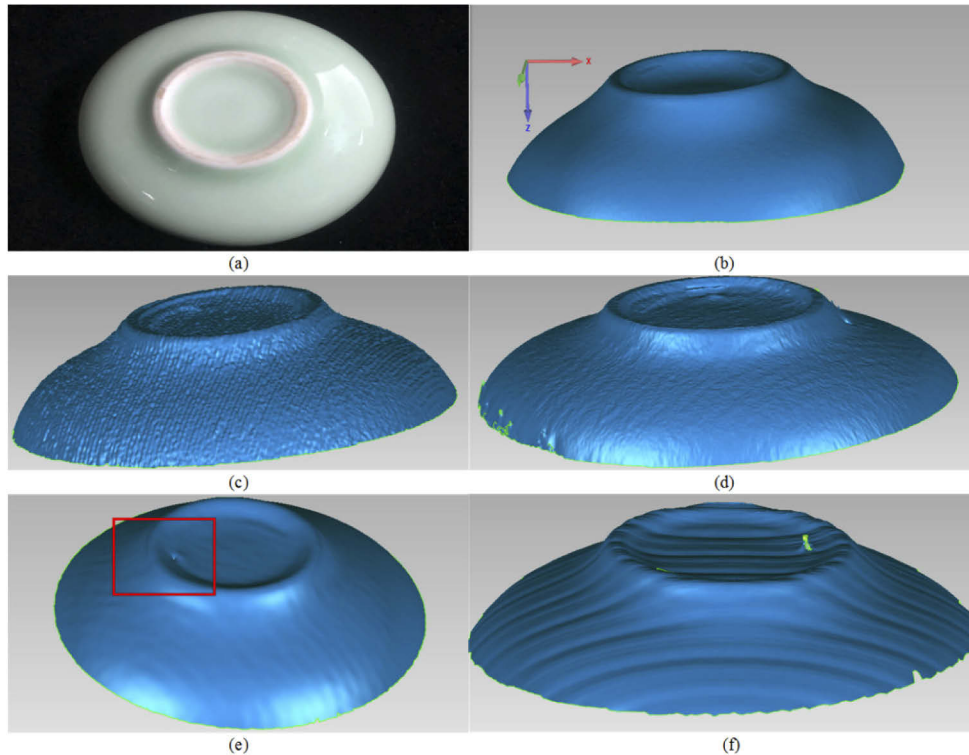


Fig. 14. Results and comparisons of a porcelain bowl. (a) A porcelain bowl was used for reconstruction. (b) The results by PS and *FC* algorithm. Results without absolute depth were acquired. (c) The results by [11]. (d) The results by our method. (e) The results by filtering algorithm. The sharp edges have been smoothed out (highlighted in red). (f) The results by three-step phase-shifting and adaptive albedo compensation algorithm in [25].

Finally, we scan complex object containing occlusion, texture and high-reflective region and demonstrate the improvement of our fusion method on the reconstruction of complex object. The 3D measurement of printed circuit board (PCB) is a challenging and ongoing problem. Two out of the 6 images of PCB illuminated from different area light source in PS system were given in Figs. 15(a)–15(b) and one of the 20 images of PCB with Gray code pattern projected in stripe-based SL is shown in Fig. 15(c). Due to high-reflective, texture and occlusion, developer was used firstly, which is time-consuming with a few details missing due to the thickness of developer. By our proposed fusion method, 9 bit Gray code with PS, more details can be acquired directly without developer as shown in Fig. 15(g). By our proposed decoding algorithm based on binary tree, noise of the point cloud reduced apparently as shown in Fig. 15(d) compared

with searching the stripe edge within the whole row in Fig. 2(c). Taking the previous location results and the type of stripe edges into consideration, only a rising or falling edge was detected in the minimum searching interval, which reduces the effect of geometric edge, edge caused by occlusion and high-reflective region on stripe detection effectively. The stripe-edge-based method for shiny surface in [11] was implemented for comparison. Compared with the reconstruction results by [11] as shown in Fig. 15(e), our proposed fusion method acquired smooth and complete reconstruction results without apparent hole. Combining normal information with depth values by piecewise integration effectively recovered the hole and preserved more details with camera-level resolution. The reconstruction result of PCB with developer by stripe-based SL was shown in Fig. 15(k). Due to the thickness of developer, some surface details near pin area was missed. In addition, the method in [8] was used for comparison and the result was shown in Fig. 15(f). The method binarizes each pixel directly by comparing the intensity values from normal and inverse

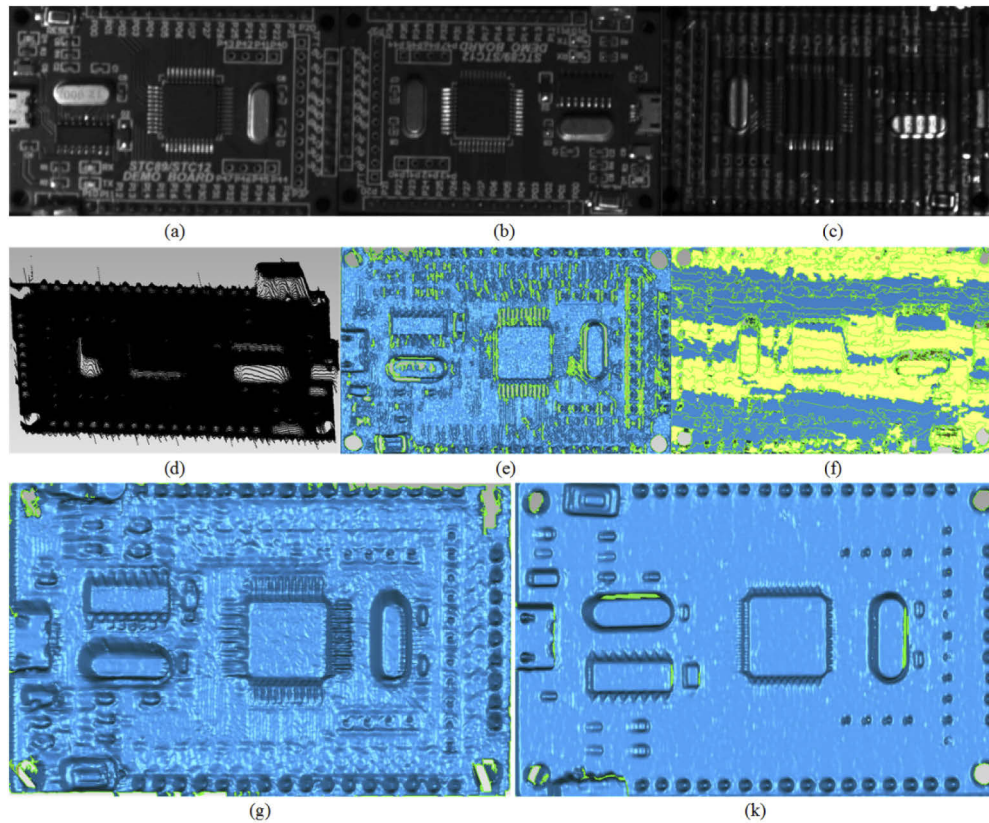


Fig. 15. Results and comparisons of PCB. (a-b) Two out of the 6 images of PCB illuminated from different area light source in PS system. (c) One of the images of PCB with Gray code pattern projected in stripe-based SL system. (d) The point cloud by the decoding algorithm based on binary tree. By searching the only rising or falling stripe edge within minimum searching interval, noise reduce apparently compared with the results in Fig. 2(c) in the whole row. (e) The results by [11]. Apparent hole exists due to high-reflective region and occlusion. (f) The results by [8]. Results with pixel accuracy were acquired. (g) The results by our method without developer. Complete scanning results were acquired with details preserved. (k) Results with developer. Several details were missed due to the thickness of developer.

camera images. In our system, as the resolution of camera is bigger than that of projector and the minimum stripe width in patterns corresponds to more than one pixel in camera image, results with only pixel accuracy can be acquired, which is not smooth and continuous.

7. Conclusion

Compared with existing depth camera, i.e. RGB-D, RealSense and Kinect, stripe based SL has great potentials for micrometer-level 3D measurement. However, some error sources, i.e. surface texture, high-reflective region, geometric edges as well as occlusion, limit its use in complex scenes. Firstly, to improve the robustness of stripe-based SL, a scene-adaptive decoding method based on binary tree was proposed. Based on the hybrid system consisting of stripe-based SL and PS, a piecewise integration was proposed and validated to enhance the reconstruction resolution from projector-level to camera-level. A regular cylinder was used firstly for experiment. The standard deviation validates the effectiveness of piecewise integration on resolution enhancement. The reconstruction results with standard deviation less than 0.035 mm can be obtained. In addition, the results of pyramid with texture and white paper with characters printed shows the improvement on reconstruction results of objects with surface texture. For complex object, a PCB containing high-reflective region, surface textures, geometric edge and occlusion, our proposed fusion method, 9 bit Gray code with PS, can acquire complete scanning results without hole compared with existing edge-based structured light decoding algorithm. Intensity-based decoding algorithm was compared with ours as well. The results illustrate the improvement of our proposed algorithm on complex scene. In addition, a crucial and practical calibration method for area light source in photometric stereo system was proposed as well. Unlike previous calibration method based on several distance assumptions, with the depth information from structured light system, an accurate calibration results can be acquired. In our system, six light sources are used for PS. In the future, for improving the robustness of PS in complex scenes further, more light sources can be used. In addition, since our proposed decoding algorithm based on binary tree is implemented for each row in camera image, it has a great potential to reduce computing time by parallel computation.

Funding

National Key Research and Development Program of China (2017YFB1103602); Science and Technology Planning Project of Guangdong Province, China (2019B010149002); Natural Science Foundation of Guangdong Province (2020A1515010486); Natural Science Foundation of Shenzhen (JCYJ20190806171403585).

Disclosures

The authors declare no conflicts of interest.

References

1. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognit.* **43**(8), 2666–2680 (2010).
2. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognit.* **37**(4), 827–849 (2004).
3. T. Bakirman, M. U. Gumusay, H. C. Reis, M. O. Selbesoglu, S. Yosmaoglu, M. C. Yaras, D. Z. Seker, and B. Bayram, "Comparison of low cost 3D structured light scanners for face modeling," *Appl. Opt.* **56**(4), 985–992 (2017).
4. S. Van der Jeught and J. J. Dirckx, "Real-time structured light profilometry: a review," *Opt. Lasers Eng.* **87**, 18–31 (2016).
5. Z. Song and R. Chung, "Determining both surface position and orientation in structured-light-based sensing," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1770–1780 (2010).
6. X. Huang, J. Bai, K. Wang, Q. Liu, Y. Luo, K. Yang, and X. Zhang, "Target enhanced 3D reconstruction based on polarization-coded structured light," *Opt. Express* **25**(2), 1173–1184 (2017).

7. C. Guan, L. Hassebrook, and D. Lau, "Composite structured light pattern for three-dimensional video," *Opt. Express* **11**(5), 406–417 (2003).
8. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus," *Int. J. Comput. Vis.* **102**(1-3), 33–55 (2013).
9. D. Kim, M. Ryu, and S. Lee, "Antipodal gray codes for structured light," in *IEEE International Conference on Robotics and Automation*, (IEEE, 2008), 3016–3021.
10. D. Moreno, F. Calakli, and G. Taubin, "Unsynchronized structured light," *ACM Trans. Graph.* **34**(6), 1–11 (2015).
11. Z. Song, R. Chung, and X.-T. Zhang, "An accurate and robust strip-edge-based structured light means for shiny surface micrometre measurement in 3-D," *IEEE Trans. Ind. Electron.* **60**(3), 1023–1032 (2013).
12. R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.* **19**(1), 191–139 (1980).
13. R. T. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Trans. Pattern Anal. Machine Intell.* **10**(4), 439–451 (1988).
14. Z. Song, Y. Nie, and Z. Song, "Photometric stereo with quasi-point light source," *Opt. Lasers Eng.* **111**, 172–182 (2018).
15. J. L. Posdamer and M. Altschuler, "Surface measurement by space-encoded projected beam systems," *Comput. Graphics Image Process.* **18**(1), 1–17 (1982).
16. S. Inokuchi, "Range imaging system for 3-D object recognition," *ICPR*, 1984, 806–808 (1984).
17. J. Gühring, "Dense 3D surface acquisition by structured light using off-the-shelf components," in *Videometrics and Optical Methods for 3D Shape Measurement*, (International Society for Optics and Photonics, 2000), 220–231.
18. Y. Ye, H. Chang, Z. Song, and J. Zhao, "Accurate infrared structured light sensing system for dynamic 3D acquisition," *Appl. Opt.* **59**(17), E80–E88 (2020).
19. M. Trobina, "Error model of a coded-light range sensor," Technical report (1995).
20. X. Chen and Y.-H. Yang, "Scene adaptive structured light using error detection and correction," *Pattern Recognit.* **48**(1), 220–230 (2015).
21. T. P. Koninckx, P. Peers, P. Dutré, and L. Van Gool, "Scene-adapted structured light," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, (IEEE, 2005), 611–618.
22. Y. Jian, G. Fu, and U. P. Poudel, "High-accuracy edge detection with Blurred Edge Model," *Image Vision Comput.* **23**(5), 453–467 (2005).
23. Z. Song, H. Jiang, H. Lin, and S. Tang, "A high dynamic range structured light means for the 3D measurement of specular surface," *Opt. Lasers Eng.* **95**, 8–16 (2017).
24. T. Qingguo, Z. Xiangyu, M. Qian, and G. Baozhen, "Utilizing polygon segmentation technique to extract and optimize light stripe centerline in line-structured laser 3D scanner," *Pattern Recognit.* **55**, 100–113 (2016).
25. M. Pistellato, L. Cosmo, F. Bergamasco, A. Gasparetto, and A. Albarelli, "Adaptive Albedo Compensation for Accurate Phase-Shift Coding," in *24th International Conference on Pattern Recognition (ICPR)*, (IEEE, 2018), 2450–2455.
26. H. Zhao, X. Liang, X. Diao, and H. Jiang, "Rapid in-situ 3D measurement of shiny object based on fast and high dynamic range digital fringe projector," *Opt. Lasers Eng.* **54**, 170–174 (2014).
27. H. Jiang, H. Zhao, and X. Li, "High dynamic range fringe acquisition: a novel 3-D scanning technique for high-reflective surfaces," *Opt. Lasers Eng.* **50**(10), 1484–1493 (2012).
28. Y. Liu, Y. Fu, X. Cai, K. Zhong, and B. Guan, "A novel high dynamic range 3D measurement method based on adaptive fringe projection technique," *Opt. Lasers Eng.* **128**, 106004 (2020).
29. H. Lin, J. Gao, Q. Mei, Y. He, J. Liu, and X. Wang, "Adaptive digital fringe projection technique for high dynamic range three-dimensional shape measurement," *Opt. Express* **24**(7), 7703–7718 (2016).
30. S. Feng, Q. Chen, C. Zuo, and A. Asundi, "Fast three-dimensional measurements for dynamic scenes with shiny surfaces," *Opt. Commun.* **382**, 18–27 (2017).
31. 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.* **126**(12), 1342–1366 (2018).
32. D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Efficiently combining positions and normals for precise 3D geometry," *ACM Trans. Graph.* **24**(3), 536–543 (2005).
33. Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Variational methods for normal integration," *J. Math. Imaging Vis.* **60**(4), 609–632 (2018).
34. M. Haque, A. Chatterjee, and V. Madhav Govindu, "High quality photometric reconstruction using a depth camera," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2014), 2275–2282.
35. A. Chatterjee and V. Madhav Govindu, "Photometric refinement of depth maps for multi-albedo objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2015), 933–941.
36. E. Bylow, R. Maier, F. Kahl, and C. Olsson, "Combining depth fusion and photometric stereo for fine-detailed 3d models," in *Scandinavian Conference on Image Analysis*, (Springer, 2019), 261–274.
37. J. J. Clark, "Photometric stereo with nearby planar distributed illuminants," in *The 3rd Canadian Conference on Computer and Robot Vision*, (IEEE, 2006), 16–16.
38. S. Ikehata, D. Wipf, Y. Matsushita, and K. Aizawa, "Photometric stereo using sparse Bayesian regression for general diffuse surfaces," *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(9), 1816–1831 (2014).
39. F. Hao, Q. Lin, W. Nan, J. Dong, and Y. Hui, "Deviation correction method for close-range photometric stereo with nonuniform illumination," *Opt. Eng.* **56**(10), 103102 (2017).