Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Three-dimensional surface reconstruction via a robust binary shape-coded structured light method

Suming Tang
Xu Zhang
Zhan Song
Hualie Jiang
Lei Nie

SPIE.

Suming Tang, Xu Zhang, Zhan Song, Hualie Jiang, Lei Nie, "Three-dimensional surface reconstruction via a robust binary shape-coded structured light method," *Opt. Eng.* **56**(1), 014102 (2017), doi: 10.1117/1.OE.56.1.014102.

Three-dimensional surface reconstruction via a robust binary shape-coded structured light method

Suming Tang,[a,b] Xu Zhang,[b] Zhan Song,[a,c,*] Hualie Jiang,[a] and Lei Nie[a]
[a]Chinese Academy of Sciences, Shenzhen Institutes of Advanced Technology, Guangdong Provincial Key Laboratory of Robotics and Intelligent System, No. 1068, Xuyuan Road, Shenzhen 518055, China
[b]Shanghai University, School of Mechatronic Engineering and Automation, Department of Mechanical Engineering and Automation, No. 149, Yanchang Road, Shanghai 200072, China
[c]The Chinese University of Hong Kong, Department of Mechanical and Automation Engineering, Shatin, New Territories, Hong Kong 999077, China

Abstract. A binary shape-coded structured light method for single-shot three-dimensional reconstruction is presented. The projected structured pattern is composed with eight geometrical shapes with a coding window size of 2×2. The pattern element is designed as rhombic with embedded geometrical shapes. The pattern feature point is defined as the intersection of two adjacent rhombic shapes, and a multitemplate-based feature detector is presented for its robust detection and precise localization. Based on the extracted grid-points, a topological structure is constructed to separate the pattern elements from the obtained image. In the decoding stage, a training dataset is first established from training samples that are collected from a variety of target surfaces. Then, the deep neural network technique is applied for the classification of pattern elements. Finally, an error correction algorithm is introduced based on the epipolar and neighboring constraints to refine the decoding results. The experimental results show that the proposed method not only owns high measurement precision but also has strong robustness to surface color and texture. [DOI: 10.1117/1.OE.56.1.014102]

Keywords: structured light; binary geometrical coding; robust decoding; three-dimensional reconstruction.

Paper 161503 received Sep. 26, 2016; accepted for publication Dec. 16, 2016; published online Jan. 6, 2017.

1 Introduction

Three-dimensional (3-D) object reconstruction is becoming an increasingly important research topic in computer vision domains and demanded by more and more real applications. Structured light-based 3-D sensing technology is considered one of the most reliable means for surface shape reconstruction.[1,2] The underlying principle of the structured light method is to project single or multiple patterns on the target surface, and the projected patterns can be used to establish the correspondences between the camera and projector. With the system calibration parameters, 3-D reconstruction can be realized via triangulation principle.[3]

Time and spatial multiplexing techniques are the two major codification strategies for existing structured light methods.[4] Temporal-based coding methods are based on the codeword created by a sequential projection of patterns onto the object surface, so the codeword associated to a position in the image is not completely formed until all patterns have been projected. Such methods can usually provide a 3-D point-cloud with high accuracy and density with a sacrifice of scanning efficiency. In comparison, spatially encoded structured light means only demand a single projection and image shot and thus are more suitable for dynamic 3-D reconstruction applications. For spatial structured light methods, the codeword of a specific position can be determined by its neighboring pattern elements, and a De Bruijn sequence,[5] pseudorandom array, or M-array[6] is usually used to construct the projected pattern. There have been a lot of studies contributed to the spatial structured light pattern codification strategies. The proposed pattern images can be classified into two types: color pattern and binary geometrical pattern. The primitive in color pattern can be coded by color multislits,[7–9] color stripes,[10–12] color grids,[13] color spots,[14,15] color diamonds,[16,17] or color squares.[18] For the binary geometrical patterns, the primitive can be represented by different geometrical shapes[19–24] or hybrid coding.[25] Compared with color coding methods, shape coding methods are more robust because they are less sensitive to surface color. In the spatial structured light patterns, a small coding window is usually expected to relieve the difficulties in the decoding procedure. However, a small coding window often causes a greater number of colors or geometrical shapes in the pattern with a given coding volume. For the color coding methods, the usage of more colors makes the shape reconstruction more sensitive to surface color or textures. In contrast, the shape coding methods usually adopt binary shapes and thus are more robust to surface color. However, the projected binary shapes are usually distorted and blended with surface textures and that brings huge difficulty for the pattern decoding algorithms.

In this paper, a robust binary shape-coded structured light method is investigated. Based on the coding scheme of pseudorandom array, eight geometrical shapes are designed to generate a binary structured light pattern with the coding window size of only 2×2. The use of binary pattern feature makes it robust to surface color, and the small coding window size makes it robust to surface discontinuities. To extract the feature points, a multitemplate-based feature detector is presented. In the decoding stage, a training dataset is first constructed by collecting a lot of pattern elements with

*Address all correspondence to: Zhan Song, E-mail: zhan.song@siat.ac.cn

various blurring and distortion. Then, a deep neural network is trained for the pattern decoding purpose. Finally, the epipolar constraint and unique window constraint are applied to refine the primary decoding results.

The rest of this paper is organized as follows. Related works are briefly reviewed in Sec. 2. In Sec. 3, the pattern design scheme is presented. The proposed feature point detection algorithm is introduced in Sec. 4. Section 5 shows how the proposed pattern can be decoded and how the decoding results are optimized. The experimental results are given and discussed in Sec. 6. Conclusions are offered in Sec. 7.

2 Related Works

Image color cues are usually used for most spatial structured light methods. Fechteler and Eisert[7] chose seven colors to generate a multislit pattern based on the De Bruijn sequence. There was a constraint that two consecutive stripes had to differ in at least two color channels. The centers of the stripes were defined as the feature points, which can provide subpixel accuracy for 3-D reconstruction. Zhang et al.[11,12] used six colors to construct a pseudorandom pattern with 128 stripes, and the window size was 1×3. Each two adjacent color stripes also conformed to the condition of being different in at least one color channel. The edge between two adjacent stripes was defined as the pattern feature point. Salvi et al.[13] introduced a color grid pattern. The pattern was composed of the projection of a grid made by color slits in such a way that each slit with its two neighbors appeared only once in the pattern. Morano et al.[14] used perfect submap to generate a color spot pattern; the centroids of the circular elements were determined as the feature points, but no quantitative experimental results were provided. Adan et al.[15] presented a color spot pattern with seven colors for 3-D tracking of dynamic targets. The proposed pattern was generated by inserting colors with an iterative algorithm, which started with a random assignment. The codeword of pattern feature was dependent on the feature color itself and its six surrounding color elements. Song and Chung[16,17] proposed a color diamond pattern with four colors. The grid-points between adjacent rhombic shapes were defined as the feature points. The pattern size was 65×63 with a window size of 2×3. The intersection points of two adjacent rhombic shapes are defined as the feature points. Chen et al.[18] designed a color square pattern with seven colors. The pattern feature was encoded by its four-adjacent colors of pattern elements. The pattern size was 38×212, and the unique window size was 2×2. This method provided a relative small coding window size, but using seven colors made it lack robustness in dealing with the surface color fusions.

To improve the robustness of color coding methods, the binary shapes can be used to replace the color cues in the pattern generation. The binary shapes can be circle, disc, stripe,[19] thickened cuneiform,[21] thinned cuneiform,[22,23] polygon,[24] or specially designed shapes.[20,25] Albitar et al.[19] adopted binary shapes instead of colors as the coding elements to generate a binary pattern based on M-array. The proposed pattern consisted of three geometrical shapes. The pattern size was 27×29, and the coding window size was 3×3. Reiss and Tommaselli[21] improved the coding volume with five different shapes; each shape owned four or six points for surface reconstruction. Maurice et al.[22,23] presented a perfect submap generation with large Hamming distance. However, the coding window size of 3×3 decreased the code-correction ability for the scenes with depth discontinuities. Xu et al.[24] utilized the corner of the chessboard as the primitive to produce the pattern. Moreover, the orientation of the corner was used to encode the primitive. Since the primitive owned perfect symmetry, the position of the feature point could be accurately located. Jia et al.[20] used five special shapes in an M-array pattern with dimensions of 79×59 with a coding window sized 2×2. This method gained a dense mass of key points because each shape had six points. Fang et al.[25] presented a symbol density spectrum (SDS) to choose geometrical shapes for improving resolution and decreasing decoding error. The proposed SDS method provided a distribution of feature points for reconstruction after 10 geometrical shapes were extracted. Then, a comparative analysis of the shape features and scene testing of shapes damage rate were conducted to choose nine geometrical shapes from one group to form a density pattern. The 3-D reconstruction experiment showed that this method owned high resolution and robustness.

Most of research has focused on how to encode the position information with color code or shape code. However, less attention is paid to another essential problem, decoding the correspondence from the captured image. As Boyer and Kak[26] pointed out, the structured light system is similar to a digital communication system; the information can be successfully transmitted to the receiver only after correctly decoding. A large amount of error in decoding can destroy the 3-D reconstruction. So decoding is more important for successful shape acquisition. For the color coding schemes, the hue, saturation, value model is usually adopted[16,17] and the simple thresholding method[10,26] is applied to identify the color of each coding element. In addition, some machine learning-based approaches are also attempted for pattern decoding. For example, Zhang et al.[8] identified the color of color multisilt using the K-means clustering algorithm on a proposed color feature named regularized RGB. Comparative experiments showed that regularized RGB has higher discriminating power in color identification than other color features, such as RGB, HSI, Nrgb, $c_1c_2c_3$, H*S*, CIElab, and so on.[9] Tang et al.[3] employed the fuzzy c-means clustering algorithm on color feature $c_1c_2c_3$ to identify the color of color stripe and further demonstrated that a color feature only related to the spectral sensitivity of red, green, and blue sensors and the albedo of the surface owns more excellent performance in color identification than that related to the spectral sensitivity of red, green, and blue sensors, the albedo of the surface, the direction of the illumination source, the normal of the surface, and the spectral power distribution of the incident light no matter what the color of the test object is. For the shape coding schemes, although the usage of binary shapes makes the system more robust to surface color or textures, the projective distortion of pattern elements also brings difficulties to the decoding task. Image segmentation is usually applied to segment each pattern element, and the template matching is usually used to identify the pattern elements.[19–25] But the performance of pattern decoding is inferior when the pattern elements are greatly affected by complex factors, such as surface color, textures, distortion, reflections, and so on.

With the above review, we can see that increasing the number of colors or pattern elements can decrease the coding window size with a given coding volume. A small coding

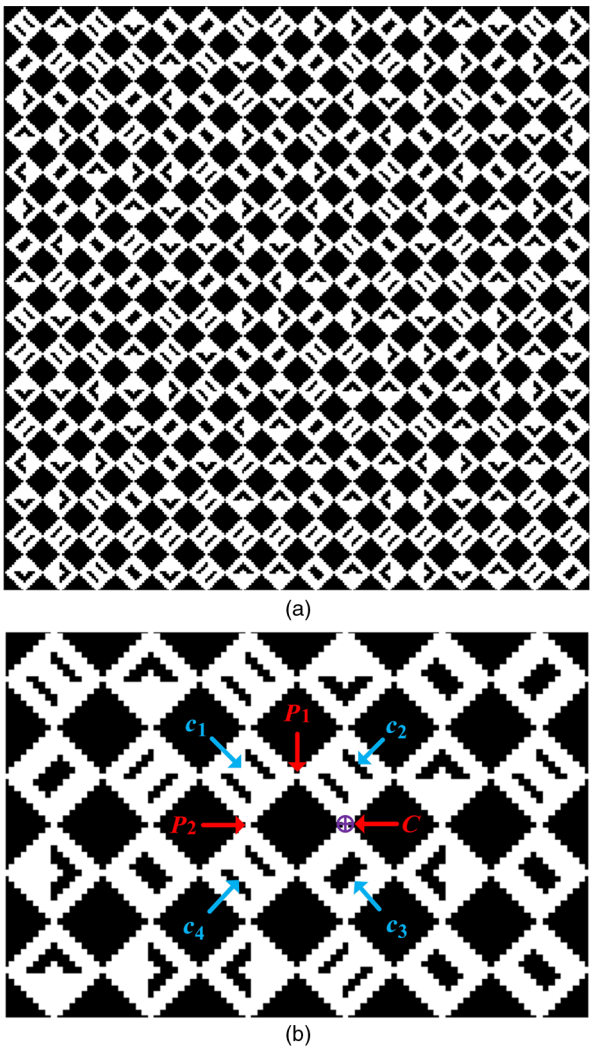

Fig. 1 The proposed binary geometrical pattern: (a) a part of the generated binary geometrical pattern and (b) indication of two types of feature points, P_1 and P_2.

window size indicates that fewer elements should be decoded to determine one codeword and thus brings benefits to the decoding stage. On the other hand, some machine learning-based approaches are also attempted for pattern decoding, but the results are still quite dependent on the surface colors and lack of robustness. To realize a robust spatial structured light method, not only the feature detection algorithm but also the decoding algorithm should be well studied.

3 Pattern Generation

The proposed pattern is pseudorandom array based. A pseudorandom array can be generated from a pseudorandom sequence with folding rule, and a pseudorandom sequence can be created by a primitive polynomial.[27] To make the pattern more robust to surface color and reflectance, shape codes are selected instead of color codes. Since small window size can alleviate the complexity of the decoding algorithm, a binary geometrical pattern with the window size of 2×2 is proposed in this paper, as shown in Fig. 1. It is obtained in the following way. A primitive polynomial $h(x)$ defined over Galois field with eight elements [GF(8)] is first used to generate a pseudorandom sequence

$$h(x) = x^4 + x + \alpha^3. \tag{1}$$

The sequence is computed using the following equation:

$$\alpha^3 + \alpha + 1 = 0, \quad \alpha^7 = 1. \tag{2}$$

Every nonzero element of GF(8) is a power of α, which is a primitive element, and each element in GF(8) is a binary linear combination of $\{1, \alpha, \alpha^2\}$. Based on the above primitive polynomial, a pseudorandom array of size 65×63 can be acquired with the window size of 2×2. Since there are eight primitives in the pseudorandom array, eight different geometric primitives are demanded to design the projected pattern. To make the pattern elements more distinguishable, the geometric primitives with great difference are designed as shown in Fig. 2 and are embed into the white rhombic shape with the color black used as the background. Moreover, the intersection points formed by two neighboring pattern elements are defined as the feature points and named as the grid-points. The grid-points include two types. The first type of grid-point is P_1, as shown in Fig. 1(b), and is constructed by two adjacent pattern elements at the horizontal direction. The other type of grid-point is P_2, which is formed by two adjacent pattern elements at the vertical direction. The two types of grid-point P_1 or P_2, as shown in Fig. 1(b), have the same code value of c_1-c_2-c_3-c_4.

4 Detection of the Grid-Points

To localize the grid-points accurately and robustly, it is essential to develop an effective grid-point detector. Inspired by the cross template feature detector,[16,17] an X-shape filter is investigated for the grid-point detection in the proposed structured light system. By filtering the image with the proposed feature template, a responding map can be generated. The centers of the shape to be detected can be found by finding the local maxima in the map. In addition, adaptive nonmaximum suppression method[28] and twofold rotation symmetry are also used to exclude the false points.

4.1 *Design of the Grid-Point Detector*

The position of the grid-point can be approximately expressed by a binary matrix. Suppose the radius of the local square centering at a grid-point is r, then the size of the matrix is $(2r+1) \times (2r+1)$. Accordingly, the (i, j) element in the local matrix for P_1 grid-point can be expressed as

$$T_1(i,j) = (i-j \geq 0 \wedge i+j \geq 0) \vee (i-j \leq 0 \wedge i+j \leq 0). \tag{3}$$

Noted that the index of the central element in the matrix is (0, 0). Similarly, the (i, j) element in the local matrix for P_2 grid-point can be expressed as

$$T_2(i,j) = (i-j \geq 0 \wedge i+j \leq 0) \vee (i-j \leq 0 \wedge i+j \geq 0). \tag{4}$$

An illustration of the proposed filters T_1 and T_2 is shown in Fig. 3. If these two filters are applied directly to the captured image, a normalized correlation[29] will be required. However, the process of normalization is time-consuming. To solve the problem, a new template is designed by combining T_1 and T_2 as

$$T_0 = T_1 - T_2. \tag{5}$$

Basic	0	1	α	$\alpha+1$	α^2	α^2+1	$\alpha^2+\alpha$	$\alpha^2+\alpha+1$
Geometric Primitives								

Fig. 2 Geometric primitives of the projected pattern.

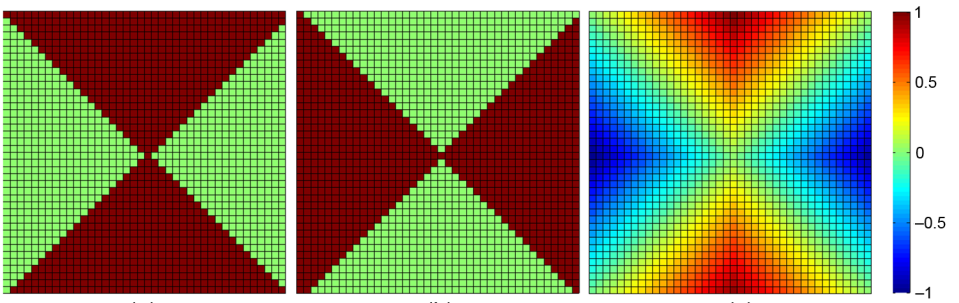

Fig. 3 Illustration of the filters T_1, T_2, and T_3: (a) local matrix of the filter T_1, (b) local matrix of the filter T_2, and (c) local matrix of the filter T_3. The radius is set to 20.

With the new template, positive maximal points will be the P_1 grid-points, and the negative ones will be the P_2 grid-points.

Considering that local areas centering at the grid-points will suffer from deformation due to the projective distortion and surface curvature, it is necessary to improve the robustness of the template. In practice, if a point in the standard local area centering at a grid-point is more distant to the two diagonal lines, its corresponding point in the captured image is less likely to change its property. Therefore, it is reasonable to increase the weight of the template elements that are distant from the two diagonal lines in the template. Consequently, the weight can be set to be linearly proportional to the distance, which can be formulated as

$$T_3 = (i \geq 0 \wedge j \geq 0) \times (i-j) - (i \leq 0 \wedge j \geq 0) \times (i-j) + (i > 0 \wedge j < 0) \times (i+j) - (i < 0 \wedge j > 0) \times (i+j). \tag{6}$$

Figure 3 visually illustrates T_3; the template T_3 is normalized by its radius. Suppose the captured image is I_0, the first step of grid-point detection is to adopt a Gaussian template to filter I_0 as a smoothing process

$$I_1 = G \otimes I_0, \tag{7}$$

where G is a Gaussian template. The next step is to use the designed template to filter I_1 as

$$H = T_3 \otimes I_1, \tag{8}$$

where H is the aforementioned responding map. Based on the map, the positive maximum points and negative maximum points can be located. Then, the adaptive nonmaximum suppression is applied to remove the false points separately. The type of a grid-point can be decided by its sign in H. Specifically, if its sign is positive it will be classified into P_1 type, otherwise P_2 type. Although the grid-points can be detected with the above operations, the false points may still exist in the candidate points. Twofold rotation symmetry is displayed at the positions of true grid-points. This can be used for confirmation of the grid-point features. For each candidate point, a circular image region C was chosen, and the coefficient of correlation between C and its 180 deg rotation was applied to measure the strength of the twofold symmetry at the candidate points as

$$\delta = \frac{\sum_m \sum_n (C_{mn} - C'_{mn})^2}{\sum_m \sum_n (C_{mn} - \bar{C})^2}, \tag{9}$$

where C is a circle region centered at a candidate point, C' is created by rotating C with 180 deg, $\bar{C}$ is the average image intensity of C, and m and n indicate the local pixel index inside C. The size of C is set to be a half of an element. The above equation uses the mean of square difference between corresponding pixels in C and C' to represent their difference. The variance distribution inside C is used to normalize the difference.

4.2 *Multitemplate Filtering Strategy*

Subject to the projective distortion and surface curvature, the projected elements are usually enlarged or compressed. Great distortions of the imaged pattern elements bring challenges to feature detection. To make the proposed feature detector more flexible and robust, a multitemplate filtering strategy is introduced, which can be performed with the following steps.

1. Apply multiple templates with a sequence of sizes to obtain the corresponding candidate point set.
2. Judge whether a candidate point is the true grid-point or not according to the number of templates detecting it. If the number is larger than a given threshold, the point is considered the true grid-point.

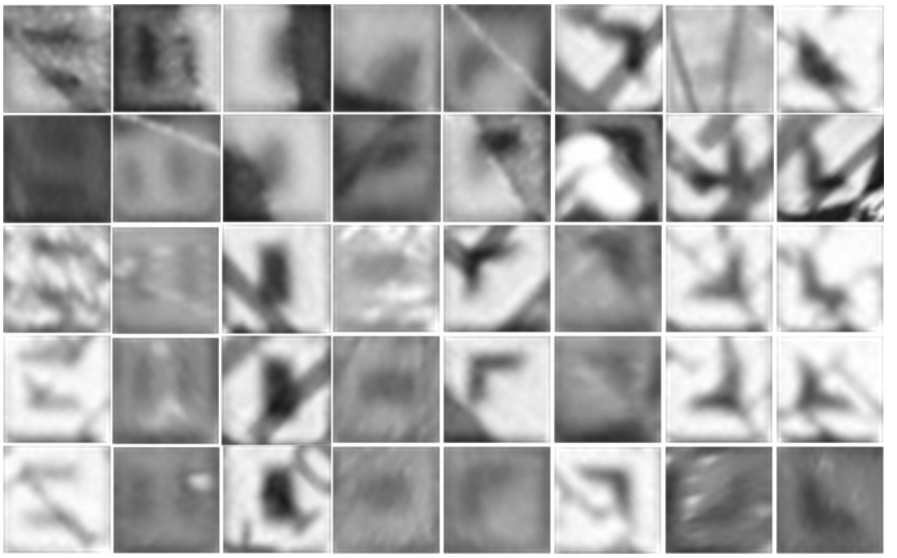

Fig. 4 Sample images of the pattern elements with blurring and distortion.

5 Deep Decoding of the Binary Structured Light Image

The pattern elements in the captured image are often blurred or distorted as shown in Fig. 4 because of some complex factors, such as plentiful color, rich texture, surface discontinuity, specular reflection, and sharp change. It is very challenging to detect and recognize the degraded pattern elements for traditional feature detectors.[19–25] Since the pattern elements are designed as a rhombic shape in our pattern, a graph can be generated by connecting four grid-points of the pattern element. Then, by collecting abundant pattern elements with blurring and distortions, an extensive training dataset can be set up for convolutional neural networks. Thus, the pattern elements can be recognized.

5.1 *Extraction of Pattern Elements*

Since the window size is only 2×2 and each grid-point is formed by two pattern elements, two adjacent grid-points can determine a unique window as well as the codeword, and two such adjacent grid-points are named as a pair-point. However, it is difficult to find a pair-point from the captured image directly because of the distortion of the pattern elements. To address this problem, a topological network is established. According to the sign of H that is computed from Eq. (8), the grid-points can be classified into two types: P_1 (blue) and P_2 (red), as shown in Fig. 5. A grid-point B is surrounded by four different type grid-points C, D, E, and F.

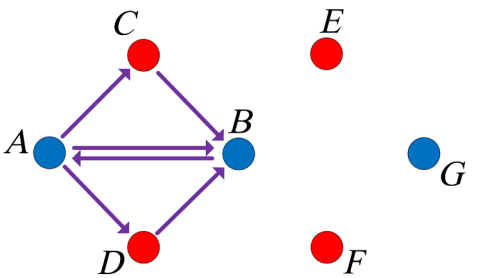

Fig. 5 A topological map of various types of grid-points.

For each P_1 type grid-point, its nearest different type grid-points can construct a quadrant. The same procedure is also applicable to P_2 type grid-points. With these quadrants, a topological network of grid-points can be constructed. From this network, the pair-point of each grid-point can be deduced. For example, if the pair-point of A is to be found, i.e., the grid-point B, the first step is to find out its different type grid-point C in the upper right corner. Then, the lower right different type grid-point of C is A's pair-point B. In this way, a topological network of all the grid-points can be established.

Based on the established grid-point topological network, each rhombic pattern element can be detected. Then, assume that the target surface is relatively smooth, i.e., the surface patch covered by one pattern element can be approximately viewed as a planar patch. On this, the distorted and blurred pattern element can be transformed into a normalized image with four grid-points around it. This procedure can be expressed as follows:

$$\begin{bmatrix} u_{\mathrm{pt}} \\ v_{\mathrm{pt}} \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{\mathrm{im}} \\ v_{\mathrm{im}} \\ \mu \end{bmatrix}, \tag{10}$$

where $u_{\mathrm{pt}}, v_{\mathrm{pt}}$ indicates the detected grid-points and $(u_{\mathrm{im}}, v_{\mathrm{im}})$ denotes the four normalized image corner points $(0, 0)$, $(a, 0)$, (a, b), and $(0, b)$. Given four pairs of points $(u_{\mathrm{pt}}, v_{\mathrm{pt}})$, $(u_{\mathrm{im}}, v_{\mathrm{im}})$, the matrix of projective transformation can be exactly solved. Then, the distorted pattern elements can be projected to the normalized image via bilinear interpolation.

5.2 *Pattern Element Identification via Deep Neural Networks*

As the pattern elements in the captured image are usually affected by various surface factors, it is necessary to collect enough labeled data for the training of deep neural networks. As a result, eight geometrical pattern elements are projected onto the experimental targets, respectively. The experimental targets include low-contrast balloon, dummy model, brilliant

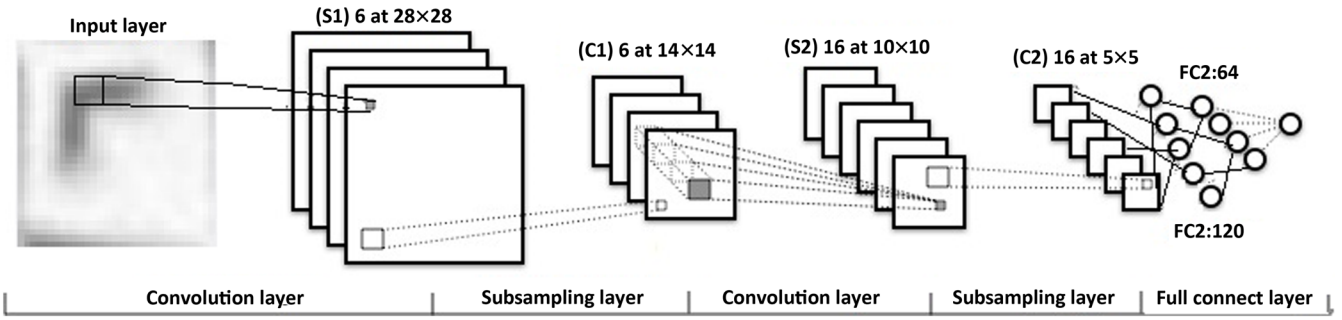

Fig. 6 The adopted network architecture for the classification of binary geometrical pattern elements.

piggy, colorful cover, dark box, textured paper, real human face, and so on. Yet, the database is still small because the pattern numbers within an image are limited. It is necessary to augment the database to achieve higher discriminating power. Our operation is described as follows:

1. Sharpness of the training samples is calculated, and Gaussian noise is added in high-contrast samples.
2. Random white/black lines are added to the samples to simulate the occlusion problem.
3. Small affine transformation is applied to simulate small localization error of grid-points.

The number of original training samples is about 80,000. With the above operations, the number of training samples can be augmented to more than 300,000. Since the illumination and contrast variations are varied for different regions in the captured image, the typical principal component analysis (PCA) whitening procedure in the deep neural network is adopted to eliminate the pixel correlation and to normalize the illumination deviation. First, the covariance matrix of the training data is computed as

$$\sum = \frac{1}{m}\sum_{i=1}^{m}(x_i - \varpi)(x_i - \varpi)^T, \tag{11}$$

where is the i'th training data and ϖ denotes the average value of training data. Then, the singular value decomposition of covariance matrix is conducted. The data are rotated and normalized to unit variance in every dimension

$$x_{\mathrm{rot},i} = \frac{U^T(x - \varpi)}{\sqrt{\lambda_i}}, \tag{12}$$

where U indicates the PCA rotation matrix and is the singular value of the training data matrix.

After collecting the training dataset, the classification of pattern elements can be conducted. Since the pattern classification task in our work is similar to the handwritten digit recognition problem, and the Lenet-5[30] has more excellent performance in dealing with such a problem than traditional shallow architectures, e.g., multilayer perceptron (MLP) and support vector machine), in this work, the Lenet-5 is adopted to classify the pattern elements. The architecture of Lenet-5 is shown in Fig. 6. The network architecture is composed of two convolutional subsampling layers (C1-6 maps with 5×5 kernel and 2×2 max pooling, C2-16 maps with 5×5 and 2×2 max pooling) and two full-connected layers (128 and 84 neuron units), and the final class probability is generated by radial basis function. With the convolutional neural networks, high recognition rate can be obtained in the decoding algorithm.

5.3 *Optimization of Decoding Result*

Subject to the surface color or textures, it is inevitable that some pattern elements are erroneously identified. Thus, the false correspondences emerge after conducting window matching.[14] To prune the false correspondences, an optimization mechanism that includes two decoding reliability terms is introduced as follows.

The first decoding reliability term is calculated based on epipolar constraint.[31] Suppose O_c and O_p express the optical centers of the camera and projector, respectively, and X_c and X_p denote two corresponding points on the camera and projector image planes, respectively. According to the epipolar constraint principle, the vectors $\overrightarrow{O_pX_p}$, $\overrightarrow{O_cX_c}$, and $\overrightarrow{O_pO_c}$ are in the same plane, which can be expressed as follows:

$$\overrightarrow{O_pX_p} \cdot [\overrightarrow{O_pO_c} \times \overrightarrow{O_cX_c}] = 0. \tag{13}$$

The intrinsic parameters O_c and O_p and rotation and translation parameters R and T can be acquired with the structured light system calibration method. By expressing X_c, X_p with the homogeneous form $\bar{X}_c$ and $\bar{X}_p$, respectively, the following equation can be obtained:

$$\bar{X}_p \cdot (T \times R\bar{X}_c) = 0. \tag{14}$$

The epipolar line $l = (a, b, c)^T$ can be expressed as

$$l = T \times R\bar{X}_c = [T] \cdot (R\bar{X}_c). \tag{15}$$

For X_p, it can be precisely localized in the projector image plane. For $X_c(u, v)$, its distance to the epipolar line can be calculated as

$$d = \frac{|au + bv + c|}{\sqrt{a^2 + b^2}}. \tag{16}$$

If d_c is larger than a given threshold value, the grid-point is viewed as a wrong decoding point.

The second term is computed based on neighboring constraint. Suppose (X_{c0}, Y_{c0}) is a grid-point in the camera image, its adjacent grid-point (X_{ci}, Y_{ci}), $i = 1 \ldots n$ can be found in a predefined local image region. Since the

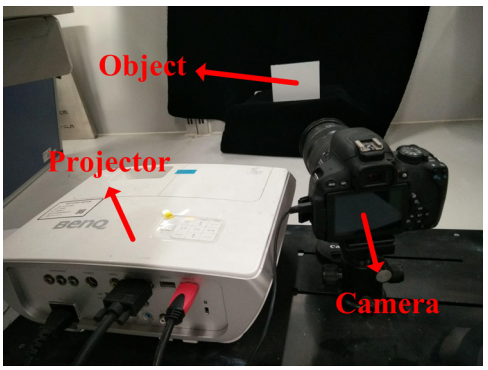

Fig. 7 The experimental structured light system setup.

codeword of several adjacent grid-points is associated, their corresponding points (X_{p0}, Y_{p0}) and (X_{pi}, Y_{pi}), $i = 1 \ldots n$ can also be found in the projector pattern. Sequentially, the correlation degree of between one grid-point and its neighboring grid-points can be calculated as

$$\sigma_i = e^{-\frac{[(X_{Pi}-X_{P0})^2+(Y_{Pi}-Y_{P0})^2]}{9}}, \quad i = 1 \ldots n. \tag{17}$$

If σ_i is a relative small value, (X_{p0}, Y_{p0}) has a long distance to its neighboring grid-point (X_{pi}, Y_{pi}), $i = 1 \ldots n$ in the projector pattern. That means decoding errors occur at the point (X_{p0}, Y_{p0}) or (X_{pi}, Y_{pi}), $i = 1 \ldots n$. Assume all neighboring grid-points (X_{pi}, Y_{pi}), $i = 1 \ldots n$ have the same influence on the point (X_{p0}, Y_{p0}), the primary decoding reliability of (X_{c0}, Y_{c0}) can be expressed as

$$\phi = \sum_{i=1}^{n} \sigma_i / n. \tag{18}$$

Each decoded grid-point can be associated with a primary decoding reliability . To improve the overall decoding reliability, for the adjacent points (X_{pi}, Y_{pi}), $i = 1 \ldots n$ of (X_{p0}, Y_{p0}), the decoding reliability of can be calculated as

$$\Phi = \sum_{i=1}^{n} \phi_i \sigma_i / \sum_{i=1}^{n} \phi_i. \tag{19}$$

According to above decoding reliability terms, most of the false correspondences can be identified and removed.

6 Experiments and Results

The experimental platform consisted of a projector with a resolution of 1920×1080 pixels (Benq W1060) and a camera with a resolution of 5184×3456 pixels (Canon EOS 700D with EFS 18- to 135-mm lens), as shown in Fig. 7. The working distance of the system is about 730 mm. In the projected pattern, the size of each pattern elements is 16×16 pixels. The collected image data are processed on a computer with Quad-Core processors (Intel Xeon E5-1620 3.60 GHz) and 8-GB RAM (DDR3 1600 MHz). The structured light system is calibrated with the method in Ref. 32. The calibration procedure mainly includes five steps. A pattern with known dimensions on the liquid crystal display (LCD) panel is first shown to the camera and imaged. Zhang's method[33] is then adopted for camera calibration. By introducing the homography constraint between camera image plane and calibration plane, the position of the calibration plane with respect to the camera is determined. With the spatial position and orientation of the LCD panel kept still, a known pattern is projected onto the LCD panel by the projector. The reflection from the panel is then imaged by the camera, and the image data are used to calibrate the projector;thus, the system calibration is accurately completed.

After system calibration, the following three experiments are conducted on the system to test the feasibility, precision, and robustness of the proposed method. The first experiment is to illustrate the proposed feature detection algorithm with a spherical surface. Then, the classification accuracy and measurement precision of our method are evaluated. Finally, some complex objects with plentiful color, rich texture, or surface discontinuity are selected to test the robustness of our method.

6.1 *Test of Feature Detection*

A spherical surface is chosen as the target to evaluate the proposed feature detection algorithm. With the X-shape template method, the grid-points can be detected as shown in Fig. 8(a). It is evident that there are some false points

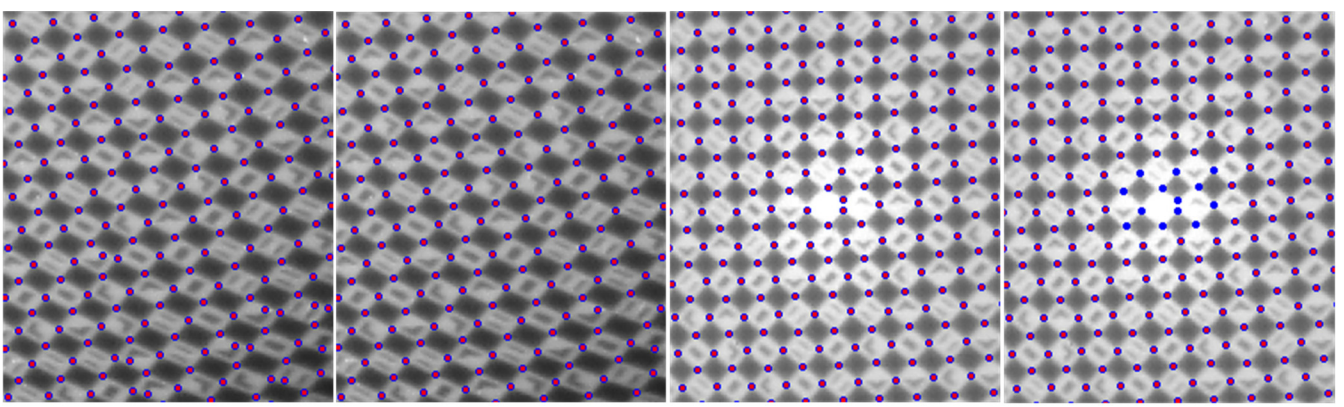

Fig. 8 Evaluation of the proposed grid-point detection method: (a) detection result with the X-shape template method, (b) detection result with twofold rotation symmetry, (c) detection result in the saturated image area, and (d) decoding performance of small window size of 2×2.

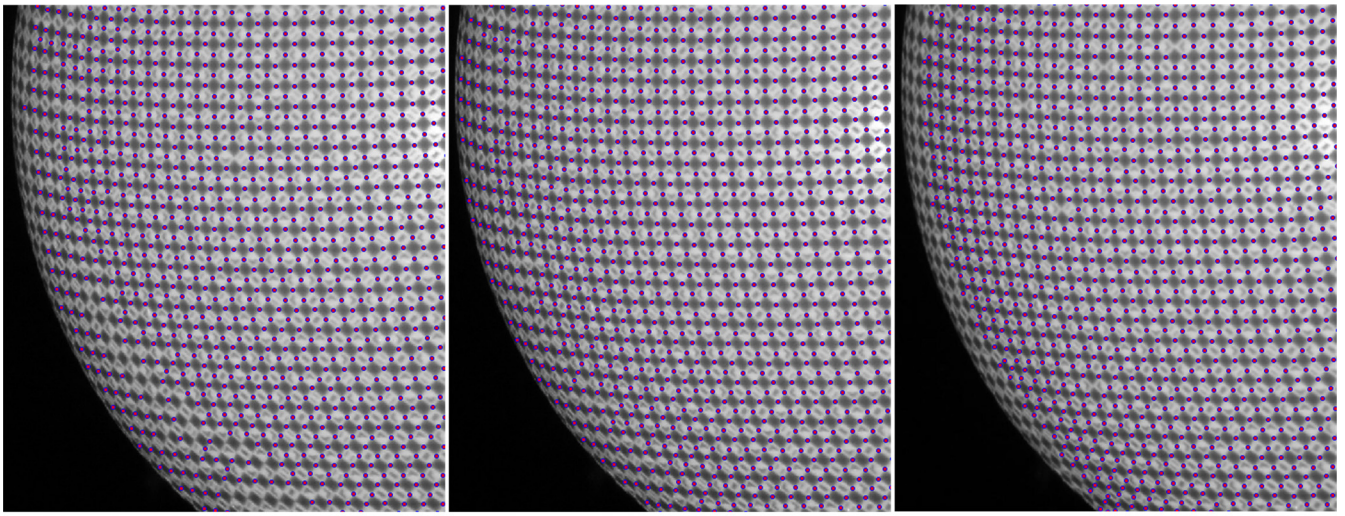

Fig. 9 Images of grid-point detection with three different detection methods: (a) detection result with the proposed single-template detection method, (b) detection result with the detection method in Refs. 16 and 17, and (c) detection result with the proposed multitemplate detection method.

among the detected points. It is because the feature detector is based on a nonmaximum suppression method. Figure 8(b) shows the result after using the rotation symmetry-based feature detector. It is obvious that most of the false points are removed. However, when the object surface owns high reflectance, the false points are hardly removed, as shown in Fig. 8(c). This is reasonable because the rotation symmetry with 180 deg is perfect in the C region. In addition, the pattern information is not absolutely clear in this saturated area. For this case, the small window size can demonstrate its advantage. Compared with a larger window size of 2×3 or 3×3, the small window size of 2×2 used in this paper

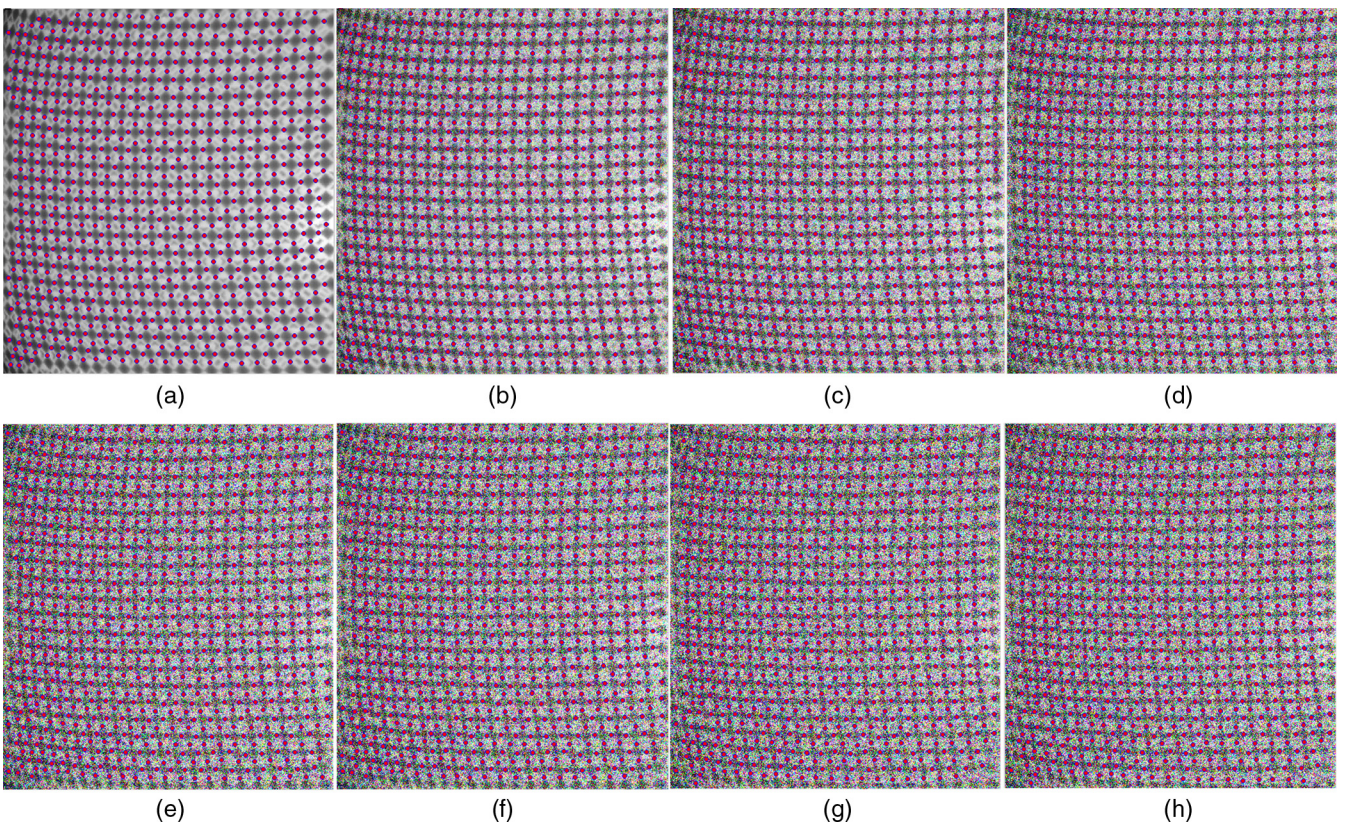

Fig. 10 Images of feature detection on different zero-mean Gaussian noises. The standard deviations of Gaussian noise from (a) to (h) are set to 0, 0.05, 0.10, 0.16, 0.20, 0.26, 0.33, and 0.41, respectively.

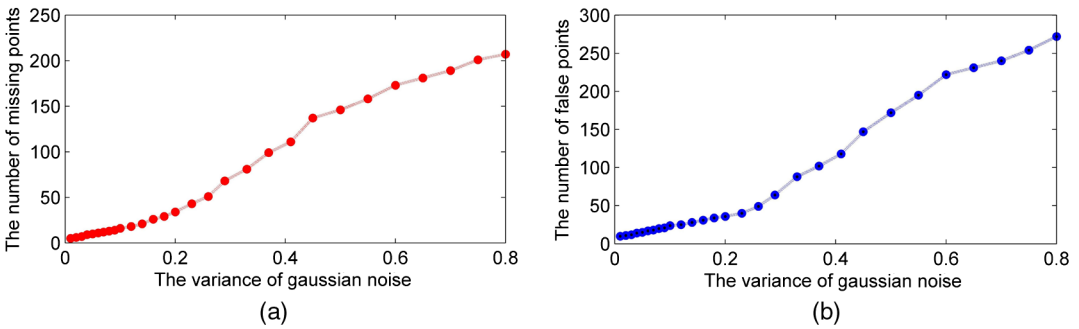

Fig. 11 Robustness evaluation of the proposed multitemplate grid-point detector. (a) The number of missing points with respect to the variance of Gaussian noise and (b) the number of false points with respect to the variance of Gaussian noise.

can be less sensitive to the surface condition. In other words, the decoding result can be less affected by this saturated image area, as shown in Fig. 8(d).

To prove the superiority of the proposed multitemplate feature detection algorithm, the method in Refs. 16 and 17 and single-template feature detection algorithm are compared. Figure 9 displays the grid-point detection results with these detection methods. It is evident that the number of detected grid-points with the multitemplate feature detection method is larger than that with other two methods. This indicates that the multitemplate feature detection method has better performance than the others. It is reasonable because the multitemplate feature detection method can provide a suitable template for grid-point detection in different regions, while the other two methods only have one template for grid-point detection in the region within a fixed surface curvature. To evaluate the robustness of our feature detection method, the extra Gaussian noise is added into the captured image. As shown in Figs. 10(a)–10(j), the standard deviations of Gaussian noise are set to 0, 0.05, 0.10, 0.16, 0.20, 0.26, 0.33, and 0.41, respectively. From these pictures, it can be seen that most of the grid-points can be successfully detected when the standard deviation of Gaussian noise is less than 0.20, and the rhombic shape can also be recognized

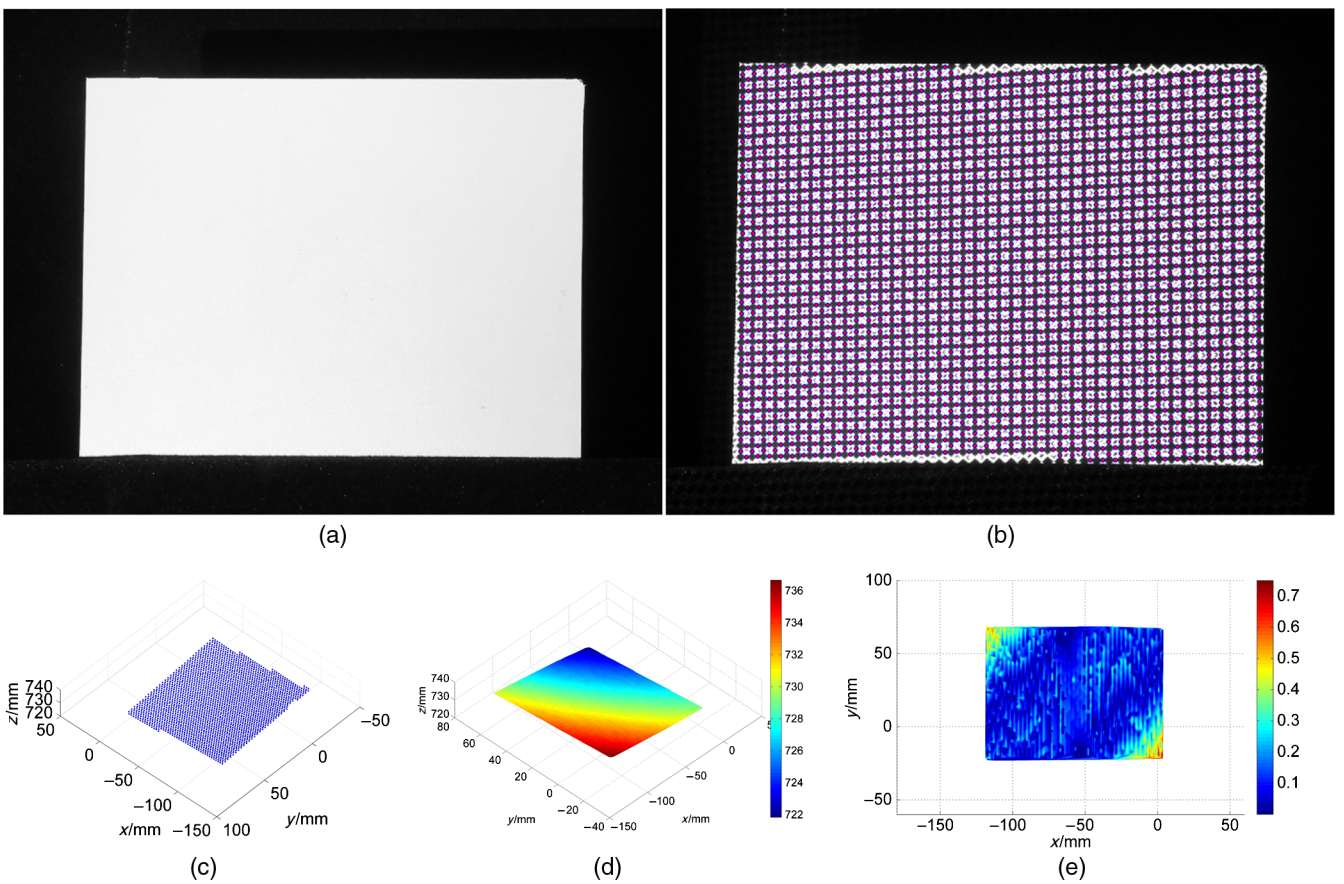

Fig. 12 3-D reconstruction of a standard plane: (a) the target, (b) result of grid-detection, (c) 3-D points, (d) result of depth reconstruction, and (e) map of depth error.

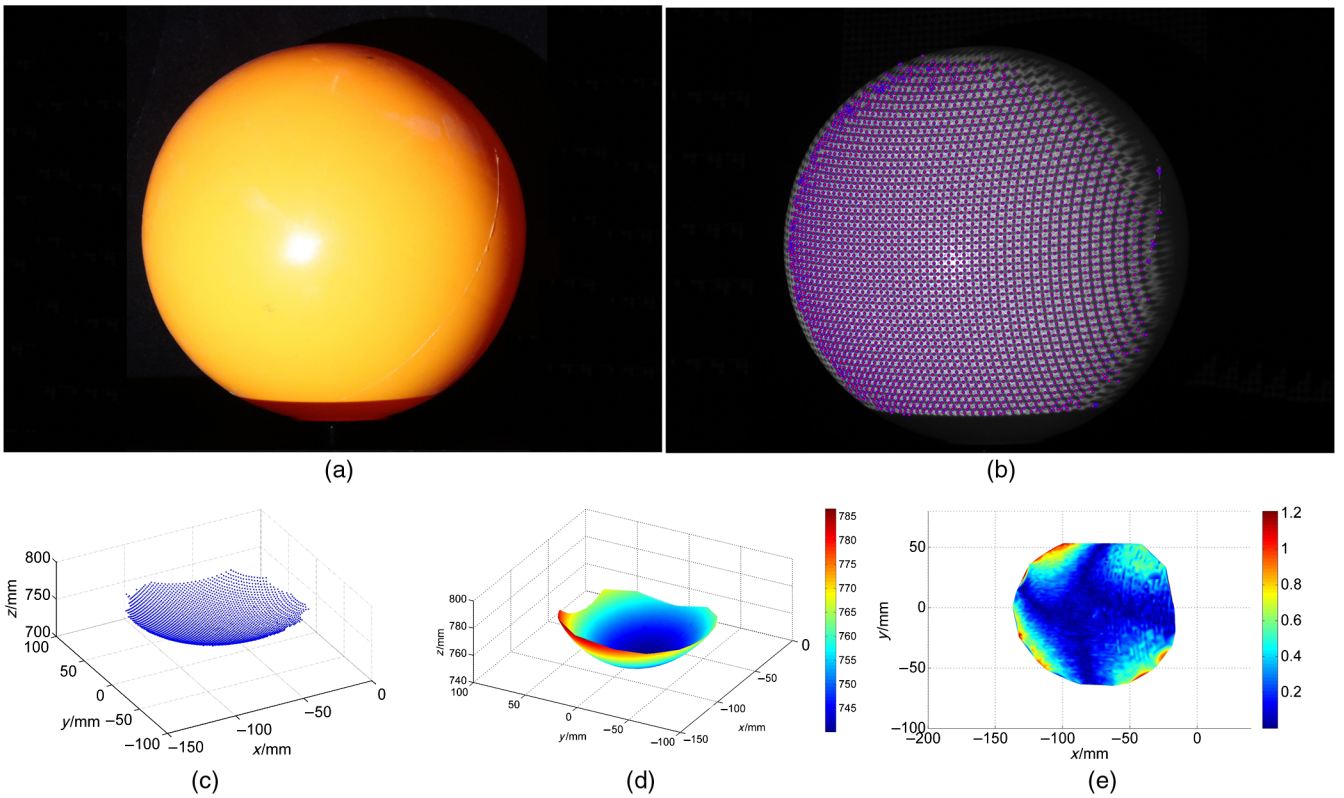

Fig. 13 3-D reconstruction of a standard sphere: (a) the target, (b) result of grid-detection, (c) 3-D points, (d) result of depth reconstruction, and (e) map of depth error.

roughly. For each detected point in the noise-free image, look for its nearest detected point in a noise image. If the distance between them is larger than 5 pixels, then the point is regarded as a missing point. If the distance between them is larger than 3 pixels, then the point is viewed as a false point. Figure 11 shows the numbers of missing points and false points in a noise image with respect to the variance of Gaussian noise. It is obvious that, with the increase of Gaussian noise, the number of missing points and false points in the given area increase, the missing rate is about 3.22%, and the false rate is about 3.74% when the standard deviation of Gaussian noise is 0.20. The experimental results show that the proposed multitemplate grid-point detection method has excellent robustness to image noises.

6.2 *Evaluation of Classification Accuracy and Measurement Precision*

As the objective of classifying the pattern elements is to identify their corresponding codeword, one way of evaluating the performance of our classification method is to calculate the classification accuracy. In the implementation, the leave-one-out method is adopted to compute the average accuracy by splitting the training dataset into 10 folds. Stochastic gradient descent is employed for the training with mini-batch 100. Weight decaying and dropout probability of 0.5 in the last full-connected layers are also utilized in the recognition. The MLP is tested with sigmoid actuation, the Lenet-5 network, and Lenet-5 on augmented training database. The experimental result shows that Lenet-5 net can obtain

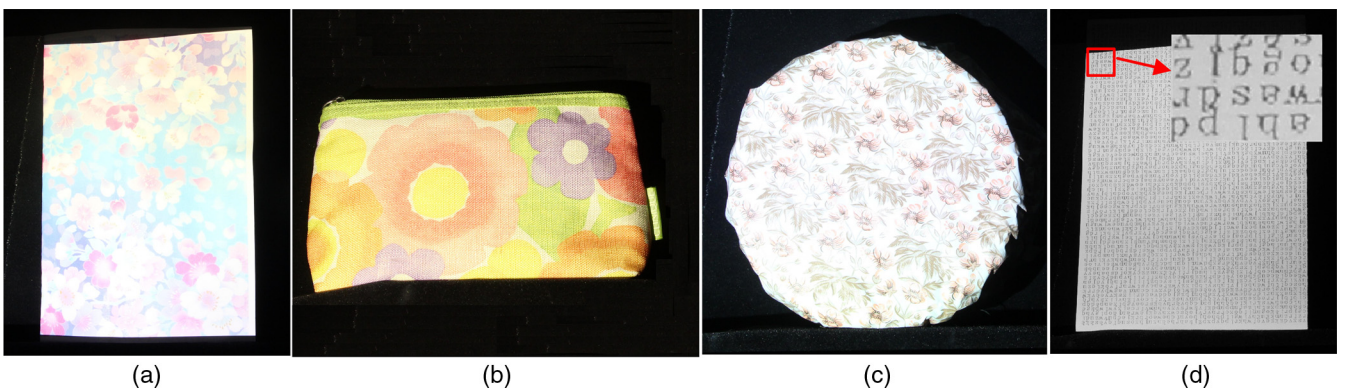

Fig. 14 Four measured objects: (a) colorful paper, (b) colorful bag, (c) colorful and textured hat, and (d) textured paper.

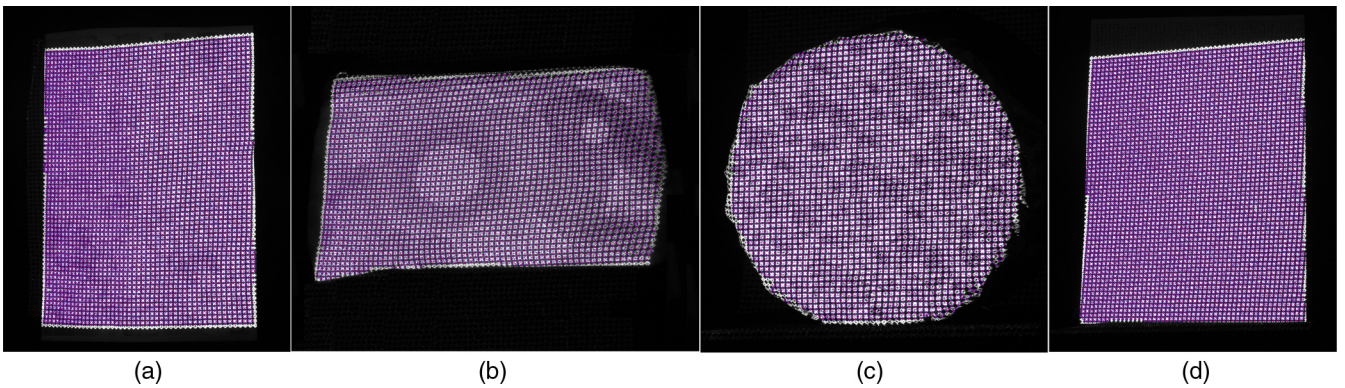

Fig. 15 Results of grid-point detection for all the measured objects: (a) colorful paper, (b) colorful bag, (c) colorful and textured hat, and (d) textured paper.

a classification accuracy of about 97.9%; in comparison, the MLP method get an accuracy of about 95.5%. With the augmented training database, the classification accuracy of Lenet-5 net can be slightly improved to 98.7%.

To evaluate the 3-D reconstruction precision, the standard plane and sphere with the radius of 81.5 mm are selected as the target objects as shown in Figs. 12(a) and 13(a), respectively. Using the proposed pattern decoding method, the correspondences for these two objects can be obtained. Then, the point-clouds can be transformed from the correspondences through Delaunay triangulation, as shown in Figs. 12(b) and 13(b). Because the obtained 3-D points, as shown in Figs. 12(c) and 13(c), are not too dense, the bilinear interpolation method is adopted to get dense point-clouds for these two objects. With the 3-D information in Figs. 12(d) and 13(d), a plane and a sphere can be fitted

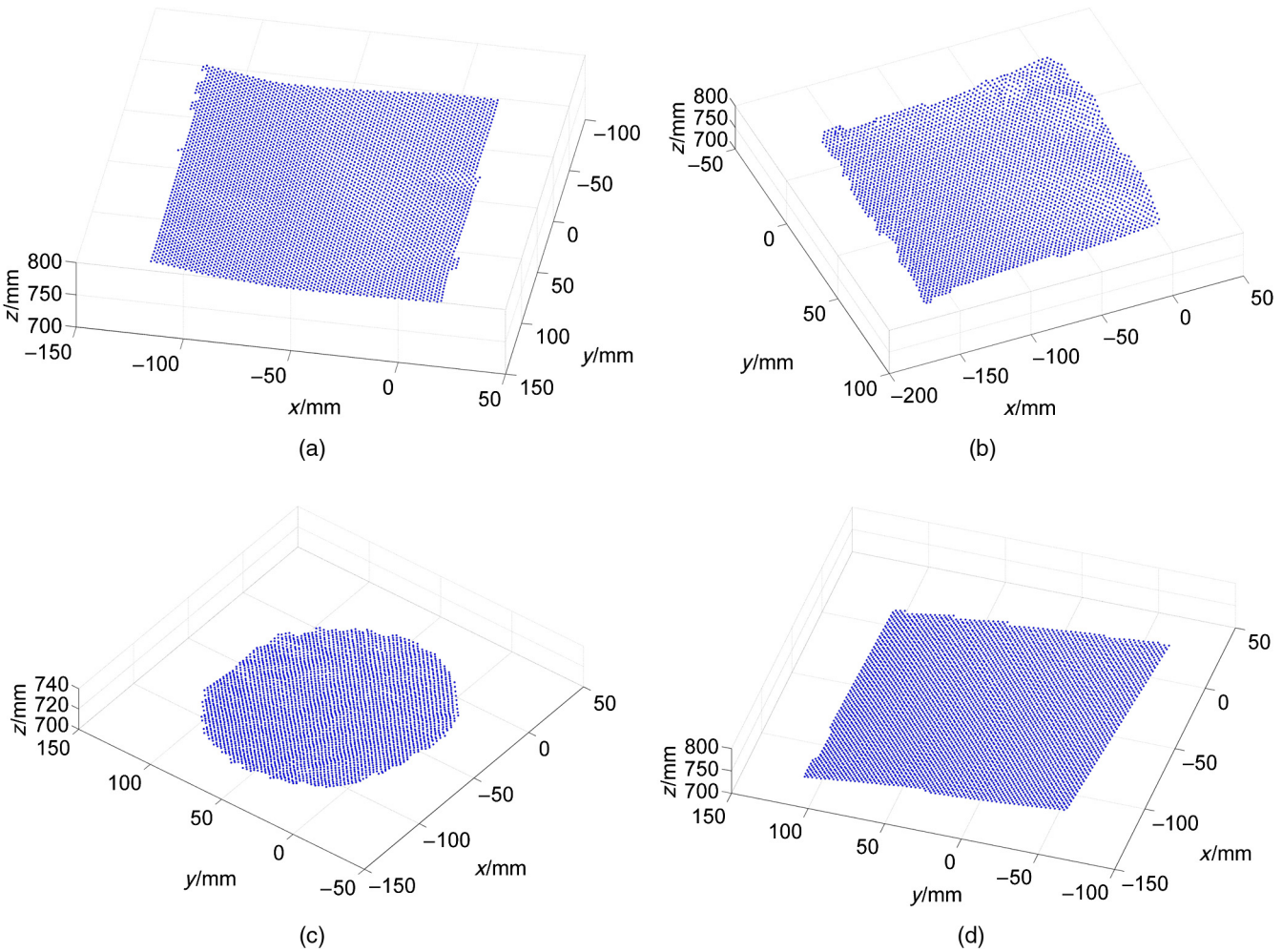

Fig. 16 3-D point-clouds for all the measured objects: (a) colorful paper, (b) colorful bag, (c) colorful and textured hat, and (d) textured paper.

Table 1 Measurement results of four complex objects.

Objects	Working distance (mm)	Measurement area (mm^2)	Number of 3-D points	Measurement time (ms)
Colorful paper	750	28,000	5835	3281
Colorful bag	735	20,900	3873	2876
Colorful and textured hat	720	19,400	3789	2592
Textured paper	733	30,400	5828	3134

Note: Measurement area denotes the actual area of the target and measurement time denotes the computation time of grid-point detection and pattern decoding without the help of GPU computing.

with the least square fitting method, respectively. The measured radius of the sphere is about 81.3124 mm. Based on the fitted plane and sphere, the depth errors for these two regular objects can be obtained, as shown in Figs. 12(e) and 13(e). Thus, the mean errors and standard deviations can be easily computed. The results show that the mean error and standard deviation of the plane are 0.1144 and 0.0917 mm, respectively, and those of the sphere are 0.2410 and 0.2008 mm, respectively.

6.3 *Three-Dimensional Reconstruction of Complex Surfaces*

Since the surface color and texture often affect the reconstruction quality for spatial coded structured light method, several complex objects are chosen to test the performance of our method in this section. The first two objects in Figs. 14(a) and 14(b) are a paper and a bag; they have plentiful color. The third one in Fig. 14(c) is a hat with light color and weak texture. The forth object in Fig. 14(d) has a rich texture. Generally, it is difficult to obtain the 3-D information of the objects with rich color or complex texture for conventional color-based structured light method because the surface color or texture always affects feature detection and pattern decoding. However, the binary geometrical pattern is not sensitive to the surface color and texture, so the feature points can still be clearly distinguished. Figure 15 shows the results of grid-point detection for all the measured objects. These results demonstrate that the proposed multitemplate feature detection algorithm has excellent robustness to the surface color and texture. With the proposed decoding method, the depth information can be acquired. Figure 16

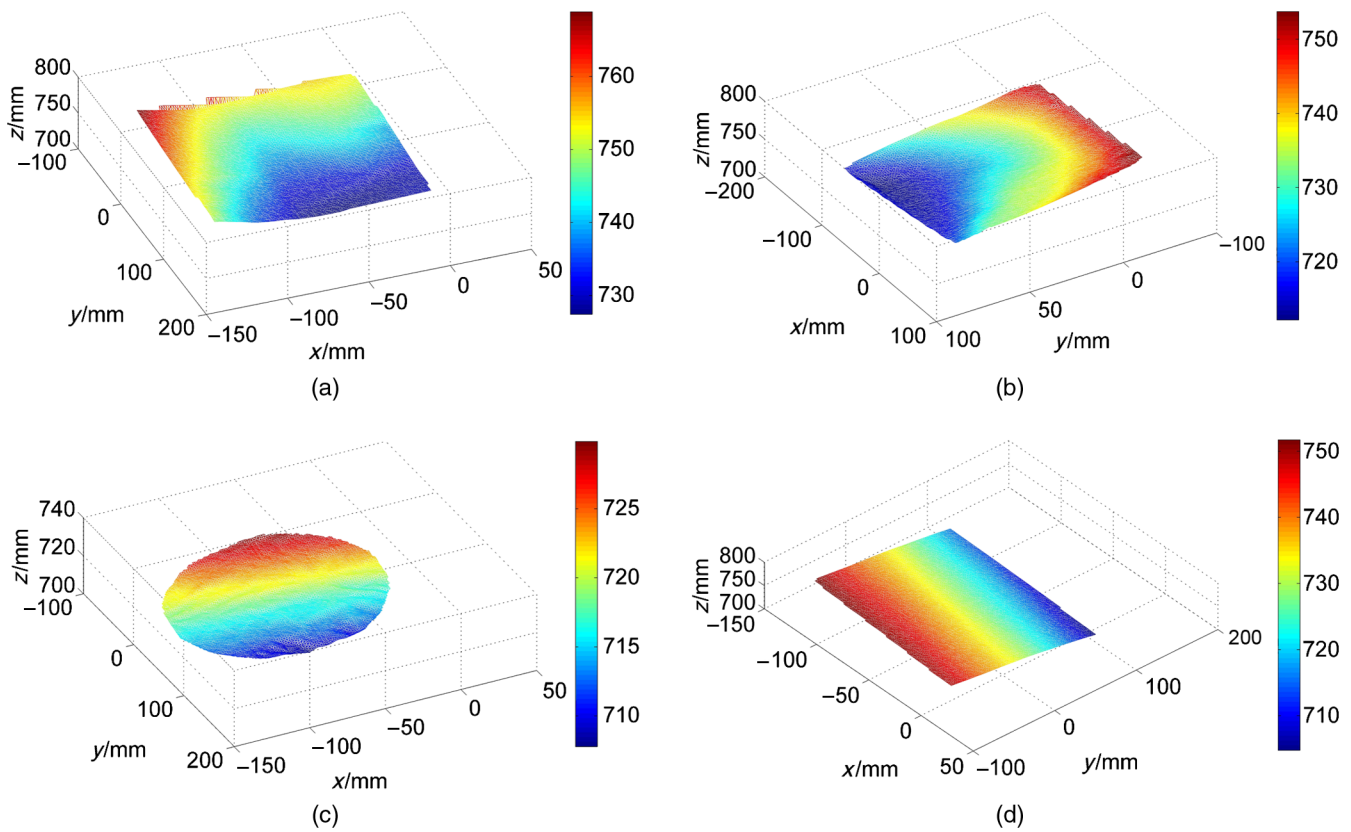

Fig. 17 Results of depth reconstruction for all the measured objects: (a) colorful paper, (b) colorful bag, (c) colorful and textured hat, and (d) textured paper.

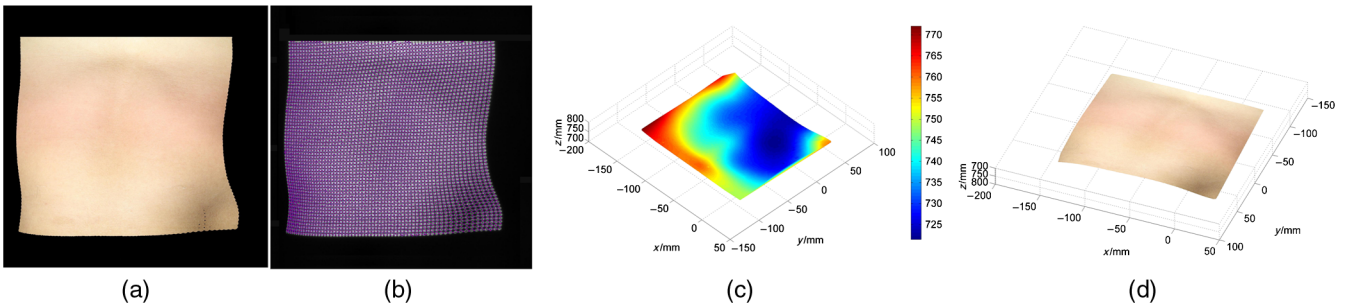

Fig. 18 3-D reconstruction of human chest: (a) the target, (b) result of grid-point detection, (c) result of depth reconstruction, and (d) 3-D model.

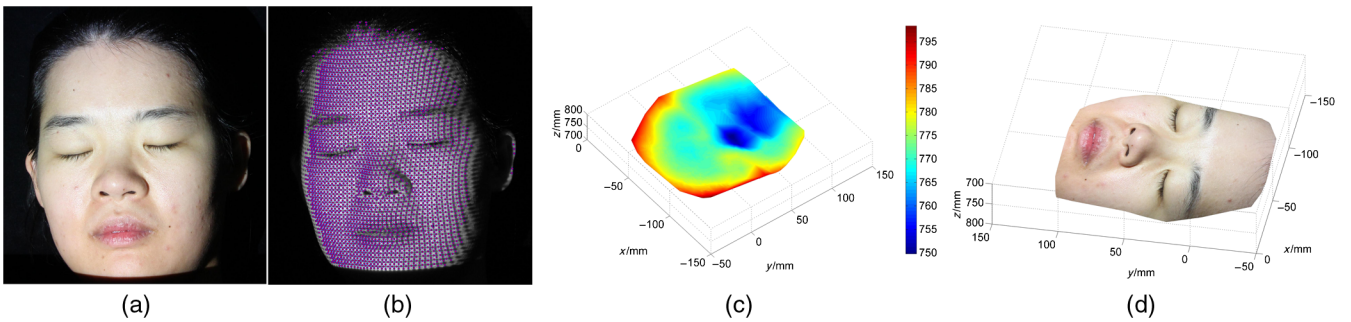

Fig. 19 3-D reconstruction of human face: (a) the target, (b) result of grid-point detection, (c) result of depth reconstruction, and (d) 3-D model.

shows the 3-D point-clouds for all the measured objects. It is clear that the point-clouds in the colorful and textured regions are very complete. It is because the pattern elements in these regions can be correctly decoded. Table 1 displays the measurement results for these four objects. According to the experimental data in this table, it can be estimated that there are about 19 3-D points in the measurement area of 100 mm^2 when the working distance is about 730 mm, and the computation time of grid-point detection and pattern decoding is about 3 s in the Visual Studio 2013 platform without the help of graphics processing unit (GPU) computing. The results of depth reconstruction after using the bilinear interpolation method are shown in Fig. 17. These results demonstrate that our method has great performance in dealing with surface color and texture.

The last experiments are conducted on a real human chest and face, as shown in Figs. 18(a) and 19(a), respectively. Figures 18(b) and 19(b) show the results of grid-point detection for these two targets. It is evident that the result of grid-point detection is great for the human chest, while it is difficult to detect the grid-points in the eyebrows, nose, and mouth areas for the human face. It is reasonable because the reflectivity in the eyebrow areas is too low and the curvature in the nose and mouth areas is too high. By applying the proposed decoding method, most of the pattern elements can be correctly recognized for these two targets when four grid-points around them could be accurately extracted. However, it is hard to correctly identify some pattern elements in the special regions. For example, in the eyebrows areas, the pattern elements are totally fused with the dark eyebrows. In the nose and mouth areas, there exist some special phenomena, such as sharp changes and surface discontinuities. These phenomena usually make the coding window broken. After using the bilinear interpolation method, the complete depth reconstruction can be achieved as shown in Figs. 18(c) and 19(c); thus, the 3-D model of the chest and face can be obtained as shown in Figs. 18(d) and 19(d), respectively.

7 Conclusions

Encoding and decoding are two major concerns involved in a spatial coding structured light system. This paper presents a robust binary coding scheme and a deep decoding method for single-shot shape acquisition. First, the binary rhombic features are chosen as the pattern elements to make the projected pattern robust to surface color and texture, and eight binary geometrical shapes are designed as the coding elements inserting into the white rhombic shapes to generate the projected pattern with a coding window size of 2×2. Second, a multitemplate-based feature detection method is developed for the extraction of the grid-points in the captured image. Based on the extracted grid-points, a topological network is established to separate the geometrical pattern elements from the structured light image. In the decoding stage, a training dataset that contains more than 300,000 samples is first constructed. Then, the deep neural network is applied for the classification of pattern elements. Finally, to refine the decoding results, an error correction algorithm is introduced based on the epipolar and neighboring constraints.

The adoption of a binary pattern element makes the method more robust to surface colors. The use of a deep neural network makes the decoding stage more accurate to surface distortion and image blurring. Extensive experiments were conducted to evaluate the proposed method from the aspects of classification accuracy, measurement precision,

and reconstruction quality. Future work will focus on how to apply the proposed method to the industrial applications with the help of GPU computing and high-speed cameras, for example, the 3-D inspection of fast moving or changing surfaces, such as the rotating blades, high-frequency vibrating films, and so on.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Nos. 61375041 and 51575332), the Shenzhen Science Plan (JCY20140509174140685, JCY20150401150223645, and JSGG20141020103440413), and Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology.

References

1. F. Chen, G. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," *Opt. Eng.* **39**(1), 8–22 (2000).
2. F. Blais, "Review of 20 years of range sensor development," *J. Electron. Imaging* **13**(1), 231–240 (2004).
3. S. Tang, X. Zhang, and D. Tu, "Fuzzy decoding in color-coded structured light," *Opt. Eng.* **53**(10), 104104 (2014).
4. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognit.* **37**(4), 827–849 (2004).
5. J. Salvi, J. Batlle, and E. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," *Pattern Recognit. Lett.* **19**(11), 1055–1065 (1998).
6. M. Williams, F. Jessie, and N. Sloane, "Pseudo-random sequences and arrays," *Proc. IEEE* **64**(12), 1715–1729 (1976).
7. P. Fechteler and P. Eisert, "Adaptive color classification for structured light system," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 1–7 (2008).
8. X. Zhang, Y. Li, and L. Zhu, "Discontinuity-preserving decoding of one-shot shape acquisition using regularized color," *Opt. Lasers Eng.* **50**, 1416–1422 (2012).
9. X. Zhang, L. Zhu, and Y. Li, "Color code identification in coded structured light," *Appl. Opt.* **51**(22), 5340–5356 (2012).
10. L. Zhang, B. Curless, and S. Seitz, "Rapid shape acquisition using color structured light and multi-pass dynamic programming," in *Proc. of the IEEE Computer Society First Int. Symp. on 3D Data Processing Visualization and Transmission*, pp. 24–36 (2002).
11. X. Zhang and L. Zhu, "Determination of edge correspondence using color codes for one-shot shape acquisition," *Opt. Lasers Eng.* **49**(1), 97–103 (2011).
12. X. Zhang, L. Zhu, and Y. Li, "Indirect decoding edges for one-shot shape acquisition," *J. Opt. Soc. Am. A* **28**(4), 651–661 (2011).
13. J. Salvi, J. Batlle, and E. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," *Pattern Recognit. Lett.* **19**(11), 1055–1065 (1998).
14. R. Morano et al., "Structured light using pseudorandom codes," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(3), 322–327 (1998).
15. A. Adan et al., "3D feature tracking using a dynamic structured light system," in *Proc. of the 2nd Canadian Conf. on Computer and Robot Vision*, pp. 168–175 (2005).
16. Z. Song and R. Chung, "Grid point extraction and coding for structured light system," *Opt. Eng.* **50**(9), 093602 (2011).
17. 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).
18. S. Chen, Y. Li, and J. Zhang, "Vision processing for real time 3-D data acquisition based on coded structured light," *IEEE Trans. Image Process* **17**, 167–176 (2008).
19. C. Albitar, P. Graebling, and C. Doignon, "Robust structured light coding for 3D reconstruction," in *Proc. of the IEEE 11th Int. Conf. on Computer Vision*, pp. 1–6 (2007).
20. X. Jia et al. "Model and error analysis for coded structured light measurement system," *Opt. Eng.* **49**(12), 123603 (2010).
21. M. Reiss and A. Tommaselli, "A low-cost 3D reconstruction system using a single-shot projection of a pattern matrix," *Photogramm. Rec.* **26**(133), 91–110 (2011).
22. X. Maurice, P. Graebling, and C. Doignon, "Epipolar based structured light pattern design for 3-d reconstruction of moving surfaces," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pp. 5301–5308 (2011).
23. X. Maurice, P. Graebling, and C. Doignon, "A pattern framework driven by the Hamming distance for structured light-based reconstruction with a single image," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2497–2504 (2011).
24. J. Xu et al., "Real-time 3D shape measurement system based on single structure light pattern," in *Proc. of the IEEE Int. Conf. on Robotics and Automation*, pp. 121–126 (2010).
25. M. Fang et al., "One-shot monochromatic symbol pattern for 3D reconstruction using perfect submap coding," *Optik* **126**(23), 3771–3780 (2015).
26. K. Boyer and A. Kak, "Color-encoded structured light for rapid active ranging," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-9**, 14–28 (1987).
27. F. MacWilliams and N. Sloane, "Pseudo-random sequences and arrays," *Proc. IEEE* **64**(12), 1715–1729 (1976).
28. M. Brown, R. Szeliski, and S. Winder, "Multi image matching using multi-scale oriented patches," in *Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR 2005)*, pp. 510–517 (2005).
29. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, pp. 131–134, Professional Technical Reference, Prentice Hall, Upper Saddle River, New Jersey (2002).
30. L. Yann et al., "Gradient-based learning applied to document recognition," *Proc. IEEE* **86**(11), 2278–2324 (1998).
31. A. Ulusoy, F. Calakli, and G. Taubin, "Robust one-shot 3D scanning using loopy belief propagation," in *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 15–22 (2010).
32. Z. Song and R. Chung, "Use of LCD panel for calibrating structured-light-based range sensing system," *IEEE Trans. Instrum. Meas.* **57**(11), 2623–2630 (2008).
33. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).

Suming Tang is a research assistant at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences (CAS). He received his bachelor's degree from Guizhou University in 2008, master's degree from Southwest Petroleum University in 2012, and his PhD from Shanghai University in 2015. His current research interests include computer vision and artificial intelligence.

Xu Zhang is an associate professor at Shanghai University. He received his BEng (with honors) degree from Northeastern University in 2005 and his PhD from Shanghai Jiao Tong University in 2011. His current research interests include range sensing and computer vision.

Zhan Song is a professor at Shenzhen Institutes of Advanced Technology, CAS. He received his PhD in mechanical and automation engineering from the Chinese University of Hong Kong, Hong Kong, in 2008. He is currently with Shenzhen Institutes of Advanced Technology, CAS, as an assistant researcher. His current research interests include structured light-based sensing, image processing, 3-D face recognition, and human–computer interaction.

Hualie Jiang is a master student at University of Chinese Academy of Sciences. He received his bachelor's degree from University of Electronic Science and Technology of China in 2014. His current research interests include computer vision and human-computer interaction.

Lei Nie is a PhD student at University of Chinese Academy of Sciences. He received his bachelor's degree from Xi'an Jiaotong University in 2008, master's degree from Beihang University in 2011. His current research interests include computer vision and machine learning.