Proceedings of The 2017 IEEE International
Conference on Real-time Computing and Robotics
July 14-18, 2017, Okinawa, Japan

# A Robust Structured Light Pattern Decoding Method for Single-Shot 3D Reconstruction*

Lifang Song, Suming Tang, and Zhan Song, *Member, IEEE*

*Abstract*— **Traditional pattern element identification methods in binary shape-coded structured light are usually lack of robustness to the surface colors and textures. This paper introduces a novel pattern decoding method for a binary structured light pattern, which is composed of eight geometrical elements. The pattern elements are designed as grid shape and the intersection of grid lines is defined as the feature point. By extracting the grid-points firstly, a topological network is constructed to separate each pattern element from the image. Then, pattern element identification is modeled as a supervised classification problem. The convolutional neural network (CNN) is applied to classify the pattern elements. The network is trained with a mass of pattern element samples with various blur and distortion. The experimental results show that the proposed pattern element identification method has strong robustness to surface color, texture, distortion and image noise.**

***Index Terms*** **- Structured light; feature identification; convolutional neural network; 3D reconstruction.**

## I. Introduction

Dynamic 3D reconstruction is an important research topic in computer vision domain, which has wide applications in the fields of industry inspection, reverse engineering, medical diagnosis, virtual reality etc. Existing techniques for dynamic 3D reconstruction can be generally classified into following categories: 1) laser speckle-based approach, like the well-known Kinect sensor; 2) time of flight (TOF) technique, which is applied as LIDAR in the pilotless navigation; 3) stereo vision, which extracts corresponding features in the scenario and solves 3D depth information via triangulation; 4) structured light-based approach, which project single or multiple patterns on the surface. The first two techniques can provide fast 3D reconstruction speed, but the measurement accuracy is low. The third technique demands plentiful features of the target surfaces, and that makes the 3D reconstruction result instable. In comparison, structured light-based approaches can provide both high measurement accuracy and dense 3D information [1]. However, the structured light-based methods are quite sensitive to the surface color and texture, which decreases the 3D reconstruction quality.

There are two usual coding strategies for existing structured light methods, i.e. the temporal coding and the spatial coding [2]. Temporal coding schemes can generate dense reconstruction, but requires multiple patterns, which is only suitable for static scene. In contrast, spatial coding schemes can acquire 3D information with only single pattern projection and thus suitable for dynamic scene and targets. The principle of spatial coding structured light method is to uniquely encode each predefined pattern feature via its surrounding pattern elements. By decoding spatial structure embedded in the illumination, the matching problem between the projected pattern and captured image can be solved. With system calibration parameters, 3D reconstruction can be implemented via traditional triangulation means [3].

In this paper, a binary spatially coded structured light method is investigated. The projected pattern is generated by pseudorandom array with the coding widow size of 2×2. The code-words are represented as grid-shapes with embedded geometrical symbols. Intersection of each two orthogonal grid-lines is defined as the feature point. 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 symmetry-based feature detector is introduced. In the decoding stage, a training dataset which contains a large number of pattern elements with blur and distortions is first established. And then a deep neural network is trained for the pattern element recognition. By the proposed structured light pattern and decoding method, surfaces with texture and colors can be robustly reconstructed.

## II. Related Works

Coding and decoding are the two major issues in spatial structured light systems. There have been a lot of coding methods to generate the projected pattern, and the pseudorandom sequence and pseudorandom array is most widely used [4, 5]. The code words in pseudorandom sequence or array are usually represented by color codes or shape codes. In comparison, binary shape coding is less sensitive to surface colors than the color coding methods.

Different coding schemes usually come up with different decoding algorithms. Ablitar et al. [6] designed three different geometrical shapes, i.e. disc, circle and dash, to generate the projected pattern based on M-array with the size of 27×29, and the coding window size was 3×3. In the decoding stage, the contour of the pattern element was detected firstly. Then, the position of the feature point was extracted by estimating the centroid of the pattern element. The pattern elements were identified based on the number of concentric circles and the distance from the contour to the center. However, only simple objects could be

*This work was supported in part by the National Natural Science Foundation of China (no. 61375041, U1613213), Shenzhen Science Plan (JCY20140509174140685, JSGG20150925164740726).

Lifang Song is with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China and Wuyi University, Jiangmen, China (lf.song@siat.ac.cn).

Suming Tang and Zhan Song are with Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. Shenzhen, China (tel: 86-755-86392100, fax: 86-755-86392299, e-mail: {sm.tang; zhan.song}@siat.ac.cn).

978-1-5386-2035-9/17/$31.00 © 2017 IEEE

reconstructed with this approach. Yang et al. [7] used three geometrical shapes (i.e. triangle, star, diamond) to represent the pattern primitives. Size of the generated pseudo-random array was 33×44 with a window size of 3×3. The symbols were identified according to the number of edges, the area of each class, the proportion of symbol area to its bounding box and symmetry. To refine the decoding result, the captured structured light pattern was projected back to the projector plane to eliminate projective distortion. Xu et al. [8] proposed a pattern based on pseudorandom sequence with length of 63. The primitives of the pattern were represented by the corner of the chessboard and encoded by the orientation of the corner. In the decoding stage, the moment and principal axes method was adopted to recognize the primitives. The decoding speed was increased by using the epipolar constrain. However, the system can only provide very sparse reconstruction. Maurice et al. [9] designed a pattern with the cuneiform features based on epipolar geometry and perfect submap with a size of 100×150. The pattern gained 15000 points for reconstruction since each geometrical shape had only one feature point. The pattern elements were identified based on their contours and orientations. The usage of epipolar geometry reduced the searching space. This method offered efficient and fast correction of mislabeled features due to blurring, spectral harmful effects and surfaces discontinuities prior to the 3D reconstruction of real scenes.

Jia et al. [10] presented a pattern based on M-array with ten special symbols, which had many turning points and the intersections. In the decoding stage, the horns and intersections of primitives were defined as feature points. Symbols were recognized by the angle variation between the pixel and its direct conjunct neighbors, the location of maximal and minimum angle variation, as well as the number of maximal and minimum angle variation. However, the symbols were difficult to recognize in complex scenarios. Fang et al. [11] used some special symbols as the primitives based on perfect submaps. Each symbol was constructed by several small rectangles. The feature points were defined as the cross points of each symbol. Symbol density spectrum (SDS) and symbols were applied to extract features of symbols in the decoding stage. The shape reconstruction experiment showed that this method owned high resolution and robustness. Minh et al. [12] adopted random illumination pattern with fiducial markers embedded into the random pattern to provide a sparse set of correspondences. Both the motion flow of the textured regions and the illumination flow were used for encoding. For the decoding stage, dense match between camera and projector was obtained by utilizing a greedy correspondence growing algorithm. However, the drawback of this case was that the method cannot deal with textured regions at occluding boundary.

With above review, binary shape coding is more robust to surface color than color coding. The decoding algorithms are dependent to the coding schemes. Most researchers utilized geometrical property of the symbol to identify its category in the decoding stage, but related methods are usually lack of robustness to surface texture and distortions. In this paper, a binary geometrical structured light pattern is introduces, and a deep neural network is designed for the robust pattern decoding.

## III. Pattern Design and Feature Detection

The geometrical pattern is usually based on the pseudo-random array with unique window property. The code-words in the array are represented by geometrical primitives. Following the coding scheme in [13], a pseudorandom array of size 65×63 can be obtained with the window size of 2×2. As shown in Fig. 1, eight geometrical primitives are designed to embed into the grids with black background. Intersections of any two orthogonal grid-lines, namely grid-points, are defined as the feature points. The code-word combination of four geometrical primitives around one grid-point is viewed as its code word.

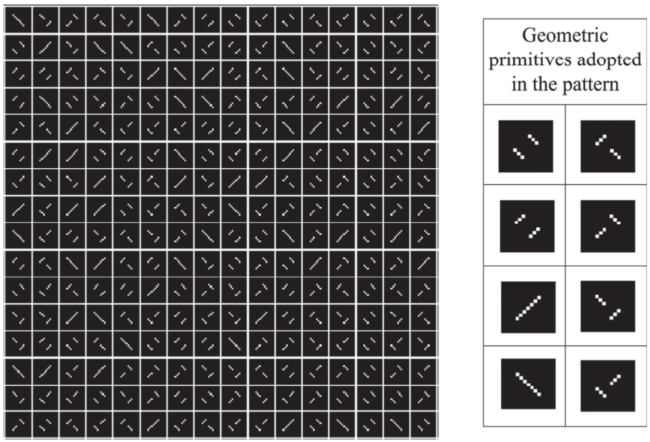


Fig. 1. The constructed geometrical patterns with 8 elements.

The binary geometrical pattern is projected onto the target by the projector, and the camera records an image with pattern illumination. To detect the feature point, the captured image is converted into gray level format firstly. Then, a cross mask is applied to locate candidate grid-points by computing the value of $H$ in neighborhood via:

$$\begin{aligned} H = &\sum_{\alpha=-w}^{w}\sum_{\beta=-w}^{w} I(i+\alpha, j+\beta) - \sum_{\gamma=0}^{l}\sum_{\eta=0}^{l}\{I(i-w+\gamma, j-w+\eta) \\ &+I(i-w+\gamma, j+w-\eta)+I(i+w-\gamma, j-w+\eta) \\ &+I(i+w-\gamma, j+w-\eta)\} \end{aligned} \tag{1}$$

where $I$ denotes the image intensity, $w$ is the radius of cross template and $l$ is set as $w/3$. $H$ indicates the sum of image intensity accumulations along the $i$ and $j$ directions. With an empirical threshold value of $H$, the candidate grid-points with self-centered property can be detected around the true grid-point positions. With an observation to the original pattern image, we can see that perfect rotation symmetry with 90° always occur at the true grid-point positions. Thus, most of the false grid-points among the candidate grid-points can be removed with this local geometrical property [14].

## IV. Pattern Decoding via CNN

Due to surface colors, textures, specular reflection, surface discontinuities of the target objects, the geometrical shapes in the acquired image are usually blurred and distorted as shown by Fig. 2. Thus, it is difficult to identify the degraded pattern elements exactly based on conventional image segmentation methods. The pattern classification task in our work can be regarded as a supervised classification problem, and convolutional neural network [15, 16] has demonstrated excellent performance in deal with such a problem. Therefore a CNN-based pattern decoding approach is investigated in this work.

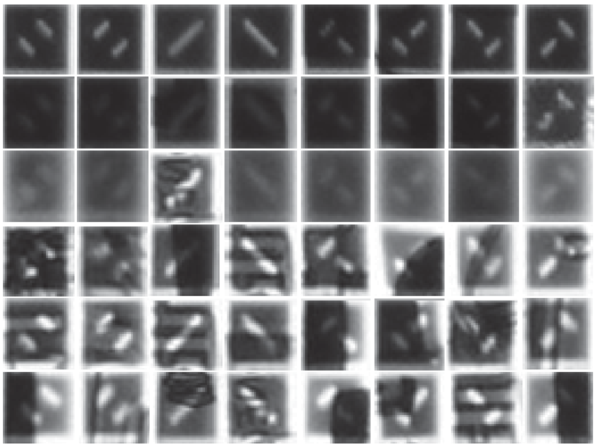

Fig. 2. The projected pattern elements are usually blurred and distorted subject to the surface shape and texture.

After detecting the grid-points, the topological network of all the grid-points can be established. Then, the pattern element can be obtained. The surface patch covered by one pattern element can be approximately regarded as a planar patch based on the assumption that small area of object surface is comparatively smooth. As image coordinates of the grid-points have been extracted, the distorted pattern elements can be transformed into a normalized image size utilizing its four-adjacent grid-points:

$$\begin{bmatrix} x_p \\ x_p \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \\ \mu \end{bmatrix} \qquad (2)$$

where $(x_p, y_p)$ denotes the detected grid-points, and $(x_n, y_n)$ represents the four normalized image corner points $(0,0)$, $(0,h)$, $(w,h)$, $(w,0)$. Given four pairs of points, the projective transformation can be exactly solved. Then the distorted pattern elements can be transformed to normalized images with bilinear interpolation.

To enhance the robustness of the decoding method, it is necessary to collect sufficient pattern element samples with various blur, discontinuity and distortion for the training of convolutional neural networks. The procedure of building the dataset is described as follows.

1) *Select the experimental objects with different characteristics, such as white box, colorful toy, light-colored cover, oil painting, textured paper, plaster model, human hand and face, etc.*
2) *Collect about 8000 training samples by projecting eight geometrical pattern elements onto the experimental targets separately.*
3) *Expand the dataset to more than 300,000 samples by following operations: a) add Gaussian noise to high contrast samples; b) add random white/black lines to the samples to simulate the occlusion problem; c) conduct small affine transformation to simulate small localization error of the grid-points; d) blur the samples with Gaussian filter to obtain blurred samples.*

With above operations, a dataset including sufficient pattern elements with blur, discontinuity or distortion can be established.

Lenet-5 is a multi-layer feed-forward neural network with a deep supervised learning architecture. It can ensure shift and distortion invariance to some extent by combining three architectural ideas: local receptive field, weight sharing technique and spatial or temporal subsampling. In fact, it has been successfully applied to deal with handwritten digit recognition tasks. Since the pattern identification problem is quite similar with handwritten digit recognition problem, the Lenet-5 is also adopted to classify the pattern elements in this work. Fig. 3 shows the structure of Lenet-5 adopted in this paper, The Lenet-5 network comprises seven layers except the input layer: three convolutional layers (C1- 6 maps with size of 28×28, C3-16 maps with 10×10, C5-120 maps with 1×1), two subsampling layers (S2- 6 maps with 14×14, S4-16 maps with 5×5), one fully connected layer (F6 with 84 neuron units) and the output layer. The output layer is a Euclidean RBF layer of 10 units, but for our work, we take the output layer of 8 units.

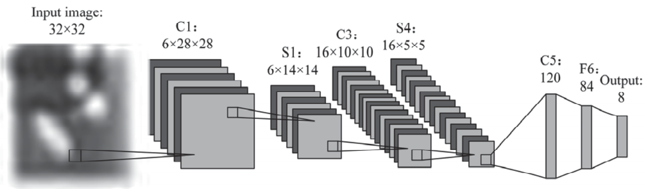


Fig. 3. Structure of Lenet-5 with 8 units output, which is used for the pattern element recognition.

Each unit of the convolutional layer can be computed from its input layers by the following equation:

$$x_j^l = f\left(\left(\sum_{i \in M_j} x_i^{l-1} w_{ij}^l\right) + b_j^l\right) \qquad (3)$$

where $x$ is the unit, $l$ denotes the layer, $w$ represents convolution kernels between two layers, $M_j$ is the number of input layers, $b$ denotes the bias and $f$ is a sigmoid function. The sigmoid function adopted in Lenet-5 is described as a scaled hyperbolic tangent:

$$f(x) = A\tanh(Sx) \qquad (4)$$

where $A$ is the amplitude of the function and $S$ determines its slope at the origin. The value of $f$ ranges from $-A$ to $+A$. The subsampling layers are generated by:

$$x_j^l = f\left(\beta_j^l down\left(x_j^{l-1}\right) + b_j^l\right) \qquad (5)$$

where $down(\cdot)$ is the downsampling function, $\beta$ is the weight and $b$ denotes the bias. With the proposed CNN decoding method, the extracted and normalized pattern elements can be recognized efficiently and accurately.

## V. Experimental Results and Discussion

In the experiments, the structured light system is composed with a digital camera with resolution of 5184×3456 pixels and a DLP projector with resolution of 1920×1080 pixels, which is calibrated with the method as described in [17]. Working distance of the system is about

800 mm, the measuring area is about 250×150 mm. In the projected pattern, the size of each grid element is 16×16 pixels. To test the accuracy, robustness and feasibility of the proposed pattern decoding method, four objects were used in the experiments, as shown in Fig. 4. Grid-point detection results of these targets are as shown by Fig. 5.

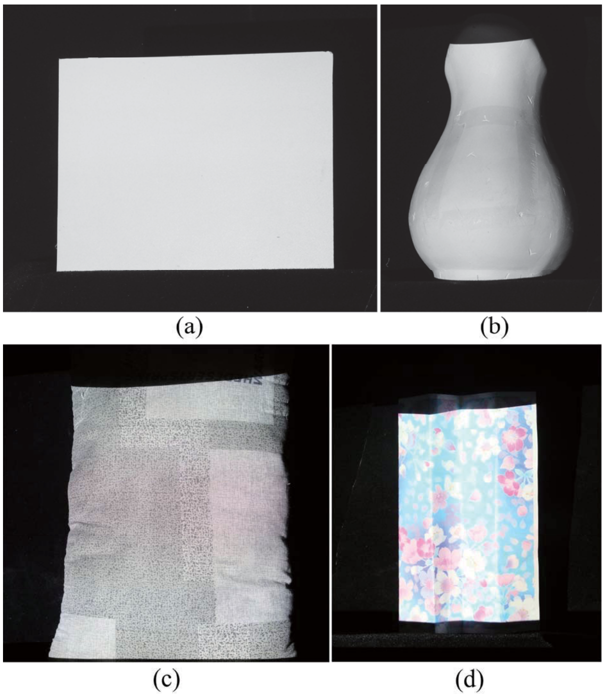

Fig. 4. Four targets with different shape, color and textures are used for the experiment.

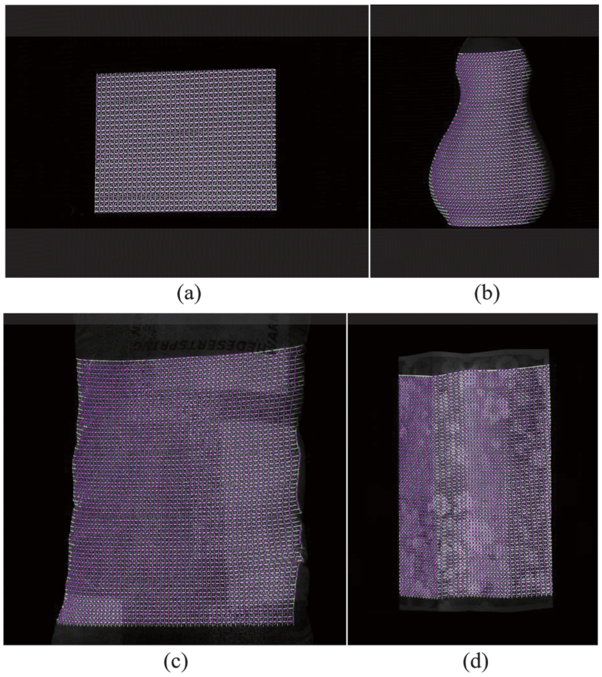

Fig. 5. Grid-point detections results of experimental targets.

To evaluate the accuracy of the proposed pattern decoding method, the training datasets are divided into 10 folds and the leave-one-out method is used. Besides, stochastic Gradient Descent is adopted for 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 experimental result shows that classification accuracy of Lenet-5 on the augmented training database can reach 98.07%.

To test the robustness of the decoding method, extra zero-mean Gaussian noise with different levels are added to the captured images for all the measured objects as shown by Fig. 6. The multilayer perceptron (MLP) method is also applied for the element recognition for comparison. The experimental results are as shown by Fig. 7. From the results we can see that, the identification accuracy on different classification methods for all the objects decreases with the increasing of Gaussian noise level. And the identification accuracy with Lenet-5 is higher than that by MLP. These experiments demonstrate that the proposed pattern identification method has strong robustness and Lenet-5 owns higher classifying power than MLP.

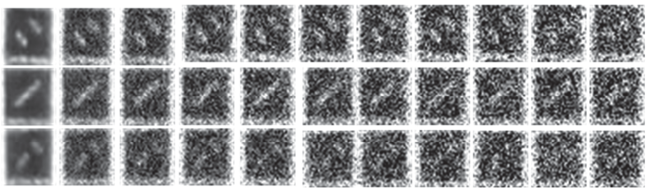

Fig. 6. Image of pattern elements with various Gaussian noise levels.

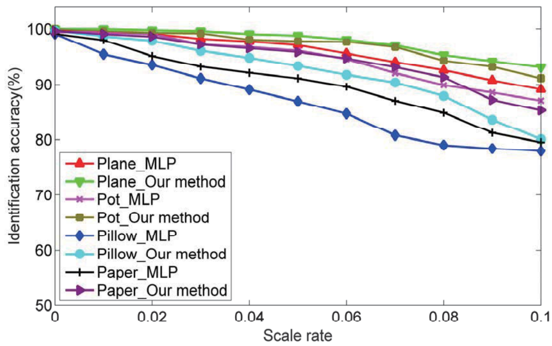


Fig. 7. Comparison of pattern element identification accuracy with respect to the variance of Gaussian noise.

After classifying the pattern elements, the code word of grid-points can be determined. Then, with the local matching method, the correspondences between the projected pattern and captured image can be found. To evaluate the quality of 3D reconstructions, the reconstructed point clouds are plot as shown in Fig. 8. By observation, most of the grid-points can be successfully decoded and reconstructed with high position accuracy. Based on the planar target, a planar fitting is applied for the measurement accuracy evaluation. The absolute mean fitting error is about 0.13 mm. Because the point clouds are still sparse, a bilinear interpolation method is applied for the sparse point clouds. The reconstructed surface models are as shown by Fig. 9. By observation, the targets with plentiful texture and colors can be well reconstructed by the proposed binary structured light pattern and decoding method.

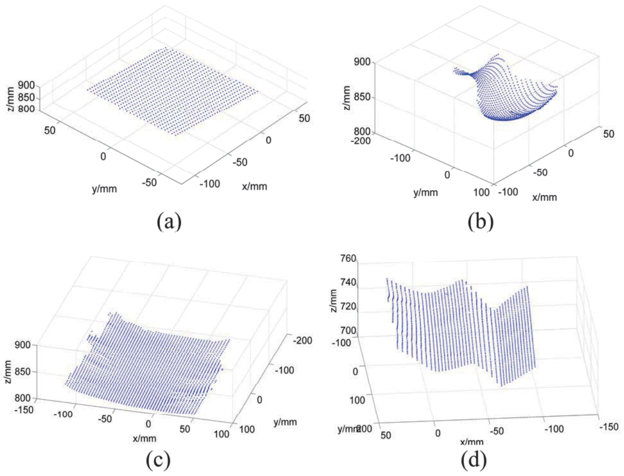


Fig. 8. The reconstructed 3D point clouds of all the experimental objects.

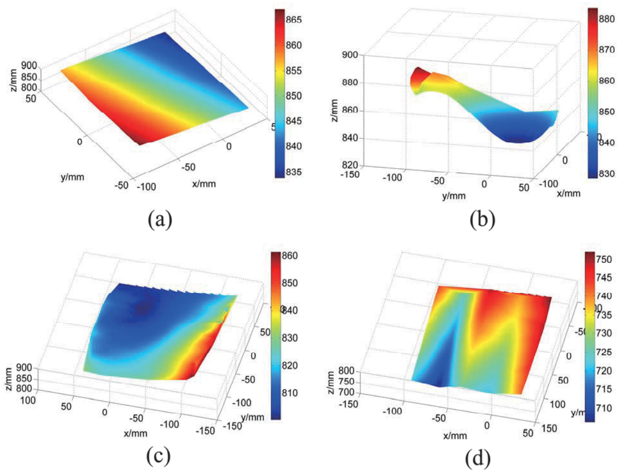


Fig. 9. The reconstructed 3D models of all the experimental targets.

## VI. Conclusion and Future Work

This paper presents a novel robust pattern identification method for binary shape-coded structured light pattern. The proposed pattern is designed as grid shape with eight embedded geometrical shapes and the intersection of each two orthogonal grid-lines is defined as the pattern feature point. By introducing a symmetry-based feature detector, a topological network is established to separate the geometrical pattern elements from the structured light image. To identify the separated pattern elements accurately, a deep learning-based pattern element identification method is studied. To make the identification algorithm more robustly and accurately, an extensive training set is established firstly. The training samples are collected from a variety of target surfaces under various lighting conditions. The adoption of deep neural network makes the pattern element identification more accurately and robust to surface distortion and blurring. Extensive experiments were conducted to evaluate the proposed method from the aspects of identification accuracy and reconstruction quality. Future work can address how to extract the feature points precisely, and how to correct the false decoded pattern elements.

## References


[1] J. Salvi, J. Batlle, and E. M. Mouaddib, "A robust-coded pattern projection for dynamic 3D scene measurement," *Pattern Recognition Letters*, vol.19, no.11, pp.1055-1065, 1998.

[2] A. O. Ulusoy, F. Calakli, and G. Taubin, "One-shot scanning using De Bruijn spaced grids," *in Proc. of 12th IEEE International Conference on Computer Vision Workshops,* pp.1786-1792, 2009.

[3] M.Williams, F. Jessie, and N. J. A. Sloane, "Pseudo-random sequences and arrays," *Proceedings of the IEEE*, vol.64, no.12, pp.1715-1729, 1976.

[4] R. A. Morano, C. Ozturk, and R. Conn, "Structured light using pseudorandom codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, no.3, pp.322-327, 1998.

[5] Z. Song and R. Chung, "Determining Both Surface Position and Orientation in Structured-Light-Based Sensing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.32, no.10, pp.1770-1780, 2010.

[6] C. Albitar, P. Graebling, and C. Doignon, "Design of a monochromatic pattern for a robust structured light coding," *in Proc. of IEEE International Conference on Image Processing*, vol.6, pp.VI-529-VI-532, 2007.

[7] Y. Lei, K. Bengtson, L. Li, and J. Allebach, "Design and decoding of an m_array pattern for low_cost structured light 3D reconstruction systems," *in Proc. of the 20th IEEE International Conference on Image Processing*, pp.2168-2172, 2013.

[8] J. Xu, "Real-time 3D shape measurement system based on single structure light pattern," *in Proc. of the IEEE International Conference on Robotics and Automation*, pp. 121-126, 2010.

[9] 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 International Conference on Robotics and Automation*, pp. 5301-5308, 2011.

[10] X. Jia, G. Yue, and M. Fang, "The mathematical model and applications of coded structured light system for object detecting," *Journal of computers*, vol.4, no.1, pp.53-60, 2009.

[11] M. Fang, W. Shen, D. Zeng, X. Jia, and Z. Zhang, "One-shot monochromatic symbol pattern for 3D reconstruction using perfect submap coding," *Optik-International Journal for Light and Electron Optics*, vol.126, pp. 3771-3780, 2015.

[12] S.G. Narasimhan, and Y. Sheikh, "Texture Illumination Separation for Single-Shot Structured Light Reconstruction," *in Proc. of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.38, no.2, pp.390-404, 2016.

[13] Z. Song and R. Chung, "Grid point extraction exploiting point symmetry in a pseudo-random color pattern," *in Proc. Of 15th IEEE International Conference on Image Processing*, pp. 1956-1959, 2008.

[14] H. Lin and Z. Song, "A twofold symmetry based approach for the feature detection of pseudorandom color pattern," *in Proc. of the 4th IEEE International Conference on International Science and Technology*, pp.623-626, 2014.

[15] Y. Le, L. Bottou, Y. Bengio, and P. Ha□ner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol.86, no.11, pp.278-2324, 1998.

[16] Y. Le, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in Neural Information Processing Systems*, vol. 2, pp. 396-404, 1997.

[17] Z. Song, and R. Chung, "Use of LCD panel for calibrating structured-light-based range sensing system," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no.11, pp. 2623-2630, 2008.