Optics and Lasers in Engineering 96 (2017) 50–62

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Robust pattern decoding in shape-coded structured light

Suming Tang[a,b], Xu Zhang[b], Zhan Song[a,c,*], Lifang Song[a], Hai Zeng[a]

[a] Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Guangdong, China
[b] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China
[c] The Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Structured light
Binary geometrical coding
Robust decoding
3-D reconstruction

ABSTRACT

Decoding is a challenging and complex problem in a coded structured light system. In this paper, a robust pattern decoding method is proposed for the shape-coded structured light in which the pattern is designed as grid shape with embedded geometrical shapes. In our decoding method, advancements are made at three steps. First, a multi-template feature detection algorithm is introduced to detect the feature point which is the intersection of each two orthogonal grid-lines. Second, pattern element identification is modelled as a supervised classification problem and the deep neural network technique is applied for the accurate classification of pattern elements. Before that, a training dataset is established, which contains a mass of pattern elements with various blurring and distortions. Third, an error correction mechanism based on epipolar constraint, coplanarity constraint and topological constraint is presented to reduce the false matches. In the experiments, several complex objects including human hand are chosen to test the accuracy and robustness of the proposed method. The experimental results show that our decoding method not only has high decoding accuracy, but also owns strong robustness to surface color and complex textures.

1. Introduction

Structured light used for 3-D shape acquisition provides several advantages including non-contact, flexibility, efficiency and high measurement precision [1,2]. A basic structured light system can be composed of one projector and one camera. The projector is used to project single or multiple patterns with coding information onto the target object, and the camera is used to capture the scene with pattern illuminations. By extracting the coding information from the captured images, the correspondence between the camera and projector can be determined. Finally, with system calibration parameters, 3-D reconstruction can be implemented via traditional triangulation means [3].

There are two usual coding structure strategies for existing structured light methods, i.e. the temporal coding and the spatial coding [4]. The temporal coding methods are usually used for static scenes or targets, since it needs multiple projections. In comparison, the spatial coding methods can be used for 3-D reconstruction of dynamic scenes since it only demands a single pattern projection. In spatial coding schemes, the label of each codeword can be determined by its neighborhood in the pattern image. Generally, spatial coding methods can only provide 3-D reconstruction results with relative low resolution and accuracy. For the coding schemes, pseudorandom sequence, De Bruijn sequences [5], pseudorandom array and M-array [6] are the four most adopted. The expressions of codeword can be luminance codes [7], color codes [8–15], and shape codes [16–29]. In comparison, the color code and the shape code are adopted more widely than the luminance code. Besides, shape coding is less sensitive to surface color compared with color coding.

The primitives in shape-coded pattern were usually coded by different binary geometrical shapes. In [16], a binary shape-coded pattern composed with eight geometrical shapes was introduced. The coding window size was designed as 2×2 based on a 63×65 M-array. The pattern element is designed as rhombic with embedded geometrical shapes. The pattern feature point was defined as the intersection of two adjacent rhombic shapes. And a neural network was trained for the pattern element identification. Albitar et al. [17,18] 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. There are three steps in the decoding method. First, the contour of the pattern elements was detected. Then, the feature point was extracted by estimating the centroid of the pattern element. Finally, the pattern elements were identified based on the number of the concentric circles and the distance from the contour to the center. Lei et al. [19] increased three geometrical shapes, i.e. triangle, star, diamond, to represent the pattern primitives. The size of the generated pseudo-random array is 33×44 with a window size of

* Correspondence to: Shenzhen Institutes of Advance Technology, Chinese Academy of Sciences, Shenzhen 518055, China.
E-mail address: zhan.song@siat.ac.cn (Z. Song).

http://dx.doi.org/10.1016/j.optlaseng.2017.04.008
Received 5 November 2016; Received in revised form 17 March 2017; Accepted 18 April 2017
Available online 03 May 2017
0143-8166/

3×3. The geometrical shapes were identified according to the number of edges, the area of each class, the distance from the centroid to the edge, the proportion of the area to its circle, and the proportion of the area to its bounding box and symmetry. To refine the decoding results, the captured structured light pattern was projected back to the projector plane to eliminate the projective distortion. Reiss et al. [20] improved the pattern resolution using five pattern primitives, each of which contained four or six points for shape reconstruction, the coding window size is 3×3. In the pattern decoding stage, the pattern elements were segmented from the background with thresholding approach and each pattern element was classified according to its class local threshold. After that, the differences between the background and the pattern elements were amplified through image enhancement processes and the similar local histograms over the image area covered by the projected points were generated. Then, the pattern elements in the enhanced image were segmented with the region-growing method and the perfect image templates representing each primitive were used to assign a label to the segmented pattern elements according to their classes. The pattern elements were matched against the five stored templates with a cross-correlation function. After this classification decision, the neighborhood relationships between segments were determined and the pattern codifications were used to assign a unique label to each pattern element. Besides, the non-recognized points were identified based on epipolar constraint and neighborhood constraint. Finally, the matches were found with local matching method. Nguyen et al. [21] adopted four symmetrical shapes in Ref. 19 for the generation of structured light pattern, and the unique code-word of each shape is defined as itself and its four neighbor shapes. To obtain the correspondence, the captured image was converted to a binary one with adaptive threshold method, and the connected components were labeled based on graph traversal method in graph theory. Then, all the connected components were recognized and classified into four separate symmetrical shapes by dividing each connected component into four quarters and calculating the average binary value of pixels in each quarter. Thereby the corresponding points between the pattern elements on the captured image and those on the projected pattern were determined based on the assumption of local spatial coherence. Xu et al. [22,23] proposed a pattern based on a pseudorandom sequences with a length of 63, the window width is 4. The pattern primitives were represented by the corner of the chessboard and encoded by the orientation of the corner. Since the pattern primitives were designed on the black background and separated enough, there was a strong gradient in the image intensity around the primitive boundary. Hence, the contour of the pattern element was detected to estimate the position of feature point. Then, the moment and principal axes method was adopted to recognize the pattern elements. The speed of decoding increased much as the pattern generated in one dimension along the epipolar line. Jia et al. [24] presented four geometrical shapes to generate the projected pattern based on M-array. These geometrical shapes were formed by two intersecting triangles, the vertices of two triangles at the primitive image center were taken as the feature points. Before decoding, some necessary preprocessings were performed on the image, like denoising and top hat transform. Then, the canny edge detector was adopted to detect the edge of the pattern elements, and the feature point was extracted based on the approximate center-of-mass coordinate. According to the angle between the centerline of the shape image and the basic shaft of the images, the pattern elements were identified. Thereby the matches between the captured image and projection pattern were found by periodically comparing the code values of these two images with a pseudorandom array template. Maurice et al. [25,26] designed a pattern with the cuneiform features based on epipolar geometry and perfect submap with a size of 100×150. The pattern gained 15,000 points for 3-D reconstruction because each geometrical shape had only one key point. The pattern elements were identified based on their contour and orientation. The use of epipolar geometry reduced the search space more. This method offered efficient and fast correction of mislabeled features due to the blurring, spectral harmful effects and surface discontinuities prior to the 3-D reconstruction of real scenes. Jia et al. [27,28] presented a pattern based on M-array with ten special geometrical shapes, which had many turning points and intersections, the horns and intersections of the primitives were defined as key points. In the decoding stage, some preprocessings, such as filtering, threshold segment and thinning, were firstly conducted to make the pattern elements distinguishable. After that, the traversal method was used to detect the pattern elements. Then, the pattern elements 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. Based on the identified pattern elements, the window matching method was performed to find the correspondence between the projected pattern and captured image. Fang et al. [29] proposed a symbol density spectrum to choose the pattern primitives for improving resolution and decreasing decoding error. The proposed symbol density spectrum provided a distribution of the feature points for 3-D reconstruction after which ten pattern primitives were extracted. Then, a comparative analysis of the pattern primitives and scene testing of the pattern primitive damage rate were conducted to choose nine pattern primitives from one group to form a density pattern. The feature points were defined as the cross points of each pattern primitive. To correctly identify the pattern elements, smoothing filtering and threshold segment were firstly performed on the image. And then, the connected components were extracted with 8-adjacent connecting area labeling algorithm. Consequently, the pattern elements were classified by recognizing the feature of the connected components. Finally, the local matching method based on window voting was implemented to find the correspondence.

The above shape-coded structured light methods [17–29] usually tend to define the geometrical shapes themselves or neighboring image points as the feature points. While the patterns are projected on the targets with plentiful color or complex textures, not only the decoding accuracy but also the feature localization precision will be greatly decreased. Moreover, the decoding methods of most shape-coded structured light systems are usually performed via simple image segmentation and template matching algorithms. But for real applications, such simple decoding methods are always impracticable because the pattern elements are indistinct and their label cannot be definitely judged when they changes drastically due to some complex factors, such as surface color, textures, distortions, reflections, discontinuities, and so on.

In this paper, a novel structured light pattern is presented, which is designed as grid-shape with embedded geometrical shapes. Based on the projected pattern, a robust decoding method is proposed to obtain the correspondence. Instead of using conventional image segmentation method, a multi-template feature detection method is introduced to detect the feature point that is the intersection of each two orthogonal grid-lines. Pattern element identification is modelled as a supervised classification problem. The deep neural network technique is applied to accurately classify the pattern elements, which is achieved by establishing a training dataset containing the pattern elements with various blurring and distortions. After matching, an error correction mechanism is proposed to refine the match result. The rest of this paper is organized as follows. Section 2 describes the framework of the proposed method. Section 3 presents the proposed pattern decoding method. The experimental results are given and discussed in Section 4. Conclusions are offered in the last section.

2. The framework of the proposed method

Our proposed one-shot shape acquisition method includes five steps, namely encoding, image capture, decoding, calibration, and triangulation, as shown in Fig. 1.

In traditional shape coding scheme, as shown in Fig. 2, the coding