Journal of Visual Communication and Image Representation 55 (2018) 229–242

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Depth sensing with coding-free pattern based on topological constraint☆

Guangming Shi, Ruodai Li, Fu Li*, Yi Niu, Lili Yang

Key Laboratory of Intelligent Perception and Image Understanding (Chinese Ministry of Education), School of Artificial Intelligence, Xidian University, No. 2 South Taibai Road, Xi'an 710071, China

ARTICLE INFO

Keywords:
Coding-free
Depth sensing
Structured light
Topological constraint
Dynamic programming

ABSTRACT

Structured light depth sensing with a single-shot pattern is widely employed to capture depth maps for dynamic scenes. For conventional structured light techniques, the projected pattern has to be coded delicately in regards to color, shape, and intensity, in order to assign each pixel with a unique label. However, using such a complicated pattern is a double-edged sword, as although it is effective in labelling pixels, it is also sensitive to environmental noise such as: ambient illumination, textures, uneven albedos, or colors of objects in a scene. In contrast, a coding-free pattern is simply constructed and also insensitive to various environmental noise. Therefore, the coding-free pattern method is capable of robustly sensing the depth for complex scenes. The main challenge in coding-free depth sensing is the 'correspondence retrieval' between the projected and captured pattern (i.e. matching pixels between the projected and captured pattern). In this study, we focused on evaluating the correspondence retrieval in a coding-free binary grid pattern. A graph based topological labelling (GBTL) algorithm is proposed to determine the topological coordinates of the intersections of the grid. Then we retrieved the correspondence by using the topology of the grid and the epipolar constraint. We also demonstrated the upper bounds of depth variance by employing the proposed method. The proposed technique alleviates many of the limitations faced with traditional correspondence retrieval. Experimental results showed that the proposed technique performed better (i.e. in terms of precision) than the popular RGB-D cameras Kinect v1 and Kinect v2. Compared with the traditional single-shot techniques, which require complicated patterns, the proposed technique significantly improved the robustness and ease of work, while achieving comparable precision. Additionally, this proposed technique could also be used for both the binary coding-free and the traditional chromatic grid patterns.

1. Introduction

Depth sensing techniques are widely used in many fields, such as robot vision, medical sciences, human-computer interaction, and entertainment, etc. Many depth sensing techniques such as time-of-flight (ToF), stereo vision, and structured light illumination (SLI) have been developed, but compared with the other techniques, SLI method, however, has many advantages in terms of higher accuracy and lower cost.

A general framework for the SLI method involves projecting single or multiple well-designed patterns onto a target scene and capturing the reflection of the projected patterns. The captured patterns may be twisted according to the depth variations of objects in the scene. The depth maps are acquired in three steps. The first step involves extracting the features from the captured images. The second step involves matching the corresponding locations of the detected features with the projected features. The final depth maps are restored based on the retrieved features via triangulation. The first two steps, which are known as feature detection and correspondence retrieval, play important roles in SLI depth sensing. The accuracies of feature detection and correspondence retrieval will directly affect the depth sensing performance.

According to the coding strategies of the projected pattern [1], SLI techniques can be roughly categorized into two classes: temporal-based and spatial-based schemes. The former scheme encodes positional information for a projector's pixels by multiple projected patterns in a time series. These techniques, such as binary code method [2] and phase shifting method [3–7], have advantages in terms of high accuracy and robustness, however, multi-shots have to be conducted to capture the pattern series, so the temporal-based techniques are not suitable for dynamic scenes.

The spatial-based scheme only projects one pattern and the depth maps are acquired with just a single-shot, which is favorable for depth sensing in dynamic scenes. For the spatial-based scheme, patterns are

---

☆ This paper has been recommended for acceptance by Wanqing Li.

* Corresponding author.

E-mail address: fuli@mail.xidian.edu.cn (F. Li).

https://doi.org/10.1016/j.jvcir.2018.06.009
Received 21 June 2017; Received in revised form 24 April 2018; Accepted 8 June 2018

usually designed delicately by using various colors, intensities, or shapes to encode pixels with unique codes. Each code manifests as a window of pixels, which is also called a feature. To retrieve the correspondence of features between the projected pattern and the captured image, the appearances of features such as colors, intensities, or shapes must be preserved in the captured image. Unfortunately, the preservation of these appearances of features may not be maintained, as they may be disturbed by environmental noise and depth variations in a scene, which is the main limitation of these spatial-based methods.

Among spatial-based methods, color-encoding scheme is one of the most commonly used strategies. Chromatic blocks, stripes, or other shapes are employed to construct patterns with unique codes such as $M-$array [8–10], pseudorandom [11,12], or De Bruijn codes [13–15]. In general, a color-encoding depth sensing method extracts the colors of pixels in captured images, and then retrieves the correspondence between the image and the pattern according to the color codes. Correctly extracting colors is vital for color-encoding methods. Unfortunately, the color of a pixel in a captured image depends on many other factors in addition to the projected color, such as shading, the albedo of target object, color cross-talk between the projector spectra and sensor filters, and sensor noise [14]. Thus, misclassifications of colors are difficult to avoid in practice, and as such color-encoding methods are usually limited to scenes with neutral colors.

Takeda [16] and Chen [17] proposed the Fourier Transform Profilometry (FTP) method, which obtains depth by using a fringe pattern with sinusoidal intensity. After a filtering in Fourier domain, the fundamental frequency was separated from the other frequency components and the phases that carry the depth information were extracted. However, these two have their limitations, as the filtering process is sensitive to spectrum aliasing, which may occur due to disturbances by uneven illumination and abrupt depth changes in the scenes.

Binary shape based methods are more robust to the effect of ambient light, uneven albedos, and colors in scenes. Albitar [18] constructed a pattern of perfect map by three different shapes: circle, disc, and stripe. Each submap consists of nine adjacent shape units, and every unit is distinguishable from each other due to the unique code of the perfect map. Maurice [19] designed a pattern of cuneiform features to form a sub-perfect map, where the features are arranged along the epipolar lines. Li [20] used unique speckles to encode pixels in the pattern. These methods encode the features by different heights, widths, orientations, or compactness for a unique identity. However, the features are not invariant to perspective distortion. Depth variations may distort the features and lead to incorrect correspondence retrieval [35].

Different from the well-designed SLI patterns, coding-free patterns (CFPs) are usually binary, and with simple and periodic constructions, such as a binary grid. Four major advantages of CFP based methods exist when using the binary grid pattern. First, compared with color-encoding methods, CFP methods are more robust to the disturbances by colors in a scene. Second, compared with intensity-encoding methods, CFP methods are free from spectrum aliasing, thus can perform well under the disturbances by ambient illumination, and uneven albedos in a scene. Third, compared with shape-encoding methods, CFP methods are insensitive to the distortions of shapes due to depth variations. Lastly, a CFP can be projected using a low-cost instrument because of its simple construction. Therefore, it is surprising that CFP methods have not gained a wider appreciation over that of traditional SLI methods.

However, the simple construction of CFPs also brings large challenges in correspondence retrieval process. Lavoie [36] claimed “When the grid is not encoded, the process (correspondence retrieval) becomes nearly impossible”. To date, only a few studies regarding this issue with CFPs have been conducted. In 2010, Song and Chung [34] proposed a coding-free depth sensing method with a ‘chessboard’-like pattern. In this study, they determined a surface orientation profile from the observed grid-lines and then converted the orientation description to an absolute depth description. However, two sacrifices were made in doing so for the conversion. One, they had to employ a weak perspective projection instead of the full perspective projection, which decreased the sensing accuracies, and two, the method could only acquire relative depth maps rather than true depth maps.

Some researchers have studied methods with a grid pattern in recent decades. One of the advantages of using a grid pattern is that the lines can be detected as features with high precision. Salvi [21] proposed a chromatic grid pattern with six colors, where three colors are assigned for horizontal lines, and the other three for vertical lines. The colors of three adjacent lines form a De Bruijn code with order 3. Previous studies [23–25] have used two colors to form a De-Bruijn sequence to achieve more accurate retrieval correspondence. Furukawa [22] proposed a grid pattern with two colors to distinguish the vertical and horizontal lines. Then coplanarity constraint and epipolar constraint are utilized to retrieve correspondence. All of these methods require projecting a chromatic pattern for correspondence retrieval.

In this study, we investigate the correspondence retrieval problem in depth sensing by using a coding-free binary grid pattern, and as such, three contributions are made in this study regarding this topic. One, the horizontal and vertical lines in a captured image cannot be separated by color or intensity because of the binary pattern. We address this issue by using a coarse-to-fine line detection strategy, by which one could suppress ambient light, separate different lines, and locate the accurate positions of the lines. Two, there is no mark on each detected line, thus it is challenging to retrieve correspondence of the lines between projected and captured grid, to resolve this issue we propose a novel *graph based topological labelling* (GBTL) method to extract the topology of the captured and twisted grid. Then the correspondence retrieval task can be accomplished by solving several linear systems of equation defined by epipolar constraint. Finally, for a comprehensive analysis, we investigate the ‘sensitivity’ of the proposed GBTL technique. We are aware that GBTL may fail on rapid varying surfaces, which may lead to line disordering in a captured grid. This study provides a quantitative analysis of the upper bound of depth variance in a scene that can be measured by GBTL.

The remainder of this paper is organized as follows. In Section 2, we introduce the framework of the proposed method. The coarse-to-fine line detection is explained in Section 3. The topology based correspondence retrieval technique, including GBTL method, absolute coordinate determination, and analysis of sensitivity are discussed in Section 4. Experimental results are presented in Section 5. We make a conclusion in Section 6.

2. Framework

A typical SLI system is composed of a camera and a projector, as shown in Fig. 1(a). The coding-free binary grid pattern is shown in Fig. 1(b), which comprises parallel lines with equal intervals in vertical and horizontal directions, respectively. The procedure of the proposed method is shown in Fig. 2.

The following procedures can be split into three major steps. First, the coarse-to-fine line detection is used to extract the horizontal and

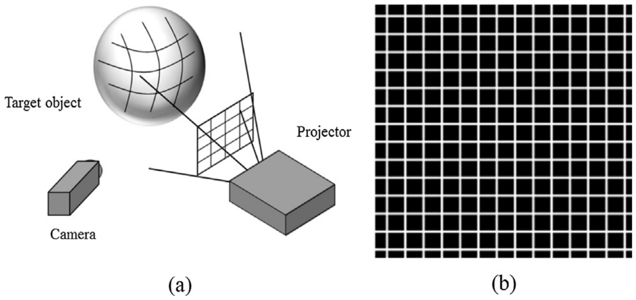

Fig. 1. (a) A typical SLI system. (b) The coding-free binary grid pattern.