J. Vis. Commun. Image R. 25 (2014) 649–658

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

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

Robust depth sensing with adaptive structured light illumination

Yueyi Zhang [a,1], Zhiwei Xiong [b,*], Pengyu Cong [c,1], Feng Wu [b]

[a] University of Science and Technology of China, No. 96, Jinzhai Road, Baohe District, Hefei, Anhui 230026, PR China
[b] Microsoft Research Asia, No. 5 Danling Street, Haidian District, Beijing 100080, PR China
[c] School of Information and Electronics, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, PR China

ARTICLE INFO

Article history:
Available online 13 June 2013

Keywords:
Depth sensing
Depth camera
Structured light
Speckle
Defocus
Overexposure
Underexposure
Adaptive illumination

ABSTRACT

Automatic focus and exposure are the key components in digital cameras nowadays, which jointly play an essential role for capturing a high quality image/video. In this paper, we make an attempt to address these two challenging issues for future depth cameras. Relying on a programmable projector, we establish a structured light system for depth sensing with focus and exposure adaptation. The basic idea is to change current illumination pattern and intensity locally according to the prior depth information. Consequently, multiple object surfaces appearing at different depths in the scene can receive proper illumination respectively. In this way, more flexible and robust depth sensing can be achieved in comparison with fixed illumination, especially at near depth.

1. Introduction

With the big success of Kinect, depth sensing, also known as range imaging and 3D surface measurement, has been attracting more and more academic and industrial attention recently than ever before. Among various depth sensing techniques, structured light is considered a promising approach to be used in future consmer-grade depth cameras, due to its easy implementation and universal applicability. Actually, the core technique behind Kinect is a variation of structured light [1]. In a structured light system, specially designed 2D patterns are projected onto the scene from a lighting engine, and the scene under illumination is captured by an imaging sensor. Depth is then measured by the shift between the projected and captured patterns through triangulation.

According to the pattern design strategy, mainstream structured light techniques can be roughly classified into two categories, time multiplexing and spatial neighborhood. Time multiplexing methods, e.g., binary coding [2], gray coding [3], n-ary coding [4] and phase shifting [5], identify each point in the scene in the temporal domain by projecting and capturing multiple patterns sequentially. On the contrary, spatial neighborhood methods only use one elaborate pattern, in which each point is uniquely coded by its neighborhood, e.g., De Bruijn sequences [6], M-arrays [7], symbol-coded [8] and color-coded [9] patterns. The pseudo-random speckle pattern used in Kinect also belongs to this category [1]. As an essential issue for structured light techniques, pattern codification receives most research efforts in this field. (Please refer to [10] for an extensive review and comparison.) Nevertheless, in practice, there are some cases beyond the capability of pattern codification. For example, suppose there are multiple objects at different depths in the scene, which is often the case in real world. If the illumination is not proper for all object surfaces, defocus blur, overexposure or underexposure may be present in the captured image, regardless of which kind of pattern is used.

In this paper, we explore the way more flexible and robust depth sensing can be achieved with a structured light system. We notice that, in most existing systems including Kinect, the illumination is fixed and not aware of the object location in space. That is to say, without loss of generality, the real depth map of the current scene, denoted as D_t, is estimated as

$$\hat{D}_t = f(P_0, D_t), \quad t = 1, 2, \ldots \tag{1}$$

where P_0 is a fixed pattern (or pattern set) and f models the general depth sensing process in a structured light system.

Intuitively, the optimum illumination for object surfaces at diverse distances to the lighting engine should be different, as shown in Fig. 1. The challenge is how to obtain this prior knowledge in the first place. One key observation is that, the latest known depth information actually provides an approximate partition of the current scene if no abrupt movement occurs. Based on this partition, it is possible to adjust the illumination locally for different object surfaces in the scene. Therefore, we propose a depth sensing framework with adaptive illumination. In comparison with (1), this new process can be modeled as

* Corresponding author.
E-mail addresses: zhyuey@mail.ustc.edu.cn (Y. Zhang), zhxiong@microsoft.com (Z. Xiong), congpengyu@bit.edu.cn (P. Cong), fengwu@microsoft.com (F. Wu).
[1] This work was done during his internship at Microsoft Research Asia.

1047-3203/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jvcir.2013.06.003

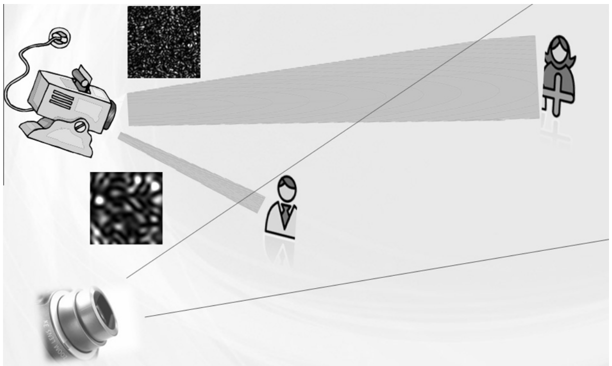

Fig. 1. Different illumination for object surfaces at different depths.

$$\hat{D}_t = f(P_t|\hat{D}_{t-1}, D_t), \quad t > 1 \tag{2}$$

where P_t is the currently projected pattern dependent on the estimate of prior depth map $\hat{D}_{t-1}$. Specifically, this work makes three contributions.

First, we establish a practical depth sensing system with adaptive illumination. In our system, a programmable projector, a CCD camera and a PC are synchronized with one another. Once the scene under current illumination is captured by the camera, the image is transmitted to the PC, where the corresponding depth map is recovered. By analyzing this depth map, the pattern for next illumination is generated on the PC and then loaded to the projector. Once the projector updates the illumination, it triggers another synchronized image acquisition.

Second, we propose a focus adaptation strategy for depth sensing. In a structured light system, the depth of field (DOF) of projector is often small as large aperture is used to produce bright images. Therefore, defocus blur may be present in the captured image when object surfaces appear at different depths. (A laser emitter such as that used in Kinect can provide large DOF, but the DOF of camera will still be limited sometimes.) By analyzing the tolerance to defocus of different speckle patterns, we generate a composite pattern for current illumination according to the prior depth information, which minimizes the defocus influence on depth sensing.

Third, we propose an exposure adaptation strategy for depth sensing. In a structured light system, the illumination intensity varies much in space, especially along the light path. However, the dynamic range of camera is always limited. When object surfaces appear at diverse distances to the lighting engine, overexposure or underexposure may be present in the captured image. Given the prior depth information, we adjust the dynamic range of the projected pattern locally to control the illumination intensity. In this way, the exposure of currently captured image is optimized.

The rest of this paper is organized as follows. The related works are briefly reviewed in Section 2. Section 3 clarifies the principles for a depth sensing system with adaptive illumination. Technical details of focus and exposure adaptation are elaborated in Sections 4 and 5, respectively. Experimental results are presented in Section 6, and Section 7 concludes the paper with some discussion on the future work.

2. Related work

Depth sensing can be performed through various methods, e.g., stereo vision [11], photometric stereo [12], shape from shading [13], depth from defocus [14], coded aperture [15], time-of-flight [16] and structured light [10]. Despite the diversity in implementation and application scenarios, structured light is considered as one of the most competitive techniques to be used in future consumer-grade depth cameras, according to the latest efforts in industry [17,18]. In this paper, we mainly focus on depth sensing with structured light. Still, the proposed adaptive principles could also be beneficial to other methods.

An early work on the programmable imaging system is introduced by Nayer et al. using a digital micro-mirror device (DMD) [19], in which several imaging functions are implemented, including high dynamic range, optical filtering and matching. Later, the operating principle of DMD has been exploited to obtain high speed projection [20], which enables high speed display, sensing and control applications [21]. Time multiplexing structured light is a typical example being speeded up, but it still requires a high speed camera for synchronized capturing [22]. In our system, however, there is no rigid requirement on the camera speed as the structured light we use belongs to the spatial neighborhood category.

Unlike the approach of depth from defocus [23], our depth sensing system is based on structured light, although a defocus model is estimated for analysis. Similar analysis has been conducted by Garcia et al. in [24], where the conclusion is drawn that signals with lower frequency content are more robust to defocus and thus selection of different patterns can increase the virtual DOF. In comparison, our work is not only different but also a step forward. On the one hand, we adopt pseudo-random speckle patterns instead of temporal dithered codes for illumination, and the optimal pattern is selected for each depth range. On the other hand, we implement this focus adaptation strategy in a practical system, and reliable depth sensing results are achieved in the presence of defocus.

In addition, the exposure variation strategy proposed by Nayer et al. in [19] is a kind of tradeoff between temporal/spatial resolution and dynamic range, whereas in our exposure adaptation strategy, the illumination is actively and locally changed. Local exposure adjustment has been investigated by Koninckx et al. in [25], where multiple patterns need to be projected to obtain a crude estimate of the scene geometry. Contrastively, we only use a single pattern composed of pseudo-random speckle, which is updated according to the prior depth information. Since our system aims for dynamic scenes and only requires multiple projections once at the initial time, it is also different from the works addressing global illumination for static scenes [26,27].

Our work shares some similarity with another work by Koninckx et al. [28], in which the current illumination pattern is adapted to the latest depth information. However, there exist two distinct differences between them. On the one hand, the target of pattern adaptation in [28] is to make it more robust against current noise levels on the decoding and to ease the detection. Specifically, increasing the density of coding cues will lower the noise level that can be tolerated. In contrast, pattern adaptation in our work aims to solve the defocus and over/under-exposure issues in case objects at diverse depths are to be measured. On the other hand, the patterns used in [28] consist of stripes and curves with variable widths and angles, while those used in our work consist of random speckles with variable sizes. So the corresponding decoding processes are totally different.

There is another pioneer work by Chien et al. [29], in which an adaptive system for 3D reconstruction is proposed and more accurate depth can be achieved for objects with various colors and details. The relationship between [29] and our work is as follows. First, they share the same observation that depth sensing should be performed more flexibly. Specifically, the illumination patterns should be changed for different target regions. Second, exposure adaptation is investigated in both works, yet in different perspectives. [29] reveals the relationship between illumination and color, while we exploit the relationship between illumination and depth.