Signal Processing: *Image Communication* 41 (2016) 28–39

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Correction and interpolation of depth maps from structured light infrared sensors

Simone Milani*, Giancarlo Calvagno

Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy

A R T I C L E I N F O

Article history:
Received 8 May 2015
Received in revised form
25 November 2015
Accepted 30 November 2015
Available online 11 December 2015

Keywords:
DIBR denoising
MS Kinect
Structured light camera
Infrared sensor
3D acquisition

A B S T R A C T

Infrared structured light sensors are widely employed for control applications, gaming, acquisition of dynamic and static 3D scenes. Recent developments have lead to the availability on the market of low-cost sensors, like Kinect devices, which prove to be extremely sensitive to noise, light conditions, and the geometry of the scene.

The paper presents a quality enhancement strategy for Kinect-acquired depth maps that corrects depth values via a set of local differential equations and interpolates the missing depth samples. The approach improves both density and accuracy of the generated 3D model under different light conditions and in the presence of cross-talk noise derived from other devices. Moreover, the paper introduces a new experimental reference dataset for Kinect denoising algorithms consisting in multiple acquisition under different noise conditions associated to a laser scan model of the scene (ground truth).

1. Introduction

The recent availability of low-cost range cameras has shaken the ICT world leading to a flourishing of new object recognition applications, human–computer interfaces, and acquisition systems of dynamic 3D scenes. Time-of-Flight cameras [1], structured light 3D scanners [2], and multicamera systems allow easy real-time acquisition of dynamic 3D scenes with both static and dynamic elements [3].

Among these, the Xbox Kinect sensor [4], which includes a standard RGB camera together with an infrared (IR) structured light scanner (see Fig. 1), has recently proved to be one of the most widely-used depth sensors thanks to its versatility, the limited cost and the improved performance in a wide range of possible applications.

In [5] Leyvand et al. discuss new possibilities in tracking persons and identifying their identity. Suma et al. present a novel toolkit to model human gestures and poses [6]. In [7] Wilson presents a touch interface implemented using a depth camera. Moreover, several navigation tools employ kinect sensor to accurately control and drive robots in an indoor environment (see Turtlebot by Willow Garage [8], as an example). Other approaches employ Kinect for modeling ground surface in precision agriculture [9], providing a reliable visual feedback to the user in virtual dressing rooms [10], managing interaction in augmented reality [11] or in virtual environments [12].

The structure of the Xbox Kinect device is reported in the diagram of Fig. 1. The implemented IR depth sensor consists in an IR projector, an IR CMOS camera, and a processing unit that controls them and elaborates the acquired signal. An IR pattern of dots is projected by the IR projector on the scene, and the IR CMOS camera acquires the reflected pattern, which will be distorted according to the geometry of the objects. The central processing unit estimates the distance of each point from the depth camera considering the distortions in the acquired dot

* Corresponding author. Tel.: +39 049 827 7641, +39 049 827 7731; fax: +39 049 827 7699.
E-mail addresses: simone.milani@dei.unipd.it (S. Milani), calvagno@dei.unipd.it (G. Calvagno).

http://dx.doi.org/10.1016/j.image.2015.11.008
0923-5965/

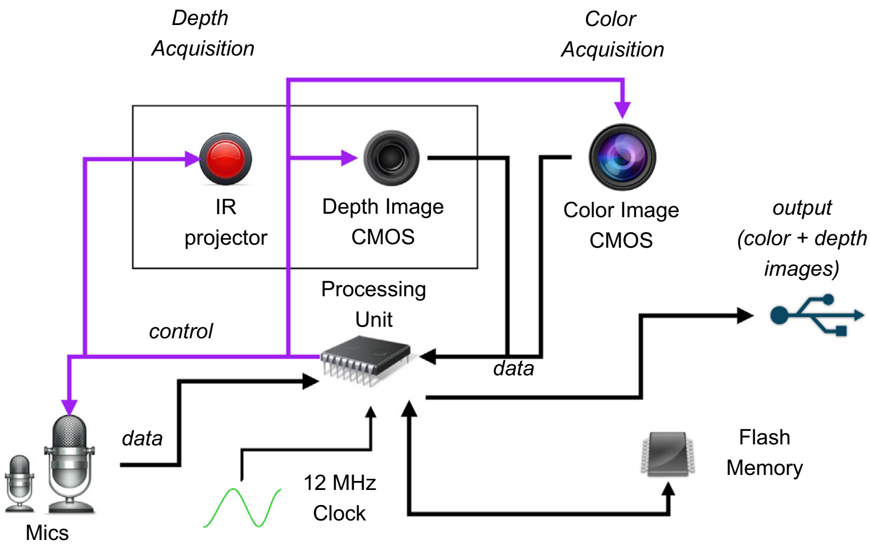

Fig. 1. Block diagram of the MS Kinect sensor.

pattern with respect to the projected one. Color information is available as well since an RGB CMOS camera permits obtaining a standard picture of the acquired scene. This information permits building a point cloud model of the 3D scene by mapping depth pixels into color pixels with a warping operation.

Unfortunately, despite the strong versatility and the wide range of new applications of these devices, the resulting depth signal is affected by a significant amount of noise.

Like many imaging systems, depth sensors present shot noise related to the radiation, A/D conversion quantization noise, and thermal noise. Moreover, in a real scene IR sensors receive a significant amount of radiation that has not been generated by the associated projector. This is the case of ambient illumination (e.g., sun or artificial lights whose electromagnetic radiations also span the IR frequency of MS Kinect) or, in case multiple similar IR structured light cameras are present in the same scene, the dot patterns produced by a different device [13]. In a controlled environment, this inconvenience can be mitigated by time division multiplexing [14], motion (shake-and-sense [15]) or by sample removal and hole filling [16]. In a distributed system, e.g., an environment where multiple Kinect sensors are operated independently, like in the case of multiple Kinect-controlled robots [17,18], this inconvenience must be addressed by effective interpolation and denoising algorithms.

The paper presents a joint denoise-interpolation algorithm for MS Kinect sensor that aims at correcting the computed depth values and interpolate them whenever they are not available because of the noise conditions. The approach relies on an initial denoising performed by matching borders between range and color images. The mismatches between edge information from RGB data and depth maps can be corrected by a set of differential local equations. Then, the resulting data are propagated to neighboring locations where depth samples are not available according to the edge information. The proposed solution improves the approach in [19] by introducing a new set of denoising and interpolating equations for depth samples and optimizing their implementation in order to run in real time. Experimental testing was improved as well using a high-quality laser-scanned 3D model as a ground truth reference in order to evaluate the algorithm performance. This activity led to the creation of a downloadable dataset with multiple acquisitions under different illumination conditions. The reported results show that the algorithm increases the number of available depth samples and refines the accuracy of the generated point clouds. Moreover, the proposed solution can run in real time allowing a frame rate of approximately 17 frame/s. In the following, Section 2 overviews some of the existing works in the literature, while Section 3 presents the characteristics and the effects of noise on depth signals. Section 4 describes the proposed algorithm in detail, with Section 4.3 reporting the depth correction process and Section 4.5 presenting the interpolation strategy. Experimental results (Section 5) and conclusions (Section 6) end the paper.

2. Related works

Several works have proposed novel denoising algorithms to improve the quality of the acquired depth maps. This is severely impaired by three main issues: the inaccuracy of depth samples due to the presence of noise, the irregularity of object boundaries, and the presence of missing depth samples (holes) in the acquired depth maps.

Depth samples are usually corrected by filtering the depth data in order to reduce the noise level. The work in [20] reviews the performance of different types of filters (median, bilateral, joint bilateral, non-local means, with adaptive thresholds, and temporally smoothing) of the data generated by a Kinect sensor. Experimental results show that temporal denoising performs extremely well in most of the situations although it requires multiple