Optik 127 (2016) 791–794

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Integral imaging display for natural scene based on KinectFusion

Jianlei Zhang, Xiaorui Wang*, Qiping Zhang, Yujiao Chen, Junhui Du, Yiqun Liu

School of Physics and Optoelectronic Engineering, Xidian University, Xi'an 710071, Shaanxi, China

ARTICLE INFO

Article history:
Received 27 January 2015
Accepted 27 October 2015

Keywords:
Integral imaging
Fusion
Three-dimensional image acquisition

ABSTRACT

In this paper, we propose an integral imaging display system based on KinectFusion, using only a moving low-cost depth camera as the pickup device. The multi-frame depth data of the observed scene streamed from a Kinect sensor is fused into a single global surface model represented by the volumetric truncated signed distance function. Thus, the inherent noise associated with single frame depth data can be eliminated. The elemental image array for display is obtained by ray casting the volumetric truncated signed distance function. To match the pickup part and the display part, the relationship between the pickup voxel and the display voxel is deduced using the ray optics theory. Experimental results show the validity of the proposed method.

1. Introduction

A considerable amount of research has been dedicated to the exploration and development of three dimensional (3D) imaging and display techniques, since two dimensional (2D) images can only provide limited information about a 3D scene due to the lack of depth cue. Among them, integral imaging (II) is considered to be one of the most prominent 3D display techniques, which is proposed by Lippmann [1]. It has attracted many attentions due to its various advantages such as quasi-continuous, full-color viewpoints within a viewing angle, full parallax and working with incoherent light [2–4].

An II system typically consists of two parts: a pickup part and a display part. In the pickup part, a lens array coupled with an image sensor is always used to capture elemental image array [5]. However, for natural scene far from the camera, the angular extension of the lens array is very small. The depth lens plus lens array referred to as far-field integral imaging is proposed to capture the far-field scene, which involves the same optical principles as the plenoptic camera [6–8]. A key benefit of the plenoptic camera is to allow post-capture refocusing [9]. However, the parallax captured by this architecture is still small, thus the captured elemental image array does not match the requirement for wide-viewing II display. Camera array is an alternative pickup method for natural scene, but the cost of the system and complex calibration process limit its extensive use.

Depth imaging technology has advanced dramatically over the last few years, finally reaching a consumer price point with the launch of Kinect [10]. A pickup method for II display using a depth camera along with color camera is proposed for real objects [11]. However, this approach creates only an incomplete mesh, from only a single viewpoint. Single frame depth measurement from a single viewpoint often fluctuates and the depth map contains numerous holes, where no reading can be obtained. In addition, one challenge for II is to realize wide-viewing angle display, thus the occlusion relation in different perspective must be known, so the depth map from a single point is not enough.

KinectFusion has been proposed for real-time 3D reconstruction [12,13]. In this paper, we apply KinectFusion technique in II as a pickup method for integral imaging display. By using only a moving low-cost Kinect depth and color camera, the multiframe depth data is fused into a single global surface model. The incomplete depth map from a single frame can be improved. By the proposed method, the wide viewing perspective for the natural scene can be captured, which can satisfy the requirement for wide-viewing integral imaging display. The matching relationship between the pickup voxel and the display voxel is deduced using the ray optics theory. Experimental results demonstrates the validity of the proposed method.

2. The proposed method

Conventional two-dimensional display emits light isotropically. In contrast, a light field display supports the control of individual rays of light, modulating radiance as a function of position and direction across its surface. II display system belongs to the light field display, which can be represented by the plenoptic function

* Corresponding author. Tel.: +86 02988201026.
E-mail address: xrwang@mail.xidian.edu.cn (X. Wang).

http://dx.doi.org/10.1016/j.ijleo.2015.10.168
0030-4026/

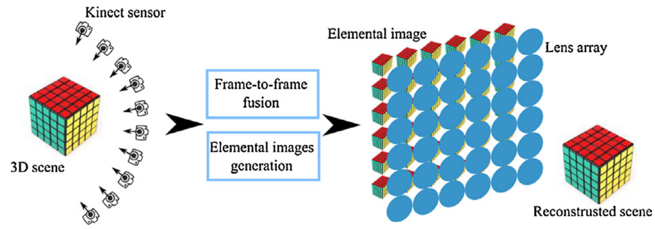

Fig. 1. Concept of the proposed method.

[14]. Lens array coupled with image sensor, camera array and the depth lens coupled with elemental lens array can capture the light field of the 3D scene for II display. However, high-resolution light field recording usually requires an excessively large bandwidth.

In this paper, a single freely moving Kinect sensor consisting of a depth camera and a color camera is used as the pickup device to build dense 3D models for large indoor environments. Fig. 1 shows the concept of the proposed method. The hand-held Kinect sensor is moved around the 3D scene to obtain multiframe depth data. Then, the multiframe depth data streamed from the Kinect sensor is fused into a single global surface model. Thus, the depth data of a single frame can be improved and the occlusion relation for wide-viewing angle is available.

KinectFusion continually tracks the 6 degrees of freedom pose of the camera and fuses live depth data from the camera into a single global 3D model. Each consecutive depth frame, with an associated live camera pose estimate, is fused incrementally into one single 3D reconstruction using the volumetric truncated signed distance function (TSDF) [15]. By ray casting the TSDF, the elemental image array for display can be obtained.

In conventional II pickup system, the viewing angle is limited by:

$$\theta = 2\arctan\left(\frac{p}{2f}\right) \tag{1}$$

where p stands for the pitch of lens array, f is the focal length of lens array. The proposed method adds the viewing angle of the II system by fusing the field of view of the multi-frame data, thus the viewing angle of the II display is no longer limited by the field of view of the pickup system.

In the input stage, the parameters of the lens array for display and the display panel should be stored in the program. According to the display parameters, the optimum pickup parameters are determined in order to match the pickup part and the display part. The lateral size of the voxel for depth camera is

$$D_c^v = \frac{p_c \times l_p}{g_c} \tag{2}$$

where p_c is the size of the pixel for depth camera, g_c the image distance for depth camera, and l_p the depth of the scene. One problem encountered with the II display for the sake of 3D display applications is the pseudoscopic nature of the 3D reconstructed image. Thus the depth of the captured scene must be converted before the generation of elemental image array [16]. In addition, the axial depth of the scene is always changed during the display stage in order to show the reconstructed scene at desired depth. Moreover, the captured scene is room scale while the II display system is only the size of a monitor. Hence, a homogeneous scaling factor exists between the captured scene and the reconstructed scene, i.e. α. So the relationship between the depth of the captured scene and that of the reconstructed scene is

$$l_d = l_o - \alpha \times l_p \tag{3}$$

Now, suppose that the size of the pixel for display panel is p_{display}, the pitch of the elemental lens is p. When the LA is positioned as $l = 0$, the lateral size of the reconstructed voxel for II at l_d becomes:

$$D_d^v = \frac{p_d \times l_d}{f} \tag{4}$$

In order to match the voxel size of the pick system and that of the display system, Eq. (5) should be satisfied:

$$\alpha D_c^v = D_d^v \tag{5}$$

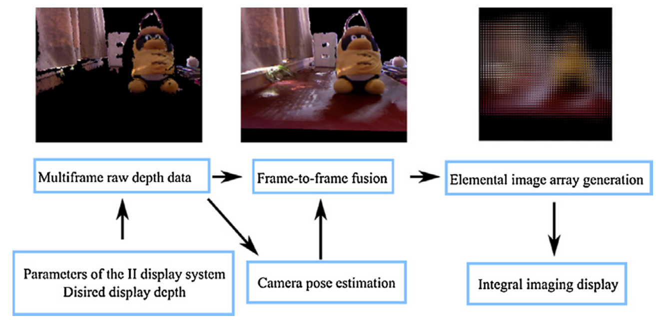

Fig. 2. Architecture of elemental images generation using the proposed method.

We now describe the components that make up the II display system based on KinectFusion. Fig. 2 provides an overview of our whole method in block form. It is comprised of the following five components:

Input of the display parameters: The parameters of the lens array and the display panel should be first inputted in the program. Thus, the viewing resolution and the viewing angle are determined by the parameters of the architecture. In addition, the desired display depth is controlled by the observer. According to the viewing parameters, the voxel size of the pickup system and the moving range of the Kinect sensor can be calculated using Eqs. (1) and (5).

Raw depth data measurement: The raw depth data of the 3D scene is extracted, which includes inherent noise. A bilateral filter is applied to the raw depth data to obtain a discontinuity preserved depth map with reduced noise [17]. In addition, a dense vertex map and normal map pyramid are generated from the raw depth measurements obtained from the Kinect sensor.

Sensor pose estimation: Live camera pose estimation involves estimating the current camera pose for each new frame depth image. In this phase, a rigid 6DOF transform is computed to closely align the current oriented points with the previous frame, using the Iterative Closest Point (ICP) algorithm [18].

Frame-to-frame fusion: Given the pose determined by tracking the depth data from a new sensor frame, the depth data is integrated into the scene model maintained with a volumetric, represented by TSDF with the surface interface defined by the zero-crossing where the values change sign.

Elemental images array generation: A virtual camera array is set to record the elemental image array (EIA), which is used in the display stage. This is performed by ray casting the TSDF into the estimated frame to provide a dense surface prediction. Before the raycasting for rendering, the volumetric representation TSDF is scaled by the homogeneous scaling factor i.e. α. In order to avoid the distortion of the reconstructed 3D scene, the viewing angle, resolution, number and arrangement of the virtual camera is the same to those of the display part.

3. Experiment and analysis

Preliminary experiments have been performed to confirm the validity of the proposed method. The prototype of integral imaging display based on KinectFusion consists of a Kinect sensor, a computer, a LCD monitor and a lens array. The Kinect sensor gives a