Computerized Medical Imaging and Graphics 40 (2015) 147–159

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Real-time computer-generated integral imaging and 3D image calibration for augmented reality surgical navigation

Junchen Wang [a,*], Hideyuki Suenaga [b], Hongen Liao [c,**], Kazuto Hoshi [b], Liangjing Yang [a], Etsuko Kobayashi [a], Ichiro Sakuma [a]

[a] Graduate School of Engineering, The University of Tokyo, Tokyo, Japan
[b] Department of Oral-Maxillofacial Surgery, Dentistry and Orthodontics, The University of Tokyo Hospital, Tokyo, Japan
[c] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China

ARTICLE INFO

Article history:
Received 22 April 2014
Received in revised form 3 September 2014
Accepted 3 November 2014

Keywords:
Integral imaging
3D image calibration
Augmented reality
Surgical navigation
Stereo tracking

ABSTRACT

Autostereoscopic 3D image overlay for augmented reality (AR) based surgical navigation has been studied and reported many times. For the purpose of surgical overlay, the 3D image is expected to have the same geometric shape as the original organ, and can be transformed to a specified location for image overlay. However, how to generate a 3D image with high geometric fidelity and quantitative evaluation of 3D image's geometric accuracy have not been addressed. This paper proposes a graphics processing unit (GPU) based computer-generated integral imaging pipeline for real-time autostereoscopic 3D display, and an automatic closed-loop 3D image calibration paradigm for displaying undistorted 3D images. Based on the proposed methods, a novel AR device for 3D image surgical overlay is presented, which mainly consists of a 3D display, an AR window, a stereo camera for 3D measurement, and a workstation for information processing. The evaluation on the 3D image rendering performance with 2560 × 1600 elemental image resolution shows the rendering speeds of 50–60 frames per second (fps) for surface models, and 5–8 fps for large medical volumes. The evaluation of the undistorted 3D image after the calibration yields sub-millimeter geometric accuracy. A phantom experiment simulating oral and maxillofacial surgery was also performed to evaluate the proposed AR overlay device in terms of the image registration accuracy, 3D image overlay accuracy, and the visual effects of the overlay. The experimental results show satisfactory image registration and image overlay accuracy, and confirm the system usability.

1. Introduction

1.1. Background

Augmented reality (AR) based surgical navigation allows virtual organs overlaid on real organs to provide surgeons with an immersive visualized surgical environment. Compared with virtual reality based surgical navigation where a flat 2D monitor is used to display a virtual surgical scene, AR navigation where a virtual surgical scene is further registered to reality could provide enhanced realism and more intuitive information for surgical guidance [1]. There are three types of AR visualization technologies in the current surgical navigation systems: video-based display [2–6], see-through display [7,8], and projection based AR [9–11]. Video-based display superimposes virtual organs on a (stereo) video stream captured by endoscopic cameras or head-mounted displays (HMD). Because camera videos cannot reproduce all the information obtained by the human visual system, see-through display and projection based AR were proposed for overlay on user's direct view using translucent mirrors or projectors. The virtual organs used in most direct view overlay systems are 2D-projected computer graphics (CG) models of organs derived from preoperative medical images. Compared with visual perception of a 3D object, 2D projection lacks two important visual clues that give a viewer the perception of depth: stereo parallax and motion parallax [12]. Depth perception in image-guided surgery enhances the safety of the surgical operation. Therefore, for those applications which superimpose virtual organs directly on a real surgical site, a 3D image is preferred so that consistent and correct parallax is maintained when observed from different locations. Depth perception can be obtained from the parallax encoded in the 3D image to give surgeons a distance sensing about surgical targets. "3D image" in this paper refers to the image encoding the parallax information which can be extracted by some optic device or human eyes to give a viewer the depth perception: the third dimension.

* Corresponding author. Tel.: +81 0358417480.
** Corresponding author.
E-mail addresses: wangjunchen@gmail.com (J. Wang), liao@tsinghua.edu.cn (H. Liao).

http://dx.doi.org/10.1016/j.compmedimag.2014.11.003
0895-6111/

Among 3D image display technologies are stereoscopy [13], integral imaging (or integral photography) [14], and holography [15]. Stereoscopy creates depth perception using two view images, like the left and right images seen by the human visual system. The disparities between the two images encode the parallax information which can be extracted using such as a parallax barrier or polarized glasses for display. However, stereoscopy has very limited data bandwidth (only two images). The parallax information is only provided at two predefined view points. That is why we cannot see more even if we shift our eyes in a 3D film cinema. On the opposite side, holography directly records the wavefront of the light from a 3D object. Since all optical data irradiated from the object can be captured and reproduced with minimal loss, the parallax information can be encoded in a complete way. However, the data bandwidth is too huge to be handled in real time using current available devices with satisfactory resolution and viewing angle. In addition, holography requires very complicated and expensive devices to capture and reproduce 3D images, which limits its application. Integral imaging exists between the stereoscopy and the holography. It has medium data bandwidth and provides stereo parallax and full motion parallax within its viewing angle. The devices for extracting the encoded parallax information are very simple: It requires only a high resolution liquid crystal display (LCD) and a lens array in front of the LCD to display 3D images. The resulting 3D image presents stereo parallax and continuous motion parallax which can be directly observed by viewers without wearing special glasses. These exciting advantages make integral imaging under active research. Although integral imaging was first invented by Lippmann more than 100 years ago [16], it began to draw great attentions in the last decade since high resolution digital cameras and LCDs became available. A thorough review on integral imaging can be found in [17] and [18]. Current study of integral imaging focuses on 3D image pickup methods [19], image reconstruction methods [20,21], and viewing quality enhancement [22,23].

1.2. Related work and limitations

Our group first introduced integral imaging into surgical navigation for AR visualization [24–26], where an autostereoscopic 3D image is displayed by a lens array monitor (3D display) using computer-generated integral imaging (CGII), and is overlaid on the surgical site by a translucent mirror (AR window). The spatial registration between the 3D image and the surgical site is performed manually using a point-based registration method involving the use of a commercial optical tracking device. In our recent publication [27], an AR solution for dental surgery using 3D image overlay was proposed, where a stereo camera was employed to replace the commercial optical tracker for automatic real-time image registration and instrument tracking. For surgical overlay, the overlaid 3D image is expected to have the same geometric shape as the original organ. However, the resulting 3D image suffers from distortion owing to the inconsistent parameters between the digital recording and the optical reconstruction. In our previous work, the 3D image deformation issue was not well addressed and there was no quantitative evaluation on the geometric accuracy of 3D images. To reduce the distortion caused by the mismatch of the lens array and the LCD monitor, it took a lot of time to manually adjust the device itself. However, there was no guarantee that the resulting distortion could be controlled under certain level due to the lack of real-time feedback of current distortion level. The manual adjustment was performed in a trial-and-error fashion, which was very time consuming.

1.3. Improvement and contribution

This paper proposes an automatic closed-loop 3D image calibration algorithm using a stereo camera to measure the distortion of 3D images as the real-time feedback. The final distortion error can be controlled under a small tolerance and the whole calibration procedure can be done within several seconds. To the best of our knowledge, there is no relevant work on compensating 3D image distortion. The second novelty of this paper is the new design of the graphics pipeline for computer-generated integral imaging. We have obtained the best rendering performance compared with our previous work. We also have made great improvement on the AR device for surgical overlay. Unlike the prototype in our previous work, the new design is compact, integrated and flexible. We performed comprehensive phantom experiments to evaluate the new AR device in terms of the accuracy and the visual effects of the 3D image overlay.

The rest of the paper is organized as follows. Section 2 describes the principle of integral imaging and introduces the new 3D image rendering pipeline together with its implementation. Section 3 presents the 3D image calibration algorithm. Section 4 describes the new AR overlay device based on the proposed methods. Section 5 presents evaluation experiments and the results. Finally, Section 6 concludes the paper.

2. CGII rendering

The basic principle of integral imaging is illustrated in Fig. 1. In the pickup procedure as shown in Fig. 1(a), the 3D object is captured by a lens array as individual elemental images recorded on the imaging plane (e.g., charge coupled device, CCD), which is located behind the lens array. For display, an LCD is used to display the previously recorded elemental images to reproduce the original routes of rays irradiated from the 3D object, causing a 3D image to be formed (Fig. 1(b)). The pickup procedure can be simulated using computer graphics (CG) techniques by tracing the desired rays if the 3D object is represented by 3D data (i.e., surface models or volumes), and we need only an LCD and a lens array to display 3D images. The problem of generating elemental images has become a CG rendering problem. Assume the LCD used for 3D display has a pixel resolution of $n_x \times n_y$ with pixel pitches of $p_x \times p_y$ mm, we want to synthesize an image (referred to as the composite elemental image, CEI) of $n_x \times n_y$ pixels, where each pixel value represents the color of the ray connecting that pixel and its assigned elemental lens. The process of synthesizing such a CEI is called CGII rendering.

2.1. Pixel assignment

Pixel assignment is to associate each pixel of the LCD screen with its corresponding lens on the lens array. The assignment is determined by the shape and pitches of the lens array. A hexagonal lens array is preferred because it can provide a denser lens arrangement (small pitches) than that of a rectangular one, as shown in Fig. 2(a). Fig. 2(b) shows the pixel assignment when using a hexagonal lens array. Assume the distance between the screen plane $(u - v)$ and the lens array plane $(s - t)$ is g, we project the lens array plane to the screen plane orthogonally along the normal direction of the screen, and a pixel is assigned to the lens whose center is nearest to that pixel. This can be expressed mathematically as

$$PA(u, v) = (s, t) \Leftrightarrow (u, v) \in Vor(s, t) \quad (1)$$

where $PA(u, v)$ is a $\mathbb{R}^2 \to \mathbb{R}^2$ mapping function that maps the pixel at (u, v) to its corresponding lens center; $Vor(s, t)$ represents the Voronoi region associated with the lens centered at (s, t). Because a pixel can be addressed by its zero-based integer indices (i, j), a