Optics and Lasers in Engineering 115 (2019) 172–178

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

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

GPU-accelerated integral imaging and full-parallax 3D display using stereo–plenoptic camera system

Seokmin Hong[a,*], Nicolò Incardona[a], Kotaro Inoue[b], Myungjin Cho[b], Genaro Saavedra[a], Manuel Martinez-Corral[a]

[a] 3D Imaging and Display Laboratory, Department of Optics, University of Valencia, 46100 Burjassot, Spain
[b] Department of Electrical, Electronic and Control Engineering, IITC, Hankyong National University, Anseong, 17579, South Korea

ARTICLE INFO

Keywords:
3D display
Integral imaging
3D data registration
Point cloud
GPU
Plenoptic camera
Stereo camera

ABSTRACT

In this paper, we propose a novel approach to produce integral images ready to be displayed onto an integral-imaging monitor. Our main contribution is the use of commercial plenoptic camera, which is arranged in a stereo configuration. Our proposed set-up is able to record the radiance, spatial and angular, information simultaneously in each different stereo position. We illustrate our contribution by composing the point cloud from a pair of captured plenoptic images, and generate an integral image from the properly registered 3D information. We have exploited the graphics processing unit (GPU) acceleration in order to enhance the integral-image computation speed and efficiency. We present our approach with imaging experiments that demonstrate the improved quality of integral image. After the projection of such integral image onto the proposed monitor, 3D scenes are displayed with full-parallax.

1. Introduction

During the last century, three-dimensional (3D) imaging techniques have been spotlighted due to their merit of recording and displaying 3D scenes. Among them, integral imaging (InI) has been considered as one of the most promising technologies. This concept was proposed first by G. Lippmann in 1908. He presented the possibility of capturing the 3D information and reconstructing the 3D scene by using an array of spherical diopters [1–3]. Nowadays, the pickup procedure is performed by placing an array of tiny lenses, which is called microlens array (MLA), in front of the two-dimensional (2D) imaging sensor (e.g. CCD, CMOS). A collection of microimages is obtained, which is referred to as integral image. Interestingly, every microimage contains the radiance (spatial and angular) information of the rays. This is because different pixels of one microimage correspond to different incidence angles of the rays passing through each paired microlens. Figs. 1 and 2 show the comparison between a conventional and an InI (also known as plenoptic of light-field) camera. Several companies announced their plenoptic camera, which is based on Lippmann's integral photography theory [4–6]. On the other hand, in the display stage the MLA is placed in front of a screen, where is projected the integral image. The microlenses integrate the rays proceeding from the pixels of the screen and thus, reconstruct the 3D scene. Consequently, when the integral image is projected onto an InI display, observers can see the 3D scene with full-parallax and quasi-continuous perspective view.

In the meanwhile, many research groups are investigating how to acquire the depth map from the plenoptic image [7–9]. Sabater et al. [7] modeled demultiplexing algorithm in order to compose a proper 4D Light-Field (LF) image, and calculate the disparities from a restored sub-images array by using block-matching algorithm. Huang et al. [8] built their stereo-matching algorithm, and utilized it into their own framework named Robust Pseudo Random Field (RPRF) to estimate the depth map from the plenoptic image. Jeon et al. [9] calculated the depth map from an array of sub-aperture images by using the derived cost volume, multi-label optimization propagates, and iterative refinement procedure. We mainly applied Jeon's approach in our experiment.

The main contribution of this paper is to utilize the stereo–plenoptic camera system in order to get dense depth map from a pair of captured plenoptic images and get rid of the constraints of monocular vision system. Normally, multiple views can enlarge the field of view and recover the occluded information by complementing each other. For this reason, we can restore the depthless areas of the scene. Another important benefit from our proposal is to yield nicer quality of the integral image using a registered pair of point clouds. Besides, the use of the GPU acceleration technique assists to enhance the integral image's generation speed.

* Corresponding author.
E-mail address: seokmin.hong@uv.es (S. Hong).

https://doi.org/10.1016/j.optlaseng.2018.11.023
Received 28 September 2018; Received in revised form 20 November 2018; Accepted 26 November 2018

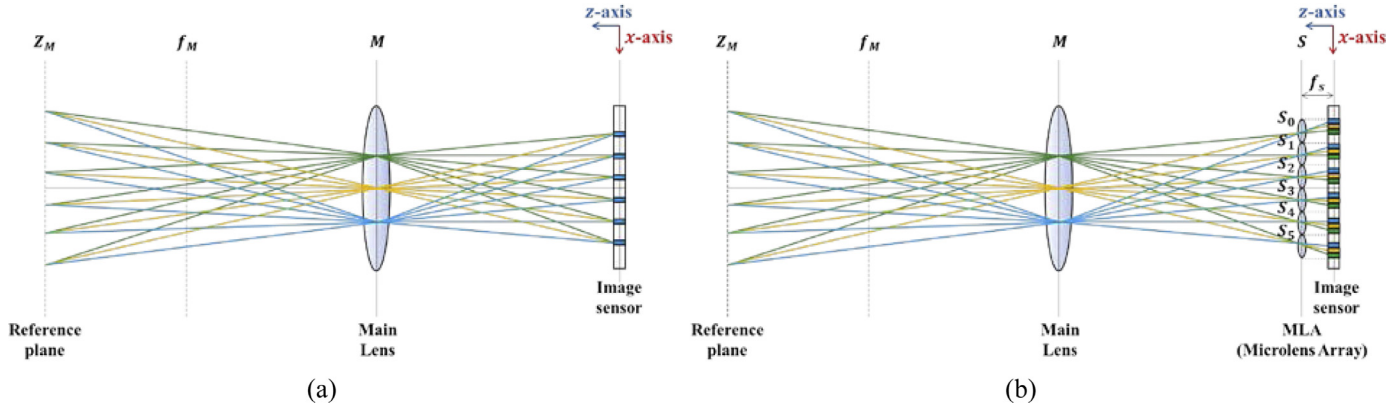

Fig 1. Scheme of image capturing system: (a) is a conventional camera; and (b) is a plenoptic camera. The pixels of (a) integrate, and therefore discard, the angular information even if they have. On the contrary, (b) can pick up both spatial and angular information thanks to the insertion of the microlens array.

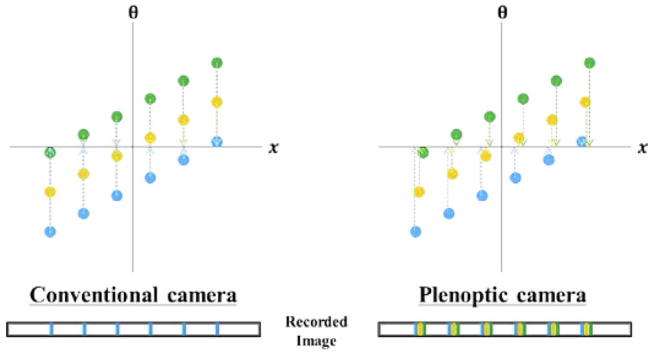

Fig 2. Illustration of the projected pixels to the imaging sensor, which are shown in the plenoptic field. The projected pixel from the conventional camera system gathers into a single pixel. However, plenoptic camera system projects all different incident information in independent pixel's position. This collected image becomes an integral image.

Fig 3. Proposed stereo–plenoptic camera system.

This paper is organized as follows. In Section 2., our previous related works are described. In Section 3., our contribution to compose and manage the point cloud from a pair of captured plenoptic images is illustrated. In Section 4. the methodology to generate an integral image from registered 3D information by using GPU acceleration technique is explained. Finally, in Sections 5. and 6. the experimental results are provided and the conclusions are carried out, respectively.

2. Related work

The closest work which is related with a stereo-type capturing and modification method has been published by our group recently. In [10] we exploited the stereo-hybrid 3D camera system composed of two Kinect sensors (Kinect v1 and v2), to take profit of different features for obtaining a denser depth map. Furthermore, we illustrated the benefit of binocular approach contrary to monocular one with some experimental results. However, the main distinction from current proposal is that [10] utilized hybrid camera set-up and obligatorily considered the remedy of the dissimilarities. Most of all, the working distance of the cameras used is restricted because of the usage of an infrared (IR) sensing technique. In this paper, we exploit the commercial plenoptic camera, named Lytro Illum. The important thing is that plenoptic cameras are passive devices in the sense that they do not need any additional light emitter. It can record the scene from the ambient light source directly. It means that the working distance of this camera is related to the camera lenses' optical properties. Furthermore, this plenoptic camera can decide the reference plane of the scene thanks to the InI's features [11,12].

In the meantime, [13] illustrated our approach to generate an integral image from a point cloud, which is ready to be projected onto an InI monitor. However, the bottleneck of this approach was that it required a long computational time. To solve this critical defect, in current approach we exploit GPU acceleration technique to generate microimages in parallel way, reducing the processing time.

3. Stereo–plenoptic image manipulation

In order to implement the stereo system, it is convenient to use two cameras of the same model. Accordingly, in our experimental system we utilized the camera slider in order to capture the scene in each different position with a single plenoptic camera, and we placed a tripod eager to configure the camera's proper position. Fig. 3 shows the camera set-up and Fig. 4 shows the overview of our experimental environment. In Section 3.1, we describe our approach to manipulate the plenoptic image and obtain the depth map from this handled image. In sequence, in Section 3.2, we explain the methodology for the arrangement and registration process of a pair of point clouds.

3.1. Plenoptic image manipulation

Our proposal in this paper is the use of commercial plenoptic camera. Its software provides various functions: it helps to choose the proper perspective view, changes the focused plane of the scene, and extracts the calculated depth map (or disparity map), color image, and an encoded raw image format [5]. Fortunately, [14,15] help to decode the

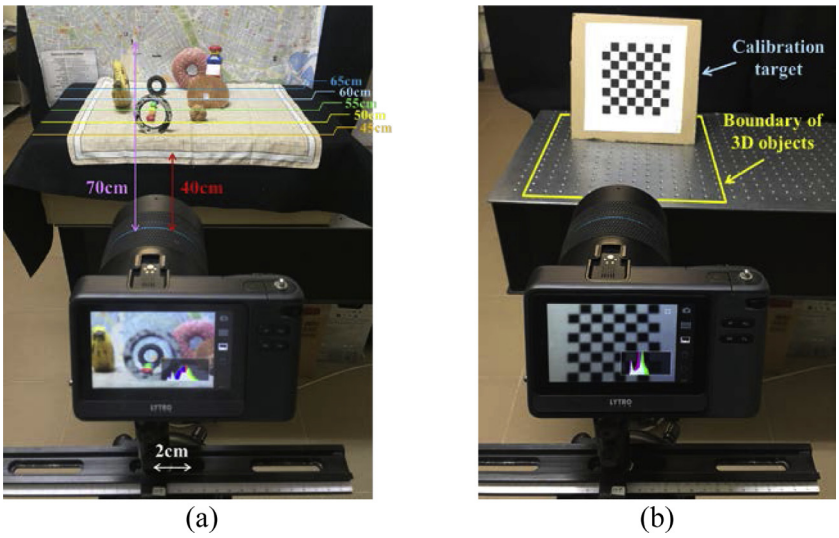

Fig 4. Overview of proposed experimental environment: (a) is capture for the main scene, and (b) is capture for the calibration process.

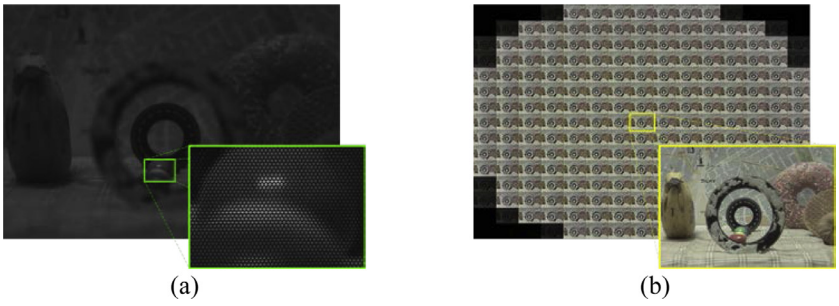

Fig 5. (a) is a raw plenoptic image from plenoptic camera, (b) is a composed sub-aperture image array from plenoptic image. See text for further details.

raw plenoptic image and extract sub-images from this encrypted data. Interestingly, this extracted raw data contains a grayscale image (see Fig. 5(a)). The main reason is that there is a Bayer color filter array over the camera's sensor to capture the colors. Thus, it must be demosaiced to get the color information back. It is noticeable that the color tones of captured images shown in Fig. 4(a) and 5(b) are different. Note, however, that the first is the image extracted from Lytro software and the other is a sub-image extracted through [14,15]. The main reason of that difference is that they use different Bayer demosaicing algorithms. We extract the sub-aperture images array in order to follow [9] approach (see Fig. 5(b)), which estimates the depth map by minimizing stereo-matching costs between sub-images with sub-pixel accuracy, and corrects the unexpected distortions. However, even after correcting the distortion problem via the referenced algorithm, the estimated depth map still has some image distortion effect. Thus, we performed the plenoptic camera calibration and rectification before the depth map calculation. The diagram of Fig. 6 shows our approach well.

Fig. 7 shows the comparison between our proposed depth map estimation strategy and the output from Lytro's software (Lytro Destktop v.5.0.1). Fig. 7(a, b) have more continuous depth levels and stable gradation than Fig. 7(c, d). On the contrary, the sharpness of the targets and the shape of the object's surfaces in the former are worse than in the latter.

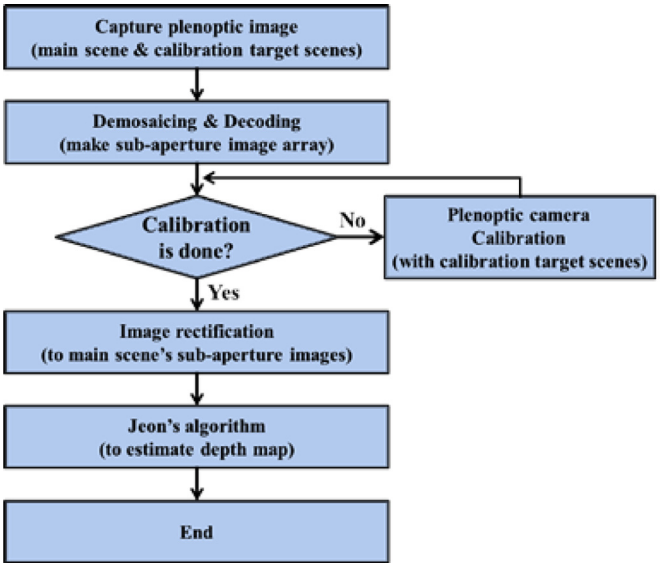

Fig 6. Flow chart of our proposed depth estimation strategy.

3.2. Point cloud modification and registration

The aim of this section is to explain how to compose the point cloud from the image, and to make registration from one point cloud to the other in order to arrange them in a proper position. In [13], we composed the point cloud from a pair of color and depth map images. We assigned six values to each point of the point cloud, namely its (x, y, z) coordinates and RGB color intensities. Each point of the RGB image corresponds to the point of the depth image having the same (x, y) coordinates. So it is sufficient to assign the corresponding depth value to all the points of the RGB image. Finally, this modified 3D information is arranged into the virtual 3D space. Afterward, we need to make registration between left and right point clouds. This is because the two scenes are mutually shifted and it is necessary to arrange them in a proper way. To solve this issue, we utilize Iterative-Closest-Point algorithm (ICP), as in [10]. ICP calculates the movement and minimizes the distance be-

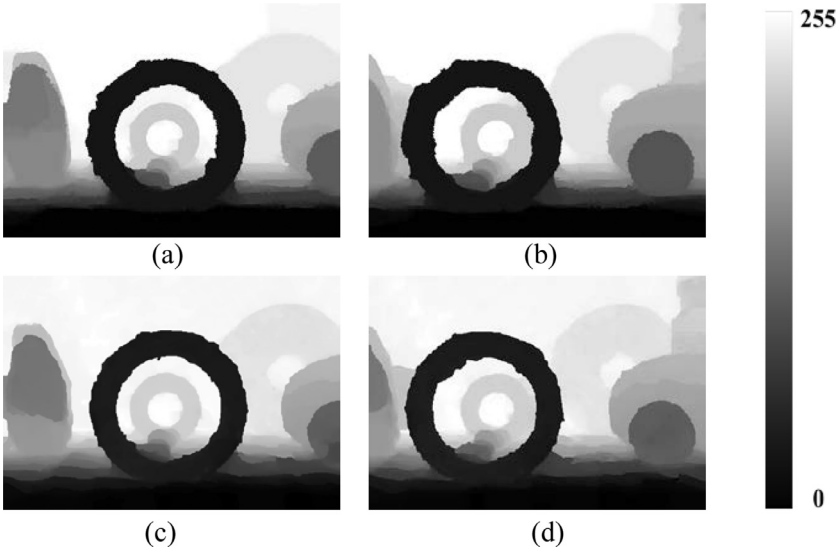

Fig 7. The depth map comparison result: top row images (a, b) are the estimated depth map result from our approach, while bottom row images (c, d) are from the output of Lytro's software. Right bar shows the depth intensity value from depth map image (0: closest area, 255: farthest area).

tween point clouds. As is well known, ICP is often used to reconstruct 2D or 3D data captured from different positions. The output of ICP algorithm is a rigid-body transformation matrix, which includes translation and rotation information [16–18]. This matrix permits to refer the position of one point cloud to the other in appropriate way.

4. Integral image generation from the point cloud with GPU acceleration

Once aligned the pair of point clouds, the resulting one is ready to generate an integral image. As we mentioned in [13], the production of an integral image is processed in a virtual 3D space using a virtual pinhole array (VPA). We place the VPA in a proper position in the virtual 3D scene, and all the points from the cloud are projected through all the pinholes by using back-projection technique, as in [19]. Interestingly, the location of the VPA will represent the position of the displayed image's reference plane. For instance, a point located behind the VPA will be reconstructed behind the MLA, while a point in front of the VPA will be reconstructed floating in front of it. Each point projected through the pinholes forms the microimages' pixels and finally, this entire back-projection mapping calculation produces the integral image.

On the other hand, we also need to consider the scale factor between input image and integral image's sizes. The main reason is that the scale factor decides the nearest-neighbor interpolation's index, as in [13]. This interpolation helps to fill the empty pixels during the back-projection mapping and as a result, proper interpolation index helps to improve the quality of the integral image. Eq. 1 and 2 show how to derive scale factors:

$$\begin{cases} Dst_w = II_w \\ Dst_h = \frac{II_w}{Org_w} \times Org_h \end{cases} \tag{1}$$

$$\lambda_{u,v} = \frac{Dst_{w,h}}{Org_{w,h}} \tag{2}$$

Where II_w is target integral image's width size, $Org_{w,h}$ is input image, $Dst_{w,h}$ is final integral image size, and $\lambda_{u,v}$ is scale factor, respectively.

However, these back-projection mapping and interpolation processes are heavy work. In order to solve this drawback, we utilize the GPU acceleration technique. The use of central processing units (CPUs) computation has the limitation due to their general purpose of usage. Even if CPUs have their own threads to compute, their performance is not sufficient to boost the computation speed because of the way of CPU's sequential implementation process and the limited number of CPU Cores (the number of threads depends on the capacity of CPU's Cores). On the contrary, GPU computation enables to execute thousands of threads to compute their mission in parallel [20,21]. It means that we can compute the integral image in a parallel way and as a result, we can speed up the computation time. Fig. 8 shows our approach and the comparison scheme between CPU and GPU computation. After this process we can get the integral image, which is ready to be displayed in an InI monitor.

5. Experimental results

In our experiment, we register the right point cloud into the space of the left one. The main reason is that the right scene not only contains the occluded information of the left scene, but also new objects appear. On the other hand, regarding the display part, we utilized the Samsung SM-T700 (14.1338px/mm) tablet as screen, and we mounted a MLA which has focal length $f = 3.3$ mm and pitch $p = 1.0$ mm (Model 630 from Fresnel Technology). We utilized 152×113 microlenses from this MLA because this is the maximum possible usage for the screen used (see Fig. 10's InI monitor set-up). A noteworthy feature is that the number of pinholes of the VPA must match the number of microlenses. The generated microimage is composed of 15×15 pixels, and thus, the full size of the integral image is 2280×1695 pixels. Finally, we need to resize the integral image to take into account the real number of pixels per microlens, so the image is finally resized to 2148×1597 (resizing factor $k = 14.1338$px/15px). Fig. 9 shows the result of produced integral images.

To show our experimental result, we composed the set-up as shown in Fig. 10. Originally, our main target are binocular observers, who can see the 3D nature of displayed scene. Unfortunately, the full-parallax effect cannot be directly demonstrated in a manuscript or even in a monocular video. In order to demonstrate this 3D effect, we replaced the binocular observer with a monocular digital camera, as recording device. A collection of pictures is obtained displacing the camera in horizontal and vertical direction. Media 1 and 2 show the result obtained with each left and right scenes, and Media 3 shows the result of the proposed method. Fig. 11 shows this experimental result with more details. Our proposed result has better quality than each, left and right, captured scenes. For instance, left and right scenes have black areas (depthless areas) caused by occlusions. On the other hand, our proposed method restores these occluded areas thanks to the registration and complementation between left and right captured scenes.

Meanwhile, we exploit the parallelism in integral image computation via NVIDIA CUDA programming model, which is a software platform for solving non-graphics problems in a parallel way [21]. Our hardware specification is the following: Intel i7 4cores in CPU, and NVIDIA

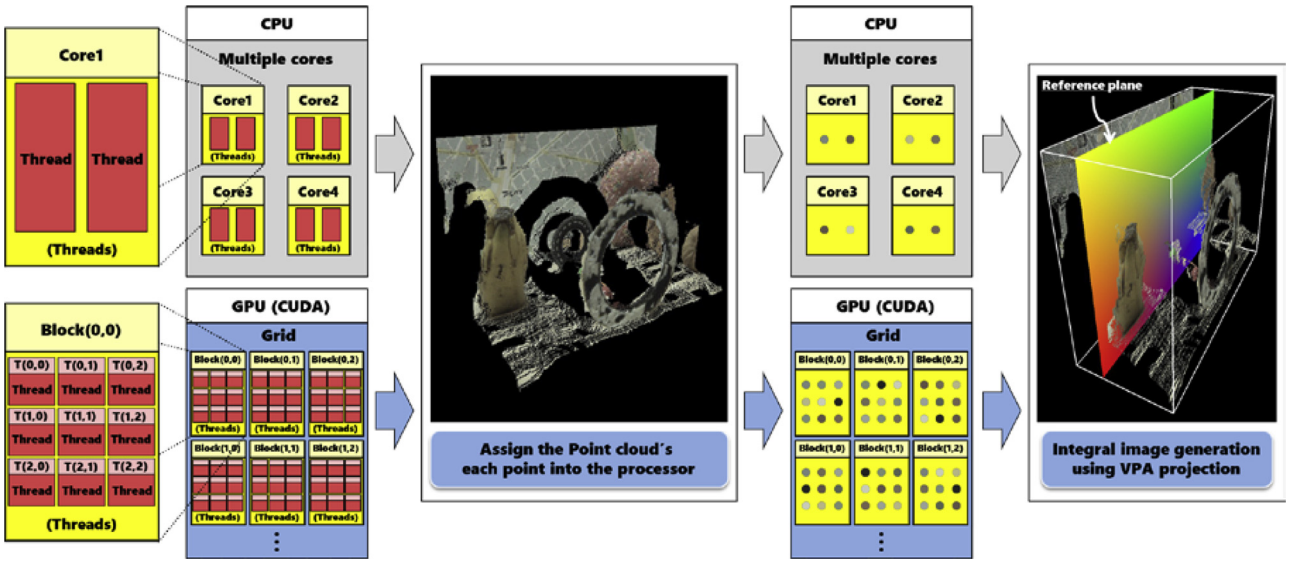

Fig 8. The comparison scheme how to compose an integral image in CPU and GPU computation. Each thread picks single 3D point from the point cloud and computes the proper pixels of an integral image using VPA projection. From the third step, GPU is able to assign thousands of points in a same time contrary to CPU.

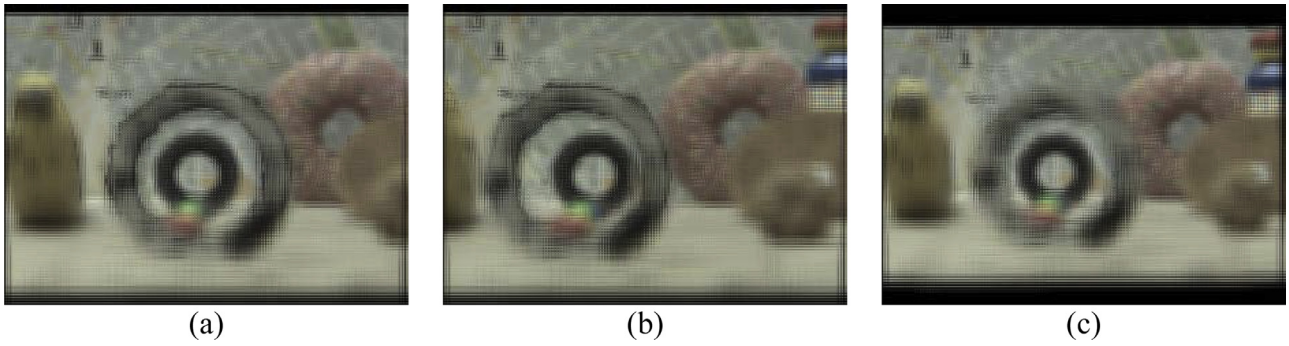

Fig 9. Composed integral image: (a) is from left scene, (b) is from right scene, and (c) image is registered scene between left and right scenes.

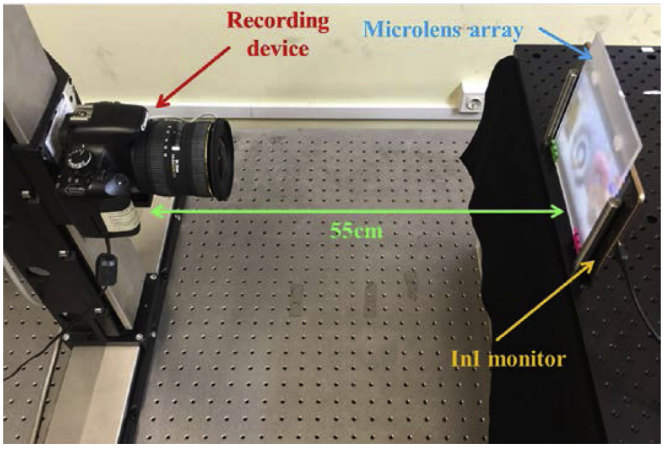

Fig 10. Overview of experimental system.

GeForce GTX 870 M in GPU. We tested the algorithm with various interpolation indices to compare the computation speeds (see Fig 12 and Table 1). We have found that the GPU implementation is much faster than CPU, especially when we increase the interpolation index. In fact, the interpolation index does not affect the computation time in the GPU implementation.

Table 1
More detail of comparison result between CPU and GPU computation time.

List	CPU(Sec.)		GPU(Sec.)	
Kind of the scene	Left, right	Registered	Left, right	Registered
0 interpolation	109.71	224.59	29.59	60.57
1 interpolation	302.39	629.87	30.47	63.46
2 interpolations	699.40	1432.56	32.68	66.94
3 interpolations	1281.99	2610.55	53.77	109.48

6. Summary and conclusion

In this paper we utilized the stereo–plenoptic camera system to display the captured plenoptic image into an InI monitor and enhance the quality of the displayed 3D image. We did a plenoptic camera calibration and rectification to solve the tilted and distorted plenoptic image's defect. Furthermore, we extracted the sub-aperture images array from the calibrated plenoptic image in order to estimate the depth map. This calculated depth map is used to compose the 3D point cloud, which is arranged into the virtual 3D space. Then we performed a registra-

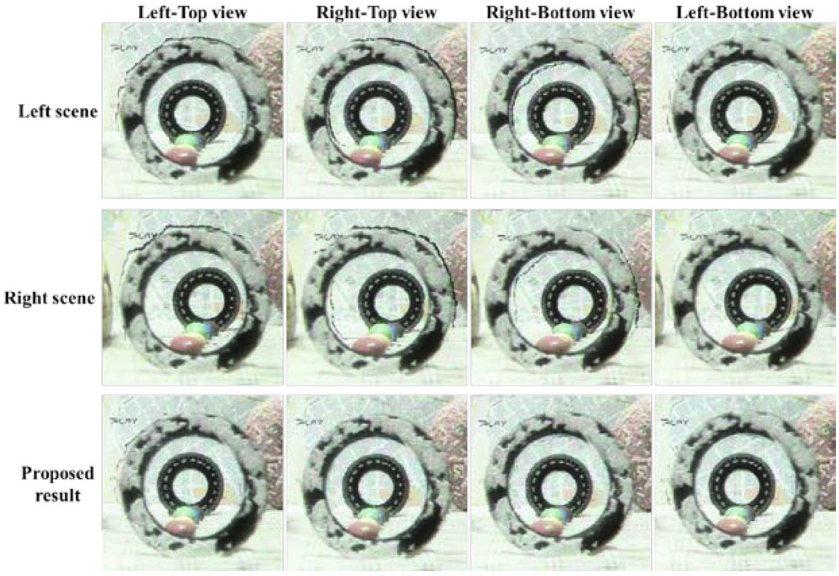

Fig 11. Comparison result between displayed integral images: first row is from left scene, second row is from right scene, and third row is our proposed result. All the images are excerpted from recorded video (Media 1, 2, and 3), and we clipped-out a specific part at the scene in order to emphasize the comparison result clearly.

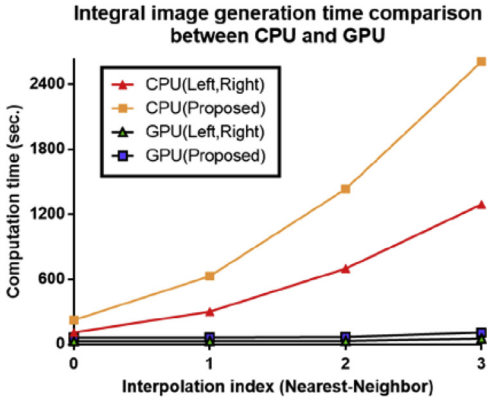

Fig 12. The integral image generation time comparison between CPU and GPU. The triangle represents the left and right scene's result, and the rectangle represents the registered scene's result.

tion between left and right scene's point clouds to arrange them in a proper position. This fused point cloud has denser 3D data and manages to recover the depthless areas properly. Finally, we generated the integral image via VPA through the back-projection method. To boost the computation time, we adopted GPU acceleration technique in this procedure. This generated integral image is displayed in our proposed integral imaging monitor and it displays an immersive scene with full parallax to the binocular observers.

In the future work, the main focus will be on the real-time implementation of the system using different and/or newer types of 3D cameras: stereo-vision camera [22,23], or even higher quality of plenoptic camera [4,6]. Another goal is to enhance the accuracy of 3D data registration using non-rigid objects mapping [24–26].

Acknowledgement

This work was supported by the Ministerio de Economía y Competitividad, Spain (Grant No. DPI 2015-66458-C2-1R), by Generalitat Valenciana, Spain (project PROMETEOII/2014/072), and by the National Research Foundation of Korea (NRF-2017R1D1A1B03030343, NRF-2017K1A3A1A19070753). S. Hong acknowledges a Predoctoral contract from University of Valencia (UV-INVPREDOC15-265754).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2018.11.023.

References

[1] Lippmann G. Epreuves réversibles photographies integrals. *Comptes Rendus de l'Académie des Sciences* 146 1908:446–51.
[2] Lippmann G. Epreuves reversibles donnant la sensation du relief. Journal de Physique Théorique et Appliquée 1908;7:821–5.
[3] Lippmann G. L'étalon international de radium. Radium (Paris) 1912;9:169–70.
[4] . Raytrix camera http://www.raytrix.de.
[5] . Lytro camera https://www.lytro.com.
[6] . PiCam: Pelican Imaging Camera http://www.pelicanimaging.com.
[7] Sabater N, Seifi M, Drazic V, Sandri G, Perez P. Accurate disparity estimation for plenoptic images. In: Proceedings European Conference on Computer Vision; 2014. p. 548–60.
[8] Huang C-T. Robust pseudo random fields for light-field stereo matching. IEEE Conf Comput Vis Pattern Recognit 2017:11–19.
[9] Jeon H, Park J, Choe G, Park J, Bok Y, Tai Y, Kweon I. Accurate depth map estimation from a lenslet light field camera. IEEE Conf Comput Vis Pattern Recognit 2015:1547–55.
[10] Hong S, Ansari A, Saavedra G, Martinez-Corral M. Full-parallax 3D display from stereo-hybrid 3D camera system. Opt Laser Eng 2018;103:46–54.
[11] Martinez-Corral M, Dorado A, Navarro H, Saavedra G, Javidi B. Three-dimensional display by smart pseudoscopic-to-orthoscopic conversion with tunable focus. Appl Opt 2014;53:19–25.
[12] Scrofani G, Sola-Pikabea J, Llavador A, Sanchez-Ortiga E, Barreiro JC, Saavedra G, Garcia-Sucerquia J, Martinez-Corral M. FIMic: design for ultimate 3D-integral microscopy of in-vivo biological samples. Biomed *Opt Express* 2018;9:335–46.
[13] Hong S, Shin D, Lee B, Dorado A, Saavedra G, Martinez-Corral M. Towards 3D television through fusion of kinect and integral-imaging concepts. J Disp Technol 2015;11:894–9.
[14] Dansereau DG, Pizarro O, Williams SB. Decoding, calibration and rectification for lenselet-based plenoptic cameras. IEEE Conf Comput Vis Pattern Recognit 2013:1027–34.
[15] Dansereau DG. Lightfield toolbox for matlab http://dgd.vision/Tools/LFToolbox.
[16] Besl PJ, Mckay ND. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14:239–56.
[17] Zhang Z. Iterative point matching for registration of free-form curves and surfaces. Int J Comput Vis 1994;13:119–52.
[18] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm. In: Proceedings Third Int. Conf. on 3-D Digit. Imaging and Mode; 2001. p. 145–52.
[19] Martínez-Corral M, Javidi B, Martínez-Cuenca R, Saavedra G. Formation of real, orthoscopic integral images by smart pixel mapping. Opt Express 2005;13:9175–80.
[20] "NVIDIA CUDA Toolkit, version 9.1", https://developer.nvidia.com/cuda-toolkit.
[21] Bilgic B, Horn BKP, Masaki I. Efficient integral image computation on the GPU. IEEE Int Intell Veh Symp 2010:528–33.
[22] Bumblebee2 .
[23] *ZED* .
[24] Hähnel D, Thrun S, Burgard W. An extension of the ICP algorithm for modeling nonrigid objects with mobile robots. Int Joint Conf Artif Intell 2003.
[25] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Comput Vis Image Underst 2003;89:2–3.

[26] Ma J, Zhao J, Jiang J, Zhou H. Non-rigid point set registration with robust transformation estimation under manifold regularization. Proc Conf Artif Intell 2017:4218–24.

Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, South Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the B.Eng. Extraordinary Award. Since 2015, he has been working with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests are image processing, computer vision, 3D display, 3D integral imaging, and applied computer science.

Nicolò Incardona received the B.Eng. and M.Sc. degrees in Electronic Engineering from Polytechnic University of Milan, Italy, in 2013 and 2016 respectively. He developed his master's thesis at Polytechnic University of Valencia, Spain. Since 2017 he has been working with the 3D Imaging and Display Laboratory, University of Valencia, Spain. His research interests are image processing, 3D display and integral microscopy.

Kotaro Inoue received the B.S. and M.S. degrees in computer science and electronics from Kyushu Institute of Technology, Fukuoka, Japan, in 2015 and 2017, respectively. He is currently a doctoral student at Hankyong National University in Korea. His research interests include visual feedback control, 3D display, 3D reconstruction, and 3D integral imaging.

Myungjin Cho received the B.S. and M.S. degrees in Telecommunication Engineering from Pukyong National University, Pusan, Korea, in 2003 and 2005, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Connecticut, Storrs, CT, USA, in 2010 and 2011, respectively. Currently, he is an associate professor at Hankyong National University in Korea. He worked as a researcher at Samsung Electronics in Korea, from 2005 to 2007. His research interests include 3D display, 3D signal processing, 3D biomedical imaging, 3D photon counting imaging, 3D information security, 3D object tracking, 3D underwater imaging, and 3D visualization of objects under inclement weather conditions.

Genaro Saavedra received the B.Sc. and Ph.D. degrees in physics from Universitat de València, Spain, in 1990 and 1996, respectively. His Ph. D. work was honored with the Ph.D. Extraordinary Award. He is currently Full Professor with Universitat de València, Spain. Since 1999, he has been working with the "3D Display and Imaging Laboratory", at the Optics Department. His current research interests are optical diffraction, integral imaging, 3D high-resolution optical microscopy and phase-space representation of scalar optical fields. He has published on these topics about 50 technical articles in major journals and 3 chapters in scientific books. He has published over 50 conference proceedings, including 10 invited presentations.

Manuel Martinez-Corral was born in Spain in 1962. He received Ph. D. degree in Physics in 1993 from the University of Valencia, which honored him with the Ph.D. Extraordinary Award. He is currently Full Professor of Optics at the University of Valencia, where he co-leads the "3D Imaging and Display Laboratory". His teaching experience includes lectures and supervision of laboratory on Geometrical Optics, Optical Instrumentation, Diffractive Optics and Image Formation for undergraduate and Ph.D. students. Fellow of the SPIE since 2010 and Fellow of the OSA since 2016, his research interest includes microscopic and macroscopic 3D imaging and display technologies. He has supervised on these topics fifteen Ph. D. students (three of them honored with the Ph.D. Extraordinary Award), published over 120 technical articles in major journals (which have received more than 2.700 citations), and pronounced over fifty invited and keynote presentations in international meetings. He is co-chair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing. He is Topical Editor of the OSA journal Applied Optics.