Towards 3D Television Through Fusion of Kinect and Integral-Imaging Concepts

Seokmin Hong, Donghak Shin, Byung-Gook Lee, Adrián Dorado, Genaro Saavedra, and Manuel Martínez-Corral

Abstract—We report a new procedure for the capture and processing of light proceeding from 3D scenes of some cubic meters in size. Specifically we demonstrate that with the information provided by a kinect device it is possible to generate an array of microimages ready for their projection onto an integral-imaging monitor. We illustrate our proposal with some imaging experiment in which the final result are 3D images displayed with full parallax.

Index Terms—Integral imaging, kinect, 3D monitors.

I. INTRODUCTION

▼ ONVENTIONAL photography is fully adapted for recording in a 2D sensor the images of the 3D world. Although the images produced by photography are essentially 2D, they carry many cues that account for the 3D nature of the recorded scenes. This is the case, among others, of the perspective rules, which make closer objects to appear bigger than further ones. This effect is due to the well-known fact that the size of the image in the photographic sensor is determined by the angular size of objects. Other cues are shadows, occlusions, or defocus. In case of video recording, the relative speed of moving objects (or static objects when the camera is moving) is also a significant depth cue. For most of applications, the capture and display of 2D images provides enough information and/or satisfaction to users and minimizes the amount of data to be stored, transmitted and displayed. This is the reason for the still massive use of 2D photography and video.

Manuscript received August 28, 2014; revised November 09, 2014; accepted November 11, 2014. Date of publication November 13, 2014; date of current version October 05, 2015. This work was supported in part by the Plan Nacional I+D+I (grant DPI2012-32994), Ministerio de Economía y Competitividad (Spain), by Generalitat Valenciana (Spain) under project PROMETEOII/2014/072, and by the IT R&D program of MKE/KEIT [10041682].

S.-M. Hong, D.-H. Shin and B.-G. Lee are with the Institute of Ambient Intelligence, University of Dongseo, 617-716 Busan, Korea (e-mail: tom-myhsm@gmail.com; shindh2@gmail.com; lbg@gdsu.dongseo.ac.kr).

A. Dorado, M. Martínez-Corral and G. Saavedra are with the Department of Optics, University of Valencia, E-46100 Burjassot, Spain (e-mail: adrian.do-rado@uv.es; genaro.saavedra@uv.es; manuel.martinez@uv.es).

This paper has supplementary material available at http://www.ieeexplore. ieee.org, provided by the author. This pack contains two videos (Media 1 and Media 2). Media 1.mov shows a 3D scene, composed by two office chairs, as displayed by an integral-imaging monitor. Media 2.mov shows a 3D scene, with a human model, as displayed by an integral-imaging monitor. The file is 15.4 MB in size.

Color versions of one or more of the figures are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JDT.2014.2370734

However, the need for capturing and displaying the 3D information of 3D scenes is increasing very fast in the 21st Century. Its potential application in, for example, microscopy [1], [2], medical imaging [3]–[6], optical inspection in production chains [7], security monitoring [8], or virtual simulators for civil or military applications [9], etc., makes the capture and display of 3D images a hot topic in the research end/or engineering for the next decade.

If we discard, at this moment, holography, which still needs coherent illumination, or stereoscopy, which does not provide real 3D experiences, we can affirm that technology based on the Integral Photography principle is in the right way of producing acceptable 3D experience. Integral Photography was proposed, in 1908, by Gabriel Lippmann [10]. His proposal intended to face the problem of conventional photographic cameras which, when working with 3D scenes, do not have the ability of recording the angular information carried by the rays of light passing through their objective [10]. Instead, the irradiance received by any pixel is proportional to the sum of radiances of all the rays, regardless of their incidence angle. To overcome this lack, Lippmann proposed to insert a microlens array (MLA) in front of the photographic film. This permits to register an array of microimages, which store a radiance map with the spatial and angular information of all the rays proceeding from the 3D scene.

The radiance map has been named in different ways, such as integral photography [10], integral imaging [12], lightfield map [13] or even plenoptic map [14], [15]. From this map it is possible, for example, to tackle the challenge of displaying 3D images with a flat monitor [16]–[19].

As for the methods for the capture of the integral imaging, some application-dependent proposals have been made along the past few years. When the 3D scene is small and close to the camera, the plenoptic architecture, in which the MLA is inserted at the focal plane of the camera lens, seems to be the best adapted [20],[21]. Also interesting is the use of a small array of tiny digital cameras [22], which can be inserted in a cellular phone. However, due to low parallax, its utility is restricted to close objects.

When the 3D scenes are much bigger, of the order of some cubic meters, a different capture rig is necessary. In this case, the most useful proposal have been based on the use of large camera arrays, arranged either in 1D or in 2D grid [23], [24]. Note that in this case, the proposed techniques need an extremely accurate synchronization between the cameras, and make use of a huge amount of data, which are unnecessary for display purposes.

1551-319X © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications standards/publications/rights/index.html for more information.

Our aim here is to propose the fusion between two concepts that are very different, but which are very successful in the area of 3D imaging and sensing. We refer to integral imaging and to kinect technology. This kind of fusion was proposed previously, but with different aim [25]. Kinect technology permits the registration, in real time, of accurate depth maps of big, opaque, diffusing 3D scenes. This is obtained with low resolution, which, however, matches perfectly with the requirements of resolution of integral-imaging monitors. Then, we propose first to capture the sampled depth map of a 3D scene with the Kinect. Second, simulate with our software, the capture of the sampled depth map with an array of digital cameras whose position, pitch and resolution are in good accordance with the characteristics of the integral-imaging monitor. And third, to project this information onto the monitor, so that the lenses of the MLA integrate the light emitted by the pixels, producing 3D scenes displayed with continuous perspective and full parallax.

II. ACQUIRING 3D POINTS CLOUD WITH KINECT

Although, as stated above, there are different methods to record the information of a three-dimensional scene, in this contribution we used a Kinect device, which was initially launched, by Microsoft¹, as an add-on accessory for the Xbox game console on 2010. Its unique features have been determinant to find applications in human's full body tracking [26], motion detection and voice recognition. However, the distinctive hallmark of this device is its capability for recording simultaneously the RGB image and the depth information in real-time. This can be made because the Kinect has two different cameras, which can operate with the same resolution [27], a RGB camera and an infra-red (IR) one. The principle behind the Kinect technology is based on depth mapping obtained from projected structured IR patterns. The Kinect's IR emitter projects a fixed pattern onto the target and both the depth distance and the 3D reconstructed map are obtained from the reflected pattern recorded by the IR camera [28], [29]. The depth information provided by the Kinect is ranged between 800–6000 mm from the sensor plane. However, data should be acquired, generally, between 1000 and 3000 mm. This is due to the quality degradation of the depth data for larger distances as result of the noise and its own low resolution [30].

Our aim here is to achieve a point cloud that includes the information of 3D position and color intensity. Although the Kinect provides such information, a limitation comes out from the fact that the two cameras (RGB and IR) are physically separated from each other and, therefore, their fields of view are different. Consequently, both scenes will not match properly, see left picture in Fig. 1.

To overcome this drawback, we use the function named 'NuiImageGetColorPixelCoordinateFrameFromDepthPixelFrameAtResolution' which is provided by the software development kit (SDK) of the Kinect, from Microsoft.². This function operates by matching depth information onto RGB image and it works in real time. Fig. 1, right picture, shows the final result after the proper matching between two cameras.



Fig. 1. (left) Raw mapping result and (right) reconstructed mapping result after our proposal method. See text for further details.



Fig. 2. Display of a 3D points-cloud. From both panels, it is clear that each point is given by its (x, y, z) position and its color intensity.

After matching the two images, we reassign the information into points located in a 3D virtual space using OpenGL environment. Now, each point is defined by six values: its (x, y, z) coordinates and its RBG color intensities. Based on their depth positions, the correct arrangement of the whole points is satisfied by using Standard Template Library (STL) and its associative container that is called 'multimap'.³, ⁴. In Fig. 2, the generated 3D point cloud is shown.

III. DEPTH ARRANGEMENT OF THE 3D POINTS-CLOUD

The next step of our procedure is to prepare the algorithm for the calculation of the microimages for their projection onto the integral-imaging monitor. To this end we need first to express the spatial 3D coordinates of the points in a homogeneous way. Take into account that coordinates of the 3D point cloud produced after the previous section, are expressed in pixels (x and y coordinates) and in millimeters (z coordinate). To make the system homogeneous we performed a calibration experiment,

⁴STL-map. [Online] Available: http://msdn.microsoft.com/library/ lfe2x6kt(v=vs.110).aspx

²Kinect SDK. [Online] Available:http://msdn.microsoft.com/en-us/library/ hh855347.aspx

³STL. [Online] Available: http://en.wikipedia.org/wiki/Standard_Template Library

¹Microsoft Kinect. [Online] Available: http://www.xbox.com/en-us/kinect/



Fig. 3. Scheme of the algorithm for calculating the microimages. The number of pixels of the integral image and the number of (x,y) pixels of the points cloud are similar.

using a chessboard as the object, and concluded that in our case one pixel was equal to two millimeters in the object space.

For the second step, the characteristics of the InI monitor need to be known and expressed in pixel coordinates. Specifically, in our experiment we used an iPad equipped with retina display (264 pixels/inch), and a MLA consisting of 147×147 lenslets of focal length $f_L = 3.3$ mm and pitch p = 1.0 mm (Model 630 from Fresnel Technology). Then, for our algorithms, any microimage was composed by 11 pixels, the gap between the microlenses and the display was fixed to g = 36.3 px, and therefore the full size of the integral image would be at most 1617×1617 .

IV. MICRO-IMAGES GENERATION

To generate the microimages we first resize laterally the 3D points cloud from 480×640 pixels to 1213×1617 pixels. Note that this change does not produce any distortion, since the ratio is still 4:3. In our computer calculation we simulate an experiment of capture of microimages. In this experiment we placed the points cloud at a certain distance from a simulated pinhole array. The distance is equal to the distance between the original scene and the kinect (see Fig. 3). Note that from now and hereafter, the plane where the synthetic pinhole array is placed will be named as the reference plane. Then we assigned the values of the pixels of the microimages by back-projection through the pinholes, as in [31].

Note that when these microimages are projected onto the monitor and displayed through the microlenses, the result will be a 3D image that is floating at a big distance from the monitor. This can reduce drastically the resolution of the displayed 3D scene, and provide a windowed aspect to it. What is more convenient is to prepare a set of microimages such that the displayed 3D image is in the neighborhood of the MLA, with some parts in front of it and some other parts behind. To obtain this directly with our algorithm, we can shift axially the 3D cloud towards the synthetic pinhole array, so that the reference plane is within the 3D scene (see Fig. 4). From the figure it is apparent that the microimages strongly depend on the reference plane position.



Fig. 4. Scheme of the back-projection algorithm. (a) The reference plane is close to the 3D point cloud; (b) The reference plane is within the 3D cloud. Pixel assignment strongly depends on the cloud position.



Fig. 5. Collection of microimages generated from the 3D points cloud captured with the kinect. (a) The full scene is in front of the pinhole array, like in the scheme shown in Fig. 3. (b) The 3D points cloud was displaced toward the pinhole array. In this case the front part of the seat of the red office chair is at the pinhole-array plane.

Finally, following Okano [32], we rotate any microimage by 180° about its center to avoid the pseudoscopic display, and resize the matrix to 1145×1527 to take into account the resolution of the retina display (10.39 px/mm).

V. EXPERIMENTAL RESULTS

First we show, in Fig. 5, the microimages calculated for two different positions of the reference plane.



Fig. 6. Single-frame excerpt from video recording of the implemented InI monitor (Media 1). The video is composed of views of the monitor obtained from different horizontal and vertical positions.





Fig. 7. Kinect output for a 3D scene with human model. (a) RGB picture. (b) Depth map from the IR picture.

Then, the microimages shown in Fig. 5(b) were displayed on iPad. The MLA was properly aligned so that pixels were close to the focal plane. Exact adjustment could not be made, due to the transparent plate that covers the retina display. This small misadjustment, about 0.5 mm, resulted in some braiding effect [33]. The InI monitor is shown in Fig. 6. It is apparent from the movie that the monitor projects a full parallax 3D image. This has been possible after a single shot capture thanks to the fusion between the kinect capture and integral-imaging processing and display.



Fig. 8. 3D points cloud of the captured scene. In green text we show the parameters for the microimages calculation and the position of the reference (or pinhole-array) plane.



Fig. 9. Integral image ready for its projection onto the InI monitor. Note that the reference plane was set just behind the back of the human model.



Fig. 10. Single-frame excerpt from video recording of the implemented InI monitor (Media 2).

To confirm the utility of our approach we did a second experiment and applied the procedure to a 3D scene with a human model. In Fig. 7 we show the RGB and the depth-map images obtained with the kinect. From this information we calculated the 3D points cloud shown in Fig. 8. In such figure we show also the parameters for the microimages calculation. Finally, in Fig. 9 we show the microimages obtained after application of our algorithm, which are ready for their projection onto the retina display of the iPad.

Again we aligned the MLA, just placed in contact with the cover plate of the retina display, and arranged the InI monitor for displaying the 3D image of the human model. Note that in this case we set the reference plane just behind the back of the model, so that all the body and mainly the left hand were reconstructed floating (about 3 cm the hand) in front of the monitor. This cannot be perceived in the video, but was clear for binocular observers. To show here the 3D nature of the displayed image we have recorded a video composed by views of the monitor obtained from different horizontal and vertical perspectives.

VI. CONCLUSION

In this paper, we have reported a novel procedure for the capture, processing and projection of integral images. Whereas the plenoptic camera is the best suited for the capture of integral images of small 3D scenes, the method proposed here can gain competitive advantage over other methods for the capture of integral images of big 3D scenes. Main advantage of fusing the kinect and the integral imaging concepts is the acquisition speed, and the small amount of handled data. Also the algorithms proposed are simple. We have demonstrated the utility of our method with two experiments, which show that full-parallax 3D images can be displayed by an InI monitor, and that calculated microimages can be adapted to the characteristics of the monitor. In further research we will combine the 3D points clouds obtained with a pair of Kinect, in order to tackle the problem of potential occlusions.

REFERENCES

- M. Martínez-Corral and G. Saavedra, "The resolution challenge in 3D optical microscopy," *Progress in Optics*, vol. 52, pp. 1–67, 2009.
- [2] M. Weber, M. Mickoleit, and J. Huisken, "Lightsheet microscopy," *Methods in Cell Biology*, vol. 123, pp. 193–215, 2014.
- [3] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao, "Augmented reality navigation with automatic marker-free image registration using 3-D image overlay for dental surgery," *IEEE Trans Biomed Eng.*, vol. 61, no. 4, pp. 1295–1304, Apr. 2014.
- [4] M. Martinez-Corral, "Multiperpective Fundus camera," Spain Patent ES 2442178 B2, Jul. 23, 2014.
- [5] J. Geng and J. Xie, "Review of 3-D endoscopic surface imaging techniques," *IEEE Sensors J.*, vol. 14, no. 4, pp. 945–960, Apr. 2014.
- [6] I. Marcus, I. T. Tung, E. O. Dosunmu, W. Thiamthat, and S. F. Freedman, "Anterior segment photography in pediatric eyes using the Lytro light field handheld noncontact camera," *J. AAPOS*, vol. 17, pp. 572–577, 2013.
- [7] U. Perwass and C. Perwass, "Digital imaging system, plenoptic optical device and data processing method," Germany Patent WO 2010121637 A1, Oct. 28, 2010.
- [8] T. G. Georgiev and A. Lumsdaine, "Methods and apparatus for rich image capture with focused plenoptic cameras," U.S. Patent 8345144 B1, Jan. 1, 2013.
- [9] C. D. Huston and C. Coleman, "System and method for creating an environment and for sharing a location based experience in an environment," U.S. 20130222369 A1, Aug. 29, 2013.
- [10] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," J. Phys., vol. 7, pp. 821–825, 1908.
- [11] R. Ng, "Digital Light Field Photography," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, USA, 2006.
- [12] H. Arimoto and B. Javidi, "Integral three-dimensional imaging with digital reconstruction," Opt. Lett., vol. 26, pp. 157–159, 2001.
- [13] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light field microscopy," ACM Trans. Graph., vol. 25, pp. 924–934, 2006.
- [14] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with plenoptic camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, pp. 99–106, 1992.
- [15] T. Georgiev and A. Lumsdaine, "The focused plenoptic camera and rendering," J. Elect. Imag., vol. 19, no. 2, Feb. 2010.
- [16] J. Geng, "Three-dimensional display technologies," Adv. Opt. Photon., vol. 5, pp. 456–535, 2013.

- [17] X. Xiao, B. Javidi, M. Martínez-Corral, and A. Stern, "Advances in three-dimensional integral imaging: Sensing, display, applications," *Appl. Opt.*, vol. 52, pp. 546–560, 2013.
- [18] J. Hong, Y. Kim, H.-J. Choi, J. Hahn, J.-H. Park, H. Kim, S.-W. Min, N. Chen, and B. Lee, "Three-dimensional display technologies of recent interest: Principles, status, issues," *Appl. Opt.*, vol. 50, pp. H87–H115, 2011.
- [19] M. Martinez-Corral, A. Dorado, H. Navarro, G. Saavedra, and B. Javidi, "3D display by smart pseudoscopic-to-orthoscopic conversion with tunable focus," *Appl. Opt.*, vol. 53, pp. E19–E26, 2014.
- [20] "Lightfield based commercial digital still camera," [Online]. Available: http://www.lytro.com
- [21] M. Miura, J. Arai, T. Mishina, M. Okui, and F. Okano, "Integral imaging system with enlarged horizontal viewing angle," *Proc. SPIE*, vol. 8384, p. 838400, 2012.
- [22] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "PiCam: An ultra-thin high performance monolithic camera array," *ACM Trans. Graphics*, vol. 32, no. 166, 2013.
- [23] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview imaging and 3DTV," *IEEE Signal Processing Magazine*, vol. 24, pp. 10–21, 2007.
- [24] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," ACM SIGGRAPH 2005.
- [25] T. Nasrin, F. Yi, S. Das, and I. Moon, "Partially occluded object reconstruction using multiple Kinect sensors," *Proc. SPIE 9117*, p. 91171G, 2014.
- [26] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, and R. Moore *et al.*, "Real-time human pose recognition in parts from single depth images," in *Proc. IEEE Computer Soc. Conf. on Computer Vision* and Pattern Recogn. (CVPR), 2011.
- [27] J. Kramer, N. Burrus, F. Echtler, H. C. Daniel, and M. Parker, "Multiple kinects," in *Hacking the Kinect, Apress*, 2012, pp. 207–246.
- [28] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth Mapping Using Projected Patterns," U.S. Patent 20080240502A1, Oct. 2, 2008.
- [29] M. Lee and J. Jeon, "Personal computer control using kinect," Kor. Inst. Inf. Scientists and Eng., vol. 39, no. 1(A), 2012.
- [30] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, pp. 1437–1454, 2012.
- [31] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca, and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," *Opt. Express*, vol. 13, pp. 9175–9180, 2005.
- [32] F. Okano, H. Hoshino, J. Arai, and I. Yayuma, "Real time pickup method for a three-dimensional image based on integral photography," *Appl. Opt.*, vol. 36, pp. 1598–1603, 1997.
- [33] H. Navarro, R. Martínez-Cuenca, A. Molina-Martín, M. Martínez-Corral, G. Saavedra, and B. Javidi, "Method to remedy image degradations due to facet braiding in 3D integral imaging monitors," *J. Display Technol.*, vol. 6, no. 10, pp. 404–411, Oct. 2010.



Seokmin Hong received the B.Eng. and M.Sc. degrees in digital and visual contents from Dongseo University, Busan, Korea, in 2012 and 2014, respectively. In 2012, Dongseo University honored him with the B.Eng. Extraordinary Award.

Since 2012, he has been working with Institute of Ambient Intelligence, Dongseo University, Busan, Korea. His research interests are image processing, computer vision and applied computer science.



Donghak Shin received the B.S., M.S., and Ph.D. degrees in telecommunication and information engineering from Pukyong National University, Busan, Korea, in 1996, 1998, and 2001, respectively.

From 2001 to 2004, he was a senior researcher with TS-Photon established by Toyohashi University of Technology, Japan. From 2005 to 2006, he was with the 3D Display Research Center (3DRC-ITRC), Kwangwoon University, Korea. He worked as a research professor at Dongseo University in Korea from 2007 to 2010. He was a visiting scholar in electrical & computer engineering department at the University of Connecticut from 2011–2012. He is currently a senior research in Institute of Ambient Intelligence, Dongseo University in Korea. His research interests include 3D imaging, 3D displays, optical information processing, and holography.



Byung-Gook Lee received the B.S. degree in mathematics from Yonsei University, Korea, in 1987, and the M.S. and Ph.D. degrees in applied mathematics from Korea Advanced Institute of Science and Technology (KAIST), in 1989 and 1993, respectively.

He worked at the DACOM Corp. R&D Center as a senior engineer from March 1993 to February 1995. He has been working at Dongseo University since 1995. He is currently a Full Professor with the Division of Computer Information Engineering at Dongseo University, Korea. His research interests

include computer graphics, computer aided geometric design, and image processing.



Adrián Dorado was born in Spain in 1988. He received the B.Sc. and M.Sc. degrees in physics from the University of Valencia, Spain, in 2011 and 2012, respectively.

Since 2010, he has been with the 3D Imaging and Display Laboratory, Optics Department, University of Valencia, Spain. His research interests include 3D imaging acquisition and display.



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 Martínez-Corral was born in Spain in 1962. He received the M.Sc. and Ph.D. degrees in physics from the University of Valencia, Spain, in 1988 and 1993, respectively. In 1993, the University of Valencia honored him with the Ph.D. Extraordinary Award.

He is currently Full Professor of Optics at the University of Valencia, where he is co-leader of the "3D Imaging and Display Laboratory". His research interest includes resolution procedures in 3D scanning microscopy, and 3D imaging and display

technologies. He has supervised on these topics seven Ph. D. theses, two of them honored with the Ph.D. Extraordinary Award. He has published over eighty technical articles in major journals, and pronounced over thirty invited and five keynote presentations in international meetings. He has been member of the Scientific Committee in over twenty international meetings.

In 2010, Dr. Martinez-Corral was named Fellow of the SPIE. He is co-chair of the Three-Dimensional Imaging, Visualization, and Display Conference within the SPIE meeting in Defense, Security, and Sensing (Baltimore). He is Topical Editor of the IEEE/OSA JOURNAL OF DISPLAY TECHNOLOGY.