

Full-parallax 3D display from single-shot Kinect capture

Seokmin Hong, Adrián Dorado, Genaro Saavedra, Manuel Martínez-Corral

Department of Optics, University of Valencia, E-46100, Burjassot, Spain
hongse@alumni.uv.es, Adrian.Dorado@uv.es, Genaro.Saavedra@uv.es, Manuel.Martinez@uv.es

Donghak Shin, Byung-Gook Lee

Institute of Ambient Intelligence, University of Dongseo, 617-716, Busan, South Korea
shindh2@gmail.com, lbg@gdsu.dongseo.ac.kr

ABSTRACT

We propose the fusion between two concepts that are very successful in the area of 3D imaging and sensing. Kinect technology permits the registration, in real time, but with low resolution, of accurate depth maps of big, opaque, diffusing 3D scenes. Our proposal consists on transforming the sampled depth map, provided by the Kinect technology, into an array of microimages whose position, pitch and resolution are in good accordance with the characteristics of an integral-imaging monitor. By projecting this information onto such monitor we are able to produce 3D images with continuous perspective and full parallax.

Key words: 3D display, Integral Imaging, Kinect

Photography is a field in continuous development that even after been researched for long time. It still keeps his place as one of the most interesting topics in research. Conventional photography is used for recording two-dimensional (2D) images from the three-dimensional (3D) world space using a 2D sensor. But, that information is incomplete. The reason is that we cannot represent the real 3D world with a 2D image. For example, with an image taken using a conventional camera is not possible to figure out about concealed objects or overlapped regions in the real world's scene. In order to solve these problems, 3D recording and displaying technics were developed, some among them are holography, stereography or Integral Imaging. 3D imaging has a lot of interesting applications in for example microscopy [1, 2], optical inspection in production chains [3], medical imaging [4-6], security monitoring [7], machine learning [8-12] or 3D display. In fact, capture and display of 3D information is a prospective topic for the next decade.

Recently, integral imaging (InI) has been considered as one of the most effective technologies to display 3D images. It was proposed, in 1908, by Gabriel Lippmann [13] under the name of integral photography. InI can produce 3D scenes displayed with continuous perspective and full parallax for the observer. On the other hand, computer vision fields can get 3D information from a 3D scene more effective and corrective than using other methods. Thus, our proposal is to use computer vision techniques for capturing the 3D scene, generating a 3D points cloud in a virtual space using the recorded 3D information and, finally generating an integral image ready to be displayed in an integral imaging monitor.

To pick up the 3D scene, we choose the Kinect device, which was launched by Microsoft in 2010 [14]. Actually, this device was an extra attachment for the x-box to detect human body's motion and track user's body continually. However, most researchers are interested in its capacity to capture, in real time, RGB image and depth information [15]. The principle of Kinect's technology is based on the projection of an infrared (IR) structured light. The Kinect's IR projector can emit a fixed IR pattern whose reflection is recorded by an IR camera. Then, the Kinect's software can calculate the depth information and generate a 3D map itself [16, 17]. However, there are some problems to process with the Kinect. In our proposal, we need to display both, RGB and depth information into a 3D virtual space in order to reconstruct the 3D points cloud information. Unluckily, Kinect cameras have different field of view (FOV) and physical locations. There are various solutions to solve these limitations, but we used a given function from the Kinect software-development-kit (SDK). This special function makes depth information mapping into the RGB image resolution size [18]. In Fig. 1, right picture, we show the final result after the adequate mapping between Kinect's two cameras. After this mapping, we define the information into each point which includes intensity of color (R, G, B) with 3D posi-

tion (x, y, z) from the Kinect respectively. Fig. 2 shows that the points cloud information can reconstruct the virtual 3D scene correctly.



Figure 1 Raw mapping between RGB image and depth information's result (left) and appropriate mapping result after our proposed method (right).

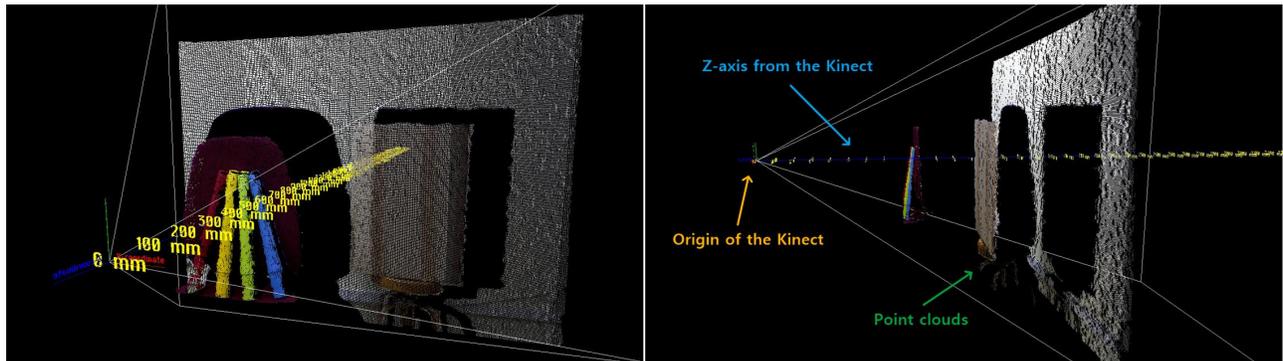


Figure 2 Reconstruction of 3D points cloud into the 3D virtual space. Both panel shows that each point has its own position and color intensity's information. The depth information is determined by the physical location of the Kinect.

Our main goal here is to manipulate the real 3D points cloud in order to build the microimages and project them onto an InI monitor. However, we had some difficulties; standard units are different to calculate in our experiment and the Kinect's camera's resolution is too low to project it into the integral imaging monitor. Therefore, we prepare some algorithms to calculate correctly the elemental images; projection onto the integral imaging monitor and resizing calculations. To solve these critical points, firstly, we need to unify the units in our system. Basically, an image resolution's unit is represented by the pixel, but a depth value of distance and physical lenslet array's units are represented by the millimeter. In order to unify these units, we choose the calibrator as a measurement object (into our experiment, we choose the chessboard) and found that one pixel becomes equal to two millimeters. To be specific, in our experiment we used an iPad equipped with retina display (264 pixels/inch) as the integral imaging monitor, and a micro-lenslet-array (MLA) consisting of 147×147 lenslets of focal length $f_l = 3.3\text{mm}$ and pitch $p = 1.0\text{mm}$ (Model 630 from Fresnel Technology). For our algorithms, any microimage was composed by 11 pixels, the gap between the microlenses and the display was fixed to $g = 36.3\text{px}$, and the full size of the InI was 1617×1617 pixels. The Kinect's usual resolution is 640×480 , and our target size is 1617×1213 . To resize the points cloud, we used the basic interpolation method which is called nearest-neighbor interpolation for each points cloud's point individually.

To generate an InI, we put a virtual pinhole array into the same virtual space of the 3D points cloud and back project each point through the pinholes, as in [19]. For the back projection from the points cloud into the pinhole array, we need to rearrangement the points to prevent overlapping each other. Fig. 3 shows the three different planes chosen

for the pinhole array, called reference plane. The integral imaging monitor is shown in Fig. 4. The monitor projects a full parallax 3D image to the observer.

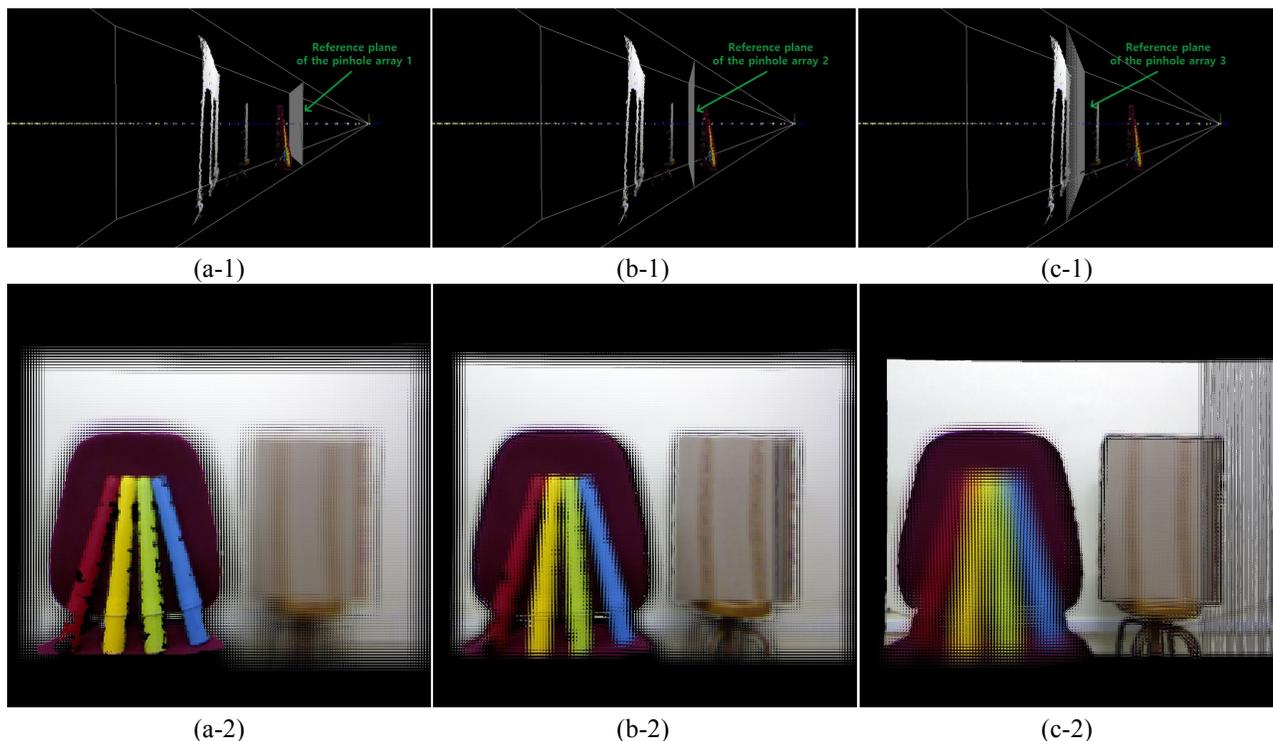


Figure 3 Three different reference planes of the pinhole array (a-1, b-1, c-1) and their back projected images (a-2, b-2, c-2).



Figure 4 Single frame extractions from the video recording of our experimentation into the integral imaging monitor. The video consists of images of the monitor obtained from different positions using a camera.

To test the usefulness of our approach, we made a second experiment. This time, the 3D scene consisted of a human model. In Fig. 5, we show the captured RGB and calibrated depth image from the Kinect simultaneously. After capturing the 3D information, we place the 3D points cloud with the virtual pinhole array into the virtual 3D space, see Fig. 6. Next, we calculate the microimages using our algorithm. Finally, the microimages are projected into our integral imaging monitor, see Fig. 7.

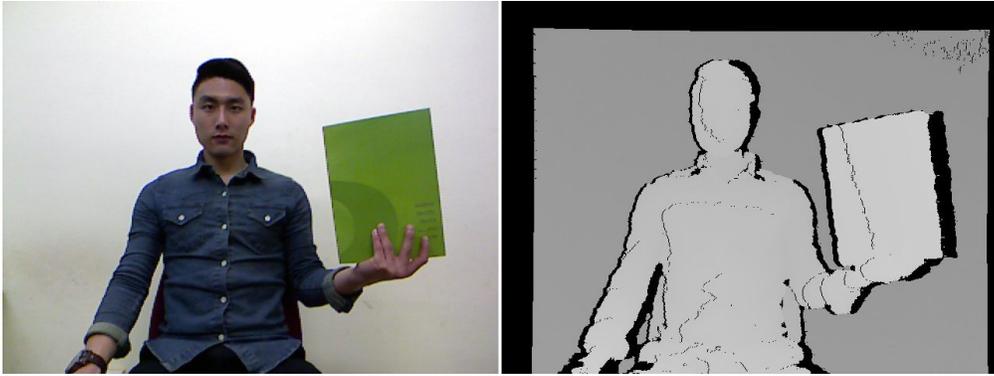


Figure 5 The Kinect capture RGB and depth image for the human model in a same time. In this scene, depth image is mapped into the RGB image's size through our suggested calibration way. Just for reference, both image's resolution size is 640×480.

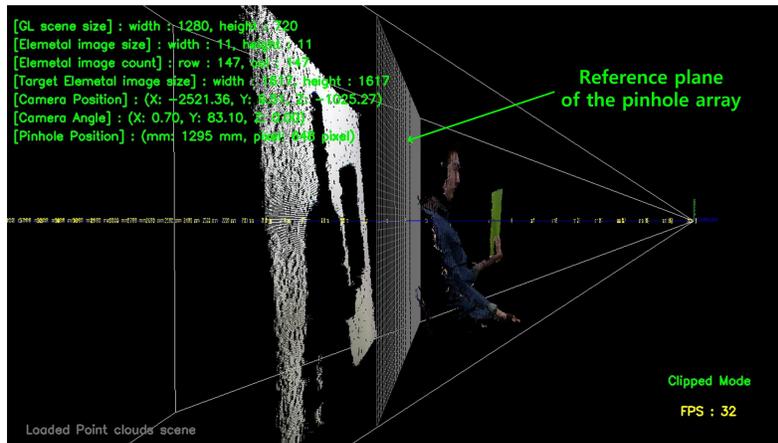


Figure 6 Captured static scene of the 3D points cloud. Through the green text we comment the parameters for the microimages' calculation and the position of the reference plane of the pinhole array.

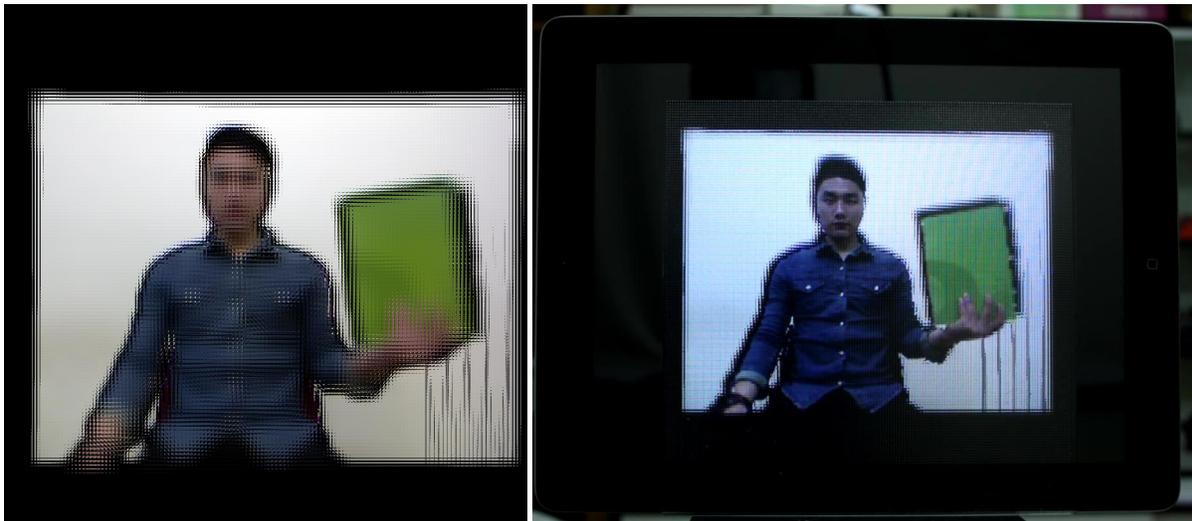


Figure 7 Final integral image's result from the pre-process (left) and our implemented result into the integral imaging monitor (right). It shows different perspective and parallaxes through our experimental system effectively. Note that the pinhole array's reference position was just behind of the human model.

ACKNOWLEDGEMENTS

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].

REFERENCES

- [1] M. Martínez-Corral and G. Saavedra, "The Resolution Challenge in 3D Optical Microscopy", *Progress in Optics* **52**, 1-67 (2009)
- [2] M. Weber, M. Mickoleit, and J. Huisken, "Lightsheet microscopy," *Methods in cell biology* **123**, 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.* **61**, 1295-304 (2014)
- [4] J. Geng and J. Xie, "Review of 3-D Endoscopic Surface imaging techniques," *IEEE Sensors Journal* **14**, 945-960 (2014)
- [5] 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* **17**, 572-577 (2013)
- [6] U. Perwass and C. Perwass, "Digital imaging system, plenoptic optical device and data processing method," Patent WO2010121637 (A1)
- [7] T. G. Georgiev and A. Lumsdaine, "Methods and apparatus for rich image capture with focused plenoptic cameras," Patent US8345144 B1.
- [8] H. Peter, M. Krainin, E. Herbst, X. Ren, D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments", *The International Journal of Robotics Research* **31**, 647-663 (2012)
- [9] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR) art. 6092378, 127-136 (2011).
- [10] J. Shotton, T. Sharp, A. Kipman, A. Fitzkipman, M. Finocchio, A. Blake, M. Cook, R. Moore, "Real-time human pose recognition in parts from single depth images", *Communications of the ACM* **56**, 116-124 (2013).
- [11] D. Hao, P. Henry, X. Ren, M. Chang, Dan B. Goldman, Steven M. Seitz, D. Fox, "Interactive 3D modeling of indoor environments with a consumer depth camera", *Proceedings of the ACM conference on Ubiquitous computing*, 75-84 (2011).
- [12] J. Biswas, M. Veloso, "Depth camera based indoor mobile robot localization and navigation", *IEEE International Conference on Robotics and Automation*, 6224766, 1697-1702 (2012).
- [13] G. Lippmann, "Epreuves reversibles donnant la sensation du relief," *J. Phys.* **7**, 821-825 (1908).
- [14] Microsoft Kinect, <http://www.xbox.com/en-us/kinect/>
- [15] S. Hong, Y. Tan, H. Yeo, B. Lee, "1-inch UniTouch System using Kinect", *International Conference on Signal Processing, Image Processing and Pattern Recognition (ICSIPR)* 6497994, 351-355 (2013).
- [16] Barak Freedman, Alexander Shpunt, Meir Machline, Yoel Arieli, "Depth Mapping Using Projected Patterns", U.S. Patent 20080240502A1, issued October 2, 2008.
- [17] Minkyu Lee, Jaebong Jeon, "Personal Computer Control Using Kinect", *The Korean Institute of Information Scientists and Engineers* **39**, 1 (2012).
- [18] Kinect SDK, <http://msdn.microsoft.com/en-us/library/hh855347.aspx>
- [19] M. Martínez-Corral, B. Javidi, R. Martínez-Cuenca and G. Saavedra, "Formation of real, orthoscopic integral images by smart pixel mapping," *Optics Express* **13**, 9175-9180 (2005)