

Regeneration of elemental images in integral imaging for occluded objects using a plenoptic camera

Min-Chul Lee¹, Kotaro Inoue¹, Cheol-Su Kim², and Myungjin Cho^{3,*}

¹Department of Computer Science and Electronics, Kyushu Institute of Technology, Fukuoka 820-8502, Japan

²Department of Electrical Energy and Computer Engineering, Gyeongju University,
188 Taejongro, Gyeongju City, KyeongsangBukdo, 38065, Republic of South Korea

³Department of Electrical, Electronic, and Control Engineering, IITC, Hankyong National University,
327 Chungang-ro, Anseong-si, Gyonggi-do 456-749, Republic of South Korea

*Corresponding author: mjcho@hknu.ac.kr

Received July 12, 2016; accepted October 28, 2016; posted online November 22, 2016

In this Letter, we propose an elemental image regeneration method of three-dimensional (3D) integral imaging for occluded objects using a plenoptic camera. In conventional occlusion removal techniques, the information of the occlusion layers may be lost. Thus, elemental images have cracked parts, so the visual quality of the reconstructed 3D image is degraded. However, these cracked parts can be interpolated from adjacent elemental images. Therefore, in this Letter, we try to improve the visual quality of reconstructed 3D images by interpolating and regenerating virtual elemental images with adjacent elemental images after removing the occlusion layers. To prove our proposed method, we carry out optical experiments and calculate performance metrics such as the mean square error (MSE) and the peak signal-to-noise ratio (PSNR).

OCIS codes: 110.0110, 110.6880.

doi: 10.3788/COL201614.121101.

Integral imaging, which was first proposed by Lippmann in 1908^[1], has been used to develop next-generation three-dimensional (3D) imaging and display techniques. To obtain and visualize 3D images, two main steps are required: pickup and reconstruction. In pickup, rays from 3D objects can be captured through lenslet array on an image sensor such as charge-coupled device (CCD). These captured rays are multiple two-dimensional (2D) images with different perspectives for 3D objects, which are referred to as elemental images. In the reconstruction or display stage, these elemental images are printed or displayed on a display device, such as a liquid crystal display (LCD), through the homogeneous lenslet array used in pickup so 3D images can be observed without special viewing glasses. Integral imaging does not require the coherent light source that is used in holography. In addition, it provides full color, full parallax, and continuous viewing points of 3D objects. Especially, it provides depth information of 3D objects using a passive imaging system. Therefore, it can be applied to occlusion removal techniques for 3D objects^[2-11].

Since integral imaging can obtain multi-view information of 3D objects, a depth map may be generated by varying the ratios of the viewing point and elemental images. Thus, the occlusion may be removed by classifying between objects and occlusion layers using the depth map and elemental images. However, this method has two main problems. The first problem is the resolutions of the elemental images and depth map are very low due to the lenslet array based integral imaging. Another problem is the information of elemental images may be lost for the occlusion removal.

In this Letter, to solve these problems, we propose an elemental image regeneration method of 3D integral imaging for occluded objects using a plenoptic camera. A plenoptic camera, which is a modified version of an integral imaging system, can record a light field (location and direction of rays) by placing the main imaging lens in front of the lenslet array. It can capture a depth map with high resolution and an all-in-focus image in a single shot, and it can simplify the process for the conventional occlusion removal technique. To record elemental images with high resolutions, in this Letter, we used synthetic aperture integral imaging (SAII)^[12] by Jang *et al.*

SAII can capture an elemental image with the same resolution as the one of the image sensor by replacing the lenslet array with a camera array for to improve the resolution of elemental images. Finally, the cracked parts of the elemental images may be interpolated from adjacent elemental images to enhance them. Since elemental images have multi-view information, it is possible to interpolate the cracked parts from adjacent elemental images, which can be carried out by inverse computational integral imaging reconstruction (CIIR)^[2,13-16].

First, we present our proposed method. A light field is a vector function presented for the position and angle of rays. In general, a light field can be defined as 5 dimensions that consist of 3D space coordinates and 2D angles. This is referred to as 5D light field. However, light intensity is invariant in an optical system, according to the brightness invariant principle. Therefore, a 5D light field can be redefined as a 4D light field. A plenoptic camera records this 4D light field so it can adjust the position of the focal plane for the image or estimate the depth map.

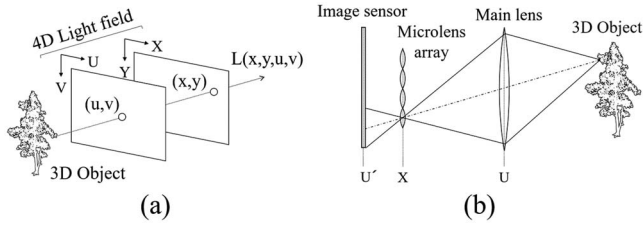


Fig. 1. 4D light field function. (a) Overview and (b) plenoptic camera.

A 4D light field function is shown in Fig. 1(a). It is rays from the objects, which can be recorded as the intersection points of two 2D planes. That is, $L(x, y, u, v)$ are coordinates (x, y) and (u, v) on the XY and UV planes, respectively. They are the same as the intersection coordinates on one of the 2D planes and the angles of two axes. The concept of a 4D light field for a plenoptic camera is illustrated in Fig. 1(b). The difference between a plenoptic camera and a conventional camera is the lenslet array placed in front of the main imaging lens. In a conventional camera, rays can be recorded on the coordinates of the image sensor (i.e., 2D information). However, in a plenoptic camera, the intersection coordinates of rays on two planes can be found by imaging object rays through both the main lens and the lenslet array.

Plenoptic cameras can reconstruct the image focused at a certain position by recording the 4D light field. This technique is called refocusing, which can be implemented by shifting the virtual image sensor plane. It is very simple and is carried out by shifting and averaging sub-aperture images, as shown in Fig. 2^[17]. It is similar to CIIR, but its equations are different because it uses the light field function.

For simplicity, let us consider the movement of 2D virtual image sensor plane X, as shown in Fig. 3. The X plane is placed a distance F from the U plane, and light field $L(x, u)$ is rays passing through the u coordinate of the U plane and the x coordinate of the X plane. For refocusing, when the X plane is moved to the X' plane, where it is a distance F' from the U plane, the recorded light field L' on the X' plane can be described as the movement of the x coordinate on L . When the expanding coefficient is $\alpha = F'/F$, L can be written as follows^[18]:

$$L'(x', u) = L\left(u + \frac{x' - u}{\alpha}, u\right). \quad (1)$$

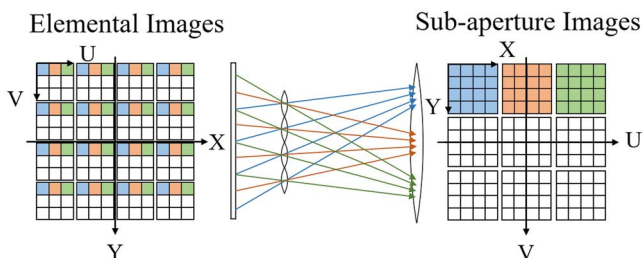


Fig. 2. Sub-aperture images.

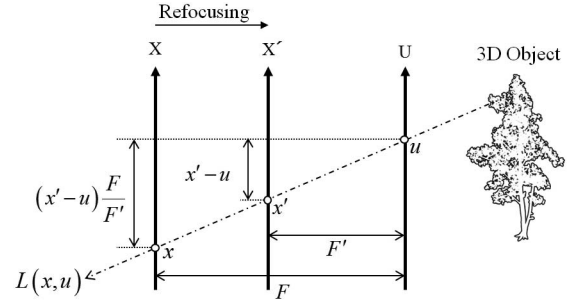


Fig. 3. Refocusing.

This equation can be extended as a 4D light field:

$$L'(x', y', u, v) = L\left(u + \frac{x' - u}{\alpha}, v + \frac{y' - v}{\alpha}, u, v\right). \quad (2)$$

Moving the virtual image sensor plane is the same as moving the position of the XY plane in the recording coordinates of the light field. It is well known that a 2D image can be transformed from a 4D light field by integrating the light field. Therefore, the image at distance F' , $E_{F'}$ can be described as follows:

$$E_{F'}(x', y') = \frac{1}{F'^2} \iint L'(x', y', u, v) du dv. \quad (3)$$

By substituting Eq. (2) into Eq. (3), we can obtain the following equation:

$$E_{F'}(x', y') = \frac{1}{\alpha^2 F'^2} \iint L\left[u\left(1 - \frac{1}{\alpha}\right) + \frac{x'}{\alpha}, v\left(1 - \frac{1}{\alpha}\right) + \frac{y'}{\alpha}, u, v\right] du dv. \quad (4)$$

From Eq. (4), it can be seen that the image can be reconstructed by shifting and averaging the x, y coordinates of light field L and expanding the image with expanding coefficient α . That is, when the u th column and v th row sub-aperture image is $L(u, v)$, the 2D image can be transformed by shifting and averaging sub-aperture images instead of elemental images in integral imaging. Therefore, Eq. (4) can be rewritten^[18]

$$E_{F'}(x', y') = \frac{1}{\alpha^2 F'^2} \iint L^{(u,v)}\left[u\left(1 - \frac{1}{\alpha}\right) + \frac{x'}{\alpha}, v\left(1 - \frac{1}{\alpha}\right) + \frac{y'}{\alpha}\right] du dv. \quad (5)$$

The depth map can be estimated using the light field. This function is included in Lytro software, but the algorithm has not been opened to the public. Thus, in this Letter, we present our own depth map estimation. The depth map has a 16-bit grayscale and its brightness is determined by the Λ used for refocusing. Λ_{Min} is brightness 0, and Λ_{Max} is brightness $2^{16}-1$.

Table 1. Definition of Parameters

Parameter	Definition
C	Size of image sensor
D	Depth maps
EI	Elemental image
M	Movement of virtual object
N	Number of pixels for each elemental image
OL	Occlusion layer
OREI	Occlusion removed elemental image
REI	Regenerate elemental image
S_g	Number of shifted pixels for regeneration
S_r	Number of shifted pixels for reconstruction
Th	Threshold for occlusion layer
VEI	Virtual elemental image
d	Reconstruction distance
f	Focal length of image sensor
p	Moving gap between image sensor

To estimate the physical distance from these Lambdas, the calibration process is required because the brightness of the depth map is different from the physical distance.

Regeneration of elemental images has two main stages: occlusion removal, and interpolation of cracked parts caused by occlusion removal. Table 1 describes the system parameters and their definition, and the regenerated images are shown in Fig. 4.

Occlusion removal can be implemented by using a certain threshold of the depth map. Let the depth map be D and the occlusion layer be OL. Then, OL can be written as

$$OL_{(k,l)}(x, y) = \begin{cases} 1, & D_{(k,l)}(x, y) \geq Th \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

where $OL(k, l)$ is the k th column and the l th row occlusion layer, x, y give the pixel position, and Th is the threshold

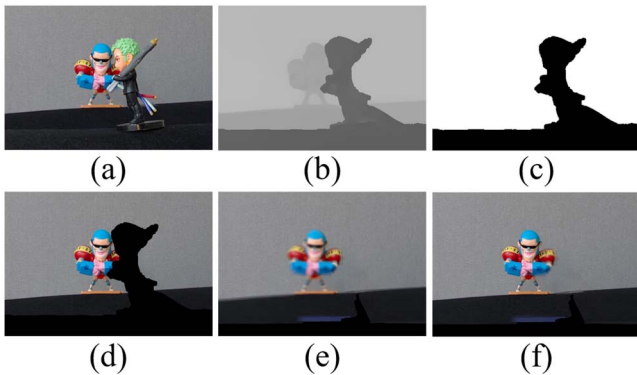


Fig. 4. Image of equation ($k = 3, l = 3$): (a) EI, (b) D , (c) OL, (d) OREI, (e) VEI, and (f) REI.

value for occlusion removal. Then, occlusions can be removed from the elemental images. Let the elemental image be EI and elemental image with occlusions removed be OREI. OREI is written as

$$OREI_{(k,l)}(x, y) = \begin{cases} EI_{(k,l)}(x, y), & OL_{(k,l)}(x, y) = 1 \\ 0, & \text{otherwise} \end{cases}. \quad (7)$$

Since the elemental image has a lot of zero brightness after occlusion removal, the visual quality of the elemental image may be degraded. Thus, in this Letter, the elemental image is interpolated and regenerated by using adjacent elemental images. Regeneration can be carried out by shifting adjacent elemental images and generating the virtual elemental image.

OREI can be split by intensities of the depth map, as follows:

$$\begin{aligned} & \text{Sliced}_{(k,l)}(x, y, t) \\ &= \begin{cases} OREI_{(k,l)}(x, y), & D_{(k,l)}(x, y) = t \\ 0, & \text{otherwise} \end{cases} \quad (Th < t \leq t_{\max}), \end{aligned} \quad (8)$$

where t is the intensity of the depth map, and t_{\max} is the maximum intensity of the depth map. Then, the movement of elemental images is calculated as depicted in Fig. 5(a). Let (k, l) be the coordinates of the current regenerated image and (m, n) be the coordinates of the image for interpolation. The movements of the elemental image in the x and y directions, M_x and M_y , are written by

$$M_x(k, m) = (k - m)p, \quad M_y(l, n) = (l - n)p, \quad (9)$$

where p is the distance between cameras for SAII. Using these movements, shifting pixels for each elemental image as shown in Fig. 5(b) can be written as follows:

$$\begin{aligned} Sg_x(k, m, t) &= \frac{EI_x f M_x(k, m)}{c_x z(t)}, \\ Sg_y(l, n, t) &= \frac{EI_y f M_y(l, n)}{c_y z(t)}, \end{aligned} \quad (10)$$

where EI_x and EI_y are the number of pixels for the image sensor in the x and y directions, and $z(t)$ is the function transformed by the intensity of the depth map in relation

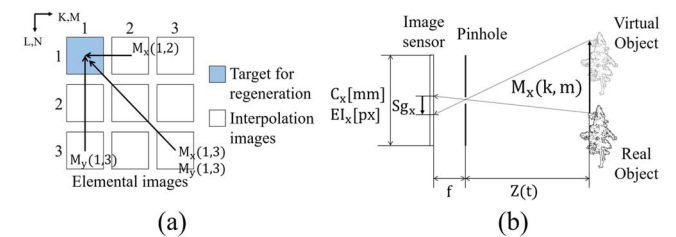


Fig. 5. Overview of the algorithm. (a) Movement of elemental images and (b) shifted pixels for regeneration.

to the physical distance. This function depends on the specifications of the plenoptic camera and the calibration methods. Thus, the virtual elemental image VEI can be written as

$$\begin{aligned} \text{VEI}_{(k,l)}(x, y) &= \frac{1}{O(x, y)} \sum_{t=\text{Th}+1}^{t_{\max}} \sum_{m=1}^K \sum_{n=1}^L \text{Sliced}_{(m,n)}(x + \text{Sg}_x(k, m, t), y \\ &\quad + \text{Sg}_y(l, n, t), t), \end{aligned} \quad (11)$$

where $O(x, y)$ is the superposition matrix for CIIR. Equation (11) is the inverse of CIIR. Finally, the regenerated elemental image can be obtained from OREI and VEI as follows:

$$\text{REI}_{(k,l)}(x, y) = \begin{cases} \text{OREI}_{(k,l)}(x, y), & \text{OL}_{(k,l)}(x, y) = 1 \\ \text{VEI}_{(k,l)}(x, y), & \text{otherwise} \end{cases}. \quad (12)$$

To prove our proposed method, we carried out computer simulations. The parameters are described in Table 1. In CIIR, the shifting pixels of each elemental image Sr_x and Sr_y are as follows:

$$\text{Sr}_x = \frac{N_x p f}{c_x d}, \quad \text{Sr}_y = \frac{N_y p f}{c_y d}. \quad (13)$$

Finally, the reconstructed 3D image at distance d can be obtained by the following equation:

$$I(x, y, d) = \frac{1}{O(x, y)} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \text{REI}_{(k,l)}(x - k\text{Sr}_x, y - l\text{Sr}_y). \quad (14)$$

Next, we show the experimental results. The depth map from the Lytro software cannot present the physical depth. Thus, we need to place the reference object at a fixed distance and measure the distance by stereo matching. Therefore, we can estimate the relation between the intensity of the depth map and the physical depth. In the pickup stage, LYTRO ILLUM is used, where the resolution of the camera is 2022(H) \times 1404(V), the focal length of camera lens $f = 70$ mm, the refocus range is 400–750 mm, and the distance between cameras $p = 10$ mm. When we know the shifting pixels (S) between the two elemental images, we can calculate the depth d using the following equation:

$$d = \frac{N_x p f}{c_x S} = \frac{2022 \times 10 \times 70}{36 \times S}. \quad (15)$$

Table 2 and Fig. 6 show the measurement results. These depths are converted to the intensity of the depth map t . As shown in Fig. 6, the relation between the intensity of the depth map and the physical depth is linear. Thus, the linear approximation can be found by the least squares

Table 2. Calibration Results

	Near	Middle	Far
S (px)	105	71	57
d (mm)	374.4	553.8	689.8
t	118	149	164

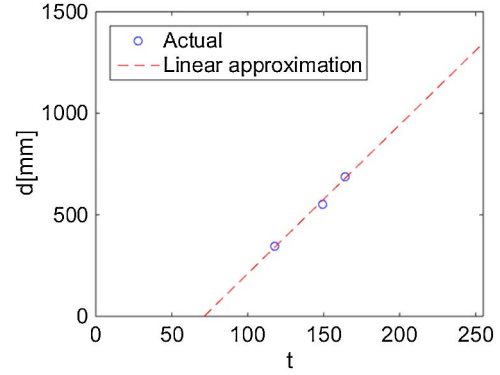


Fig. 6. Linear approximation results.

method. Therefore, the experimental equation for the transformation between the intensity of the depth map and the physical depth is

$$z(t) = 7.33t - 522.1. \quad (16)$$

This equation is used to calculate Sg_x and Sg_y in the regeneration of elemental images.

In our experiment, there are two pickup scenarios: with occlusion and without occlusion. 3D objects without occlusion are used to calculate the mean square error (MSE) and the peak signal-to-noise ratio (PSNR) as follows:

$$\text{MSE} = E[(\text{Ref} - I_{\text{input}})^2], \quad (17)$$

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right), \quad (18)$$

where $E[\cdot]$ is the expectation operator, Ref is the reference image, I_{input} is the reconstruction results, and MAX_I is the maximum pixel intensity of the image. The occlusion is placed in front of the left shoulder of the object.

Figure 7 shows the reconstructed 3D images at the reconstruction distance $d = 620$ mm for each method. The reconstructed 3D images using the elemental images without occlusion shown in Fig. 7(a) are the references for the MSE and PSNR. Figures 7(b) and 7(c) show the reconstructed 3D images using elemental images with the occlusion removed conventionally and the reconstructed 3D images by our proposed method, respectively. As shown in the enlarged figure, the characters “BF-37”

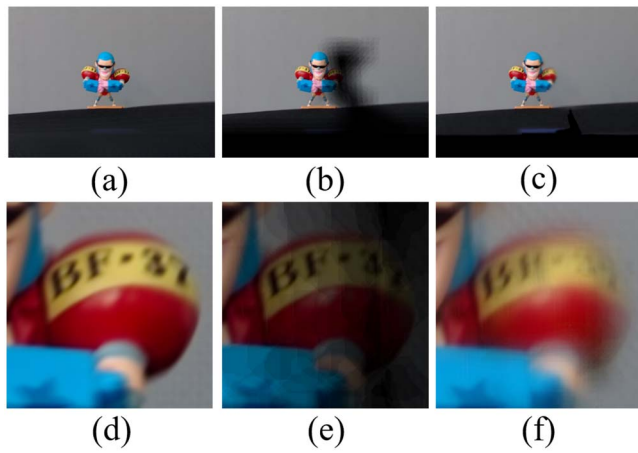


Fig. 7. Experimental results at $d = 620$ mm: (a) original, (b) conventional occlusion removal, (c) proposed method, (d)–(f) enlarged views of (a)–(c), respectively.

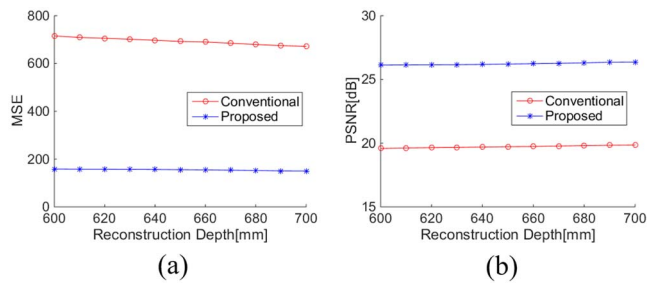


Fig. 8. Quality of image using MSE and PSNR.

on the shoulder of the object can be easily recognized in Fig. 7(f). To evaluate the visual quality of the reconstructed 3D images, we calculate the MSE and PSNR as shown in Fig. 8. Thus, we see our proposed method can obtain better results. (The MSE is improved by 60%, and the PSNR is improved by -7 dB).

In this Letter, we propose a regeneration technique for elemental images in integral imaging using

a plenoptic camera after removing the occlusions. In conventional methods, the image information may be lost after occlusion removal. On the other hand, in our proposed method, the image information can be interpolated by adjacent elemental images after occlusion removal. However, our method has some problems. The visual quality of the regenerated elemental images depends on the accuracy of the depth map or calibration. In addition, our method uses an averaging process for the 3D reconstruction, so high spatial frequencies may be lost. We will look for a solution to these problems in the future.

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2015R1A2A1A16074936).

References

1. G. Lippmann, *C. R. Acad. Sci.* **146**, 446 (1908).
2. S.-H. Hong, J.-S. Jang, and B. Javidi, *Opt. Express* **12**, 483 (2004).
3. S.-H. Hong and B. Javidi, *Opt. Express* **14**, 12085 (2006).
4. M. Cho and B. Javidi, *Opt. Lett.* **33**, 2737 (2008).
5. D.-H. Shin, B.-G. Lee, and J.-J. Lee, *Opt. Express* **16**, 16294 (2008).
6. B.-G. Lee, L. Lilliana, and D.-H. Shin, *Opt. Comm.* **283**, 2084 (2010).
7. M. Zhang, Y. Piao, and E.-S. Kim, *Appl. Opt.* **49**, 2571 (2010).
8. J.-J. Lee, B.-G. Lee, and H. Yoo, *Appl. Opt.* **50**, 1889 (2011).
9. Z. Zhou, Y. Yuan, X. Bin, and Q. Wang, *Chin. Opt. Lett.* **9**, 041002 (2011).
10. Y. Piao, M. Zhang, and E.-S. Kim, *Opt. Lasers Eng.* **50**, 1602 (2012).
11. M. Zhang, Y. Piao, J.-J. Lee, D. Shin, and B.-G. Lee, *Opt. Commun.* **313**, 204 (2014).
12. J.-S. Jang and B. Javidi, *Opt. Lett.* **27**, 1144 (2002).
13. D.-H. Shin, E.-S. Kim, and B. Lee, *Appl. Phys.* **44**, 8016 (2005).
14. B. Javidi, R. Ponce-Díaz, and S.-H. Hong, *Opt. Lett.* **31**, 1106 (2006).
15. D.-H. Shin and H. Yoo, *Opt. Express* **15**, 12039 (2007).
16. H. Yoo and D.-H. Shin, *Opt. Express* **15**, 14107 (2007).
17. R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, in *Computer Science Technical Report CSTR 2.11*, p. 1 (2005).
18. R. Ng, "Digital light field photography," Ph. D. dissertation (Stanford University, 2006).