DOI 10.1007/s10946-020-09891-9
Journal of Russian Laser Research, Volume 41, Number 4, July, 2020

3D INTEGRAL IMAGING DISPLAY PROCESSING USING THE SIMILARITY OF CORRESPONDING POINTS IN AXIALLY RECORDED IMAGES

Jin-Xiao Yang and Yu Wang*

School of Electronics and Information Engineering
Changchun University of Sciences and Technology
Changchun 130022, China

*Corresponding author e-mail: wangyulfy@cust.edu.cn

Abstract

In this paper, we propose a three-dimensional (3D) integral image display method that uses the similarity of corresponding points in a series of axially recorded images. First, we calculate the corresponding points of a 3D object, in view of the proportional relationship of the distances of points in the elemental images collected, using axially distributed sensing (ADS). Next, we extract the depth map using the minimum error of the color values of the corresponding points. Finally, we use the color image and depth map to generate an elemental image array without the need for a lens array. This approach can display a 3D image in integral imaging. To show the usefulness of the method, we obtain the elemental images using ADS through the 3ds Max, and the experimental results demonstrate that the method proposed can extract a depth map for the 3D integral image display.

Keywords: corresponding points, ADS, elemental images, depth map.

1. Introduction

Recently, various three-dimensional (3D) images, games, and videos have increasingly become available to the public. Three-dimensional imaging and display technology [1], which is the next-generation display technology actively being developed internationally, is gradually becoming a compelling medium. At present, most 3D display technologies are based on parallax or are real 3D display technologies such as integral imaging [2], holographic 3D display [3], and volumetric 3D display [4]. In contrast to holographic and volumetric 3D display technologies, which have limited space–time resolution, integral imaging can present a full-color image with continuous parallax over a wide field of view, and it is relatively less restricted by the display device. This approach has good potential, but there remains much room for improvement, and it has hence become a popular research topic [5–7].

Integral imaging is a glass-free 3D display technology that uses microlens arrays to capture and display 3D information [8]. Here, the recorded images are called elemental images, and the elemental image array is used to display the 3D scene. According to the recording method, integral imaging can be divided into microlens array, synthetic aperture [9, 10], axial distributed sensing (ADS) [11, 12], and off-axis distributed sensing [13] integral imaging.

In ADS, longitudinal perspective information is obtained by translating a camera along its optical axis [14]. Using this longitudinal information, the depth information of the object can be extracted,

Manuscript submitted by the authors in English on May 25, 2020.

which can be used to generate the elemental image array for 3D display. Hong et al. [14] proposed a method for extracting the depth of 3D objects using ADS. In this method, the camera is moved along the optical axis to record high-resolution elemental images that are then used to generate slice images of 3D objects using a computational reconstruction algorithm based on the ray back-projection. Then, to extract the depth of the 3D objects, they perform a simple block comparison algorithm between the first elemental image and set of 3D slice images. This method can provide the depth image of a 3D object. This depth image can then be used to display 3D information of the object.

In this paper, we propose a depth extraction method based on the similarity of corresponding points in axially collected high-resolution elemental images and then combine them with the corresponding color image to generate the elemental image array for integral imaging 3D display. This method of image processing works when the perpendicular distance of the object from the optical axis is much smaller the total distance moved by the camera. Moreover, to obtain the elemental images of real 3D objects, the lens should be carefully corrected for off-axis aberrations.

2. Depth Extraction Using the Corresponding Points in Axially Recorded Images

The method for axially recording multiple images is as follows. The camera moves along the direction of the optical axis of the camera at a certain distance from the object and collects multiple high-resolution elemental images at fixed intervals Δz. A schematic diagram of the light field acquisition system for ADS is shown in Fig. 1. In this system, only one camera sensor is required. This simple movement extends the ranging ability of an existing single-camera system. As shown in Fig. 1, the camera is farthest from the 3D object in the initial position. As the camera moves closer to the 3D object along the optical axis, the elemental images obtained have different magnification ratios. These differences in the magnification ratio of the recorded 2D images are useful for estimating or extracting the depth information of the 3D object [11].

The light from a captured target point falls on the same pixel in two adjacent elemental images, if the camera's stepwise movements are less than a certain value. In this case, the relative position of the object point in two adjacent elemental images is zero, hence the parallax information is zero, and this makes depth extraction impossible. Therefore, according to [11], the minimum step of the camera in an

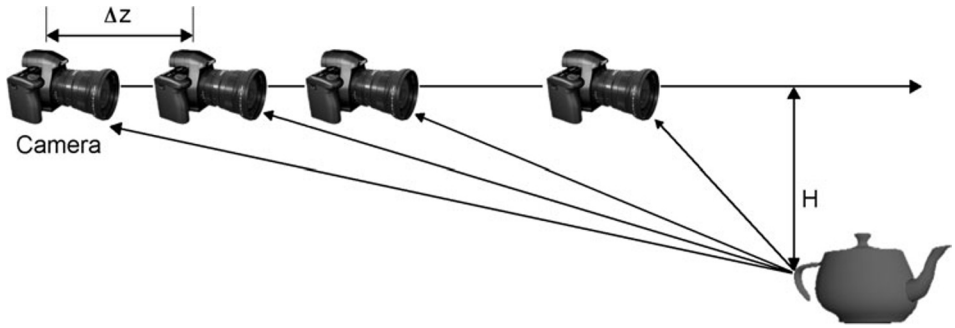

Fig. 1. Light field acquisition for ADS.

ADS acquisition system can be calculated as follows:

$$\min \Delta z_i = \frac{\mu z_i^2}{\mu z_i + Hg}\,. \tag{1}$$

We assume the perpendicular distance from the point of the 3D object to the camera's optical axis is H, where z_i denotes the axial distance from the ith camera sensor to the target object in space, g is the focal length of the camera, μ represents the sensor pixel size, and "min" means the minimum distance between adjacent camera sensors, that is, the minimum step distance of the camera in the axial direction, at which the parallax information of the object points can be recorded by the adjacent elemental images.

In an ADS system, according to the principle of lens imaging, any object point in a real scene is imaged by a lens, and a series of projection points formed by the object point can be obtained on the axially distributed elemental image. Moreover, the 3D light field information of the object is included in this series of projection points. The projection points obtained by imaging the same object point through the lens are called the corresponding points, as indicated in Fig. 2 by the gray dot on the camera sensor. The calculated color values of the corresponding points in each elemental image are highly similar only when the object point is at the correct depth position, that is, the statistical error of the color values of the corresponding points is the smallest. According to this principle, for different depth values within the depth range of a 3D object, the statistical error of the color value of a corresponding point in each elemental image can be calculated separately. The depth corresponding to the minimum error value is the depth of the object point. We can obtain a depth map of a 3D object using this method.

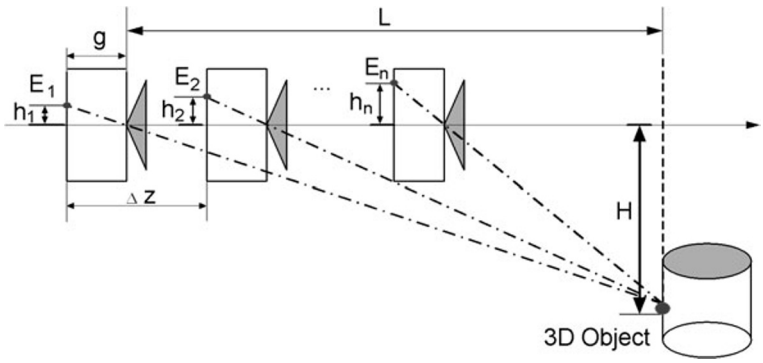

Fig. 2. Corresponding points in axially recorded multiple images.

As shown in Fig. 2, the camera is at a certain distance from the 3D object and gradually moves closer to the object in increments of Δz along the direction of the optical axis to acquire the elemental images. We define the focal length of the camera as g. Then, the camera moves $(N-1)$ times in the axial direction, and records N elemental images as E_n $(n = 1, 2, \ldots, N)$. When a 3D object point is located on the camera's optical axis, the depth information of that point cannot be obtained from those N elemental images, so the object should have a certain perpendicular distance from the camera's optical axis [11]. We denote the perpendicular distance from the point of the 3D object to the camera's optical axis as H and the distance from the initial position of the camera as L $(L > H)$. In the nth elemental image E_n collected along the optical axis, the perpendicular distance between the corresponding points and the image center of the object point is h_n, and the following expressions can be derived from the triangular relationship shown in Fig. 2:

$$\frac{h_1}{H} = \frac{g}{L}\,, \tag{2}$$

$$\frac{h_2}{H} = \frac{g}{L - \Delta z}\,, \tag{3}$$

$$\frac{h_n}{H} = \frac{g}{L - (n-1)\Delta z}\,. \tag{4}$$

From Eqs. (2) and (3), the following relationship between the distance h_1 of the corresponding point of the first elemental image and the distance h_2 of the corresponding point of the second elemental image at an interval of Δz can be obtained:

$$\frac{h_1}{h_2} = \frac{L - \Delta z}{L}. \tag{5}$$

In a similar way, from Eqs. (2) and (4), we have

$$\frac{h_1}{h_n} = \frac{L - (n-1)\Delta z}{L}. \tag{6}$$

The distance from the corresponding point to the optical axis in the elemental image collected by the ADS has a proportional relationship, as Eq. (6) describes. Therefore, using the distance relationship between the corresponding points in the axial acquisition, the distributions of the positions of the other image points in the elemental image can be obtained. Moreover, from Eq. (6), we know that the distance is related to the distance L from the object to the camera, that is, the depth of the object, and this feature can be used to search for object depth information.

The pixel coordinates of the first elemental image are denoted as x_{E_1} and y_{E_1}, and the captured elemental image size is $p \times q$. Then, for the corresponding points x_{E_n}, y_{E_n}, in the remaining $(N-1)$ elemental images E_n $(n = 2, 3, \ldots, N)$ can be formulated as follows:

$$x_{E_n} = \frac{p}{2} - \left(\frac{p}{2} - x_{E_1}\right) \frac{L}{L - (n-1)\Delta z}, \tag{7}$$

$$y_{E_n} = \frac{q}{2} - \left(\frac{q}{2} - y_{E_1}\right) \frac{L}{L - (n-1)\Delta z}. \tag{8}$$

To obtain the depth of a point, we need the minimum sum of absolute value difference (SAD) value of the color of the corresponding points for the pixel in elemental image E_1. Using Eqs. (7) and (8), a series of corresponding points for any pixel in image E_1 can be calculated. This is because only at the correct depth position of an object point, the colors of the corresponding points in each elemental image are highly similar [15], that is, the statistical error of the color values of the corresponding points is the smallest. Therefore, we use the SAD to calculate the error in color between the corresponding points. The SAD is a local image matching metric that is often used for image block matching to evaluate the similarity of image blocks. We denote the distance between an object point and the initial position of the camera as L, which is the absolute value of the projection point of the object point in E_1 and the corresponding point in E_n $(n = 2, 3, \ldots, N)$. The sum of the value differences is expressed as SAD_L, and it is calculated as follows:

$$\mathrm{SAD}_L = \sum_{i=1}^{b} \sum_{j=1}^{b} \left(\sum_{n=2}^{N} \left| E_1(x+i, y+j) - E_n(x'+i, y'+j) \right| \right). \tag{9}$$

Here, $b \times b$ is the matching window size, N is the number of elemental images collected by the ADS system, and $E_n(x', y')$ is the color of the point corresponding to $E_1(x, y)$.

For different values of L in a 3D object depth range, each SAD_L is calculated. The value of L with the smallest SAD value is the depth $\hat{L}$ of the object point; it reads

$$\hat{L}(x, y) = \arg\min_{L} \mathrm{SAD}_L(x, y). \tag{10}$$

This operation is repeated for each pixel in elemental image E_1 to obtain the depth map corresponding to E_1.

Because the corresponding points algorithm is mainly discussed as a theoretical model, it assumes the light propagates along a straight line. Hence, the method will not work well for objects that have color components that depend on angle.

3. Pixel Mapping to Generate the Elemental Image Array

In this section, we describe how we use a pixel mapping algorithm [16] to generate the elemental image array for the 3D display from the calculated depth map and color image. First, we use the camera to extract the depth value and color information of each pixel on the surface of the object. Then, a computer virtual synthesis method based on ray optics is used to obtain the corresponding elemental image array. In the process of virtual synthesis, to avoid the problem of depth inversion, the depth map obtained is first inverted before it is synthesized. Figure 3 shows the pixel mapping process during the generation of the elemental image array. The rays from the 3D object point pass through the center of each microlens array, and then they are recorded in the corresponding pixels in the elemental image array. The corresponding pixel coordinates are given by the following relationships:

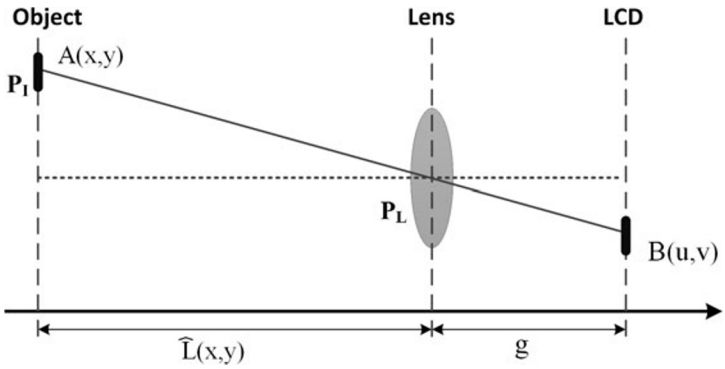

Fig. 3. Elemental image array generation.

$$u = P_L x_L - (xP_I - P_L x_L)\,\frac{g}{\hat{L}'(x,y)}\,, \tag{11}$$

$$v = P_L y_L - (yP_I - P_L y_L)\,\frac{g}{\hat{L}'(x,y)}\,. \tag{12}$$

Here, $\hat{L}'(x,y)$ is a version of $\hat{L}(x,y)$ in Eq. (10) rescaled to meet the requirements of the 3D images in the display space, g is the focal length of the lenslet, and u and v are the pixel indices corresponding to the elemental image array. Furthermore, x_L and y_L are the index values corresponding to the camera coordinate system. The size of the lens array is P_L, and P_I is the size of the pixels in the object image.

4. Experimental Results

To demonstrate the method proposed, we carried out an experiment. The axial image acquisition in this experiment was completed using 3DS Max, where a simulated camera was moved along the optical axis in a one-dimensional direction to obtain the elemental images. We used a simulation camera with a resolution of 3,872×2,592 pixels. The imaging lens with a focal length $g = 40$ mm was used in this experiment. The experimental scene consisted of a teapot and a ring, and the two models were 150 mm apart from each other and 350 and 500 mm away from the first camera. The experimental setup is shown in Fig. 4. In the ADS process, the objects should not be on the optical axis because it is difficult to collect

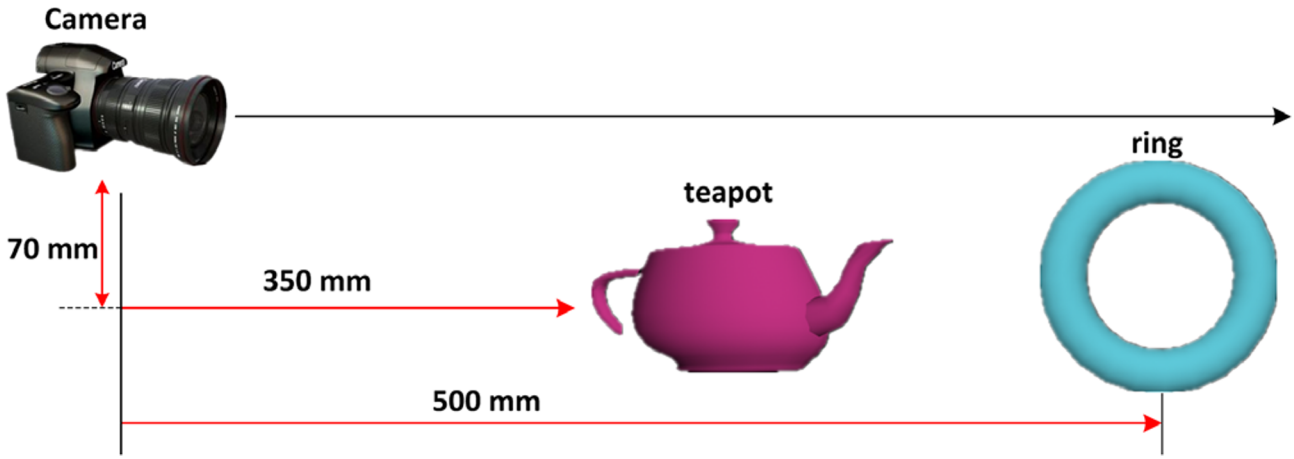

Fig. 4. Experimental setup.

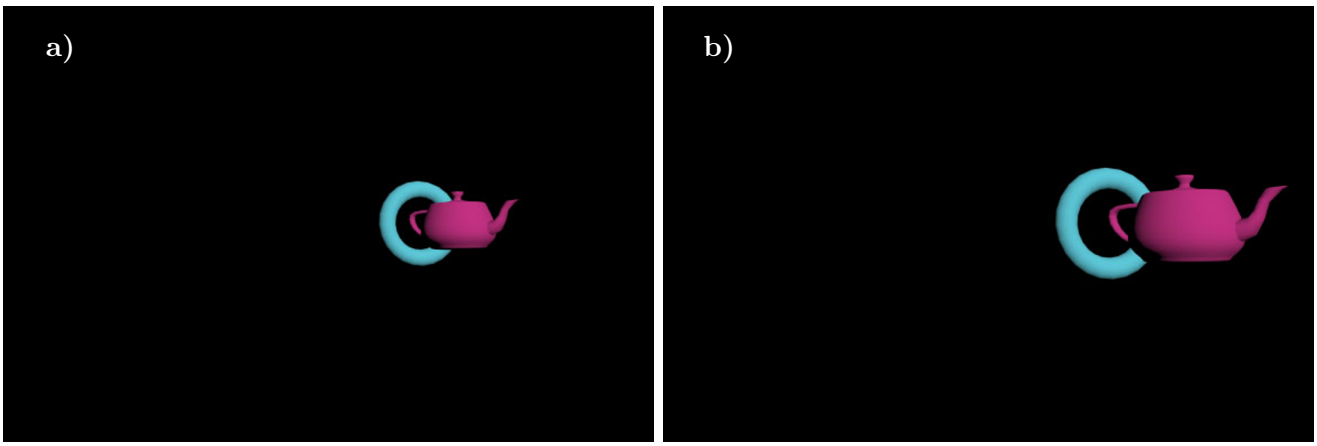

Fig. 5. Examples of elemental images for E_1 (a) and E_{41} (b).

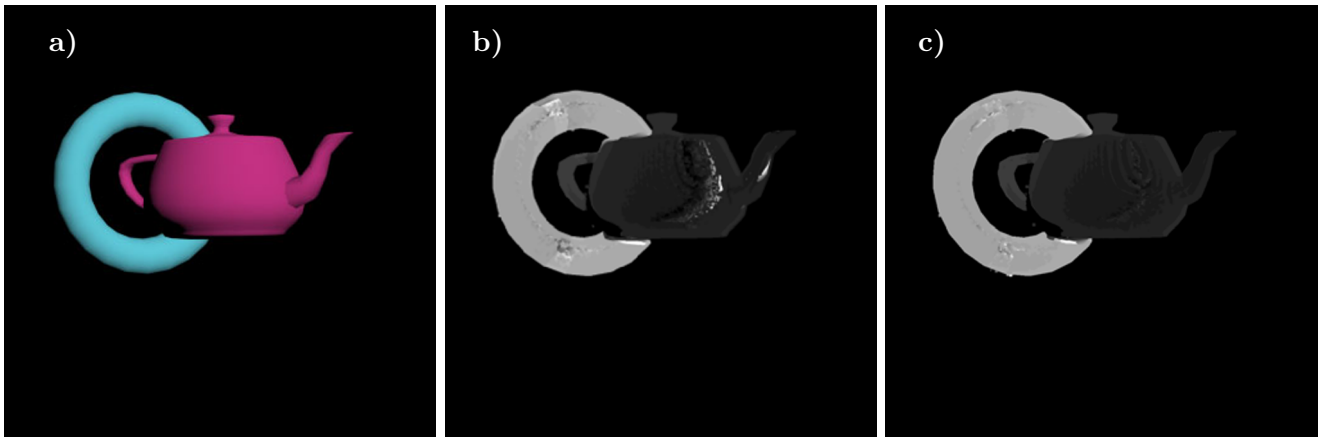

Fig. 6. Color image (a), the depth map obtained from the method of [14] (b), and the depth map obtained from the method proposed in this paper (c).

perspective information from 3D objects near the optical axis [11], so the 3D objects were approximately 70 mm from the optical axis in this setup.

We used Eq. (1) to calculate the minimum step distance of the camera movement and obtained $\Delta z = 3$ mm for this experiment. A total of $N = 41$ elemental images were collected along the optical axis with a total displacement distance of 120 mm. Some examples of recorded elemental images are shown in Fig. 5.

The elemental images collected by the moving camera had a matching window size of 8×8, and the search range of the object depth L ranged from 300 to 600 mm in steps of 10 mm. The depth map obtained by the method in [14] for the image Fig. 6 a is shown in Fig. 6 b.

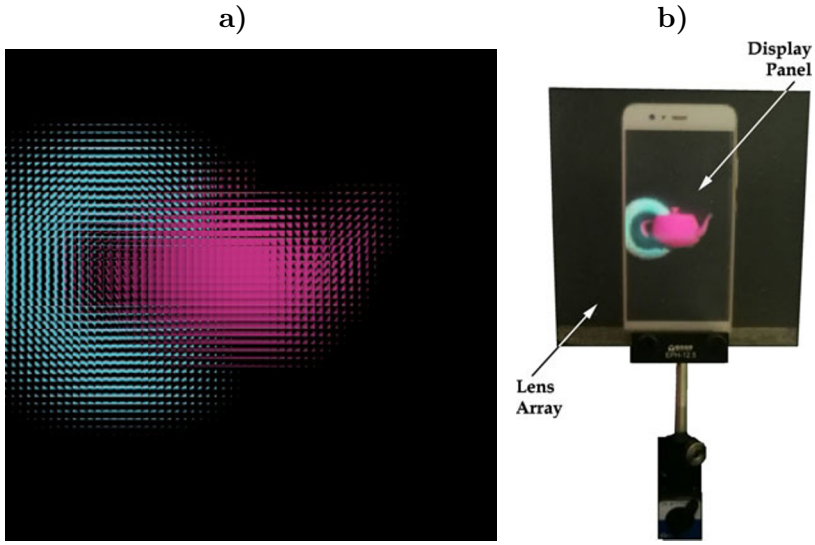

Fig. 7. Elemental image array (a) and the display setup (b).

We used the same elemental images and, for a certain pixel point in the first elemental image E_1, calculated the depth of the point according to the minimum SAD of the color of the corresponding points. We used a search range for object depth L of 300 – 600 mm. The accuracy of the extracted depth information depends on the moving step. According to the formula for determining the minimum moving step of L in [11], the steps of L in this experiment should be 10 mm. The SAD calculation used an error window of size 8×8. We obtained the corresponding points in the elemental images from Eqs. (7) to (8). The color error SAD_L of the corresponding points at different depths was calculated by Eq. (9). Then, for different values of L in the depth range of a 3D object, the L value with the smallest SAD value was used as the depth of the point to obtain elemental image E_1. The elemental image E_1 is shown in Fig. 6 a, and the corresponding depth map is shown in Fig. 6 c. The quality of the depth maps extracted using the proposed method is notably better than those obtained by the method described in [14].

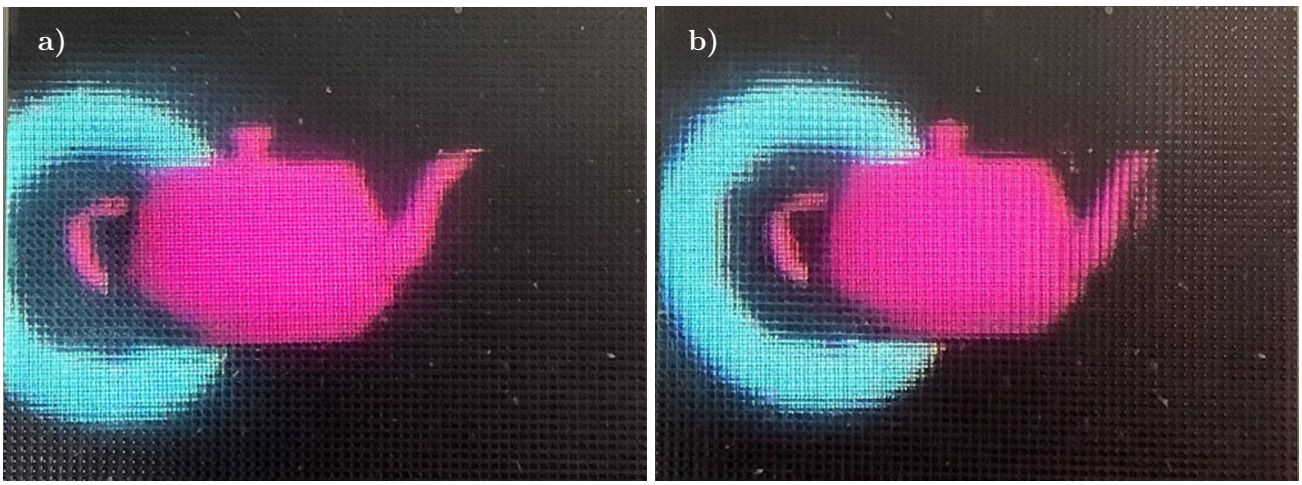

Fig. 8. Experimental result; here, left view (a) and right view (b).

To synthesize the elemental image array, we extracted the depth map and color image. The elemental image array consists of 60 rows and 60 columns of unit images, and the resolution of each unit image is 40×40 pixels, as shown in Fig. 7 a.

To display 3D images, we carried out an optical experiment using the elemental image array shown in Fig. 7 a. The setup is shown in Fig. 7 b, where we used a square microlens array with a focal length of 3 mm and a width of 1 mm to display the image on an LCD display with a pixel size of 0.06 mm. The distance between the microlens array and the elemental image plane is approximately 3.5 mm. As shown in Fig. 8, the change in viewing the angle change can clearly be seen. This demonstrates that the method proposed in this paper can be used for integrated imaging display based on axial viewpoint image acquisition.

5. Summary

In this paper, we presented a depth extraction method using the error statistics of the color values of a series of corresponding points obtained using the longitudinal parallax information of multiple recorded high-resolution elemental images. In the method proposed, we considered the spatial variation in pixel intensity in the local area around the object point. Extracting the depth information using the color similarity of a single pixel alone generates noise. Therefore, we combined this approach with a SAD local window matching algorithm, where the color similarity in a local window around the pixel is used to search for the depth information of the point, which enhances the robustness of the algorithm and can lead to an accurate depth map. Using the acquired coordinate information of the depth map and color map, we generated an elemental image array according to the mapping relationship of each pixel through the lens and optically displayed it using a microlens array and LCD display.

The experimental results show that the method proposed in this paper can generate an elemental image array for displaying integrated imaging based on ADS.

Acknowledgments

This work was supported by the 13th Five-Year Plan Scientific and Technological Research Project of the Education Department of Jilin Province under Grant No. JJKH20200779KJ. We thank Kimberly Moravec, PhD from Liwen Bianji, EDANZ Editing China (www.liwenbianji.cn/ac) for editing the English text of a draft of this manuscript.

References

1. B. Javidi and A. Murat Tekalp, *Proc. IEEE*, **105**, 786 (2017).
2. Y. Piao, M.Zhang, D. Shin, and H. Yoo, *Opt, Lett.*, **38**, 3162 (2017).
3. P. A. Savas Tay, R. Blanche, A. V. Voorakaranam, et al., *Nature*, **451**, 694. (2008).
4. G. E. Favalora, *Computer*, **38**, 37 (2005).
5. V. J. Traver, P. Latorre-Carmona, E. Salvador-Balaguer, et al., *IEEE Signal Process. Lett.*, **24**, 171 (2017).
6. M. Cho, H. Yun, K. Inoue, and B. Cho, *Three-Dimensional Imaging, Visualization, and Display*, International Society for Optics and Photonics (2018).
7. Y. Kim, K. Hong, and B. Lee, *3D Res.*, **1**, 17 (2010).
8. A. Stern and B. Javidi, *Proc. IEEE*, **94**, 591 (2006).

9. J. S. Jang and B. Javidi, *Opt. Lett.*, **27**, 1144. (2002).
10. Y. Piao, L. Xing, M. Zhang, and B. G. Lee, *Opt. Lasers Engin.*, **88 Compl.**, 153 (2017).
11. R. Schulein, M. Daneshpanah, and B. Javidi, *Opt. Lett.*, **34**, 2012 (2009).
12. M. LaRosa and B. Javidi, "Experiments with three-dimensional optical microscopy using axially distributed sensing," in: *Proceedings of 11th Euro-American Workshop on Information Optics (WIO 2012), 20–24 August 2012, Quebec City, Quebec, Canada*, IEEE Publ. (2013).
13. X. Xiao, M. Daneshpanah, M. Cho, and B. Javidi, *J. Display Technol.*, **6**, 614 (2010).
14. S. P. Hong, D. Shin, B. G. Lee, et al., *Opt. Express*, **20**, 23044 (2012).
15. D. C. Hwang, D. Shin, S. C. Kim, and E. Kim, *Appl. Opt.*, **47**, D128 (2008).
16. G. Li, K. C. Kwon, G. H. Shin, et al., *J. Opt. Soc. Korea*, **16**, 381 (2012).