

Depth extraction in computational integral imaging based on bilinear interpolation

HUIHUA HUANG, PENG GE*

School of Physics and Opto-electronic, South China University of Technology, Guangzhou, 510640, China

*Corresponding author: scpge@scut.edu.cn

We proposed a method using a merit function to determine the depth of objects in computational integral imaging by analyzing the existing methods for depth extraction of target objects. To improve the resolution of reconstructed slice images, we use a digital camera moving in horizontal and vertical direction with the set interval to get elemental images with high resolution and bilinear interpolation algorithm to increase the number of pixels in slice image which improves the resolution obviously. To show the feasibility of the proposed method, we carried out our experiment and presented the results. We also compared it with other merit functions. The results show that merit function SMD2 to determine the depth of objects is more accurate and suitable for real-time application.

Keywords: computational integral imaging, depth extraction, bilinear interpolation algorithm.

1. Introduction

Depth extraction is a crucial point for three-dimensional reconstruction in modern society. Researchers have proposed some methods for depth extraction such as monocular vision, binocular vision, light field cameras, structured light, integral imaging technique and so on [1–7]. Among them, monocular vision, binocular vision and multi-vision are the most commonly used methods. Monocular vision is simple in structure and fully developed in algorithm, but the prior knowledge of the 3D scene is necessary. Binocular vision requires two cameras to capture two images from different views simultaneously. Besides, it needs a series of processing procedures such as camera calibration, feature extraction, stereo matching, to extract the depth of the scene. However, due to unfavorable factors such as illumination, noise and lens distortion, it is hard to make images unambiguously matched. The multi-vision uses three or more cameras to acquire the images of different views, which is usually used to solve the problem of matching ambiguity in binocular stereo vision. Compared with aforementioned methods, integral imaging proposed by LIPPMAN in 1908 [8] is more straightforward for depth extraction.

Integral imaging is regarded as a promising technique in three-dimensional imaging and display technique, which has triggered much interest for many researchers. Researchers focused on it in the past twenty years. LEE *et al.* used block matching with sum of absolute differences for reconstructed slice images in synthetic aperture integral imaging in order to attain the depth of objects [9]. A method that combines the correlation between an elemental image and a periodic function was proposed by JANG *et al.* [4]. YOO used block matching between the slice image pairs to obtain it via the windowing technique [10]. The aforementioned methods mostly use block matching, which would be affected by the block size. ZHAO *et al.* used integral imaging technology and the sum of an absolute differences evaluation function to achieve the depth of target objects underwater, which can get the depth of target objects conveniently with an appropriate time cost of calculation [11]. Nevertheless, the sensitivity is not high enough near the focus (*i.e.*, the function is too flat near the extreme point and the unimodality is not well). To get the depth more accurately, we use another merit function, which is the morph of sum of absolute differences (SMD2) for reconstructing slice images. SMD2 [12] is an evaluation function that is based on a gray scale difference. It uses the product of the adjacent pixels difference in horizontal and vertical directions as the core function, which maintains a good computing performance and steep unimodality.

One of the drawbacks that limit the further development of integral imaging is low resolution [13]. Interpolation algorithm is an effective way to increase the resolution. There are mainly three kinds of the interpolation algorithm, such as nearest neighbor interpolation, bilinear interpolation and cubic convolution interpolation. The nearest neighbor interpolation has many advantages such as faster operation, but the accuracy is poor. The accuracy of cubic convolution interpolation is the highest but it is the most computationally intensive and the most complex. Taking the computation time into account, we choose the bilinear interpolation algorithm.

In this article, we put forward a method that uses computational integral imaging with bilinear interpolation algorithm and evaluation function SMD2 for depth extraction. In the proposed method, we first capture elemental images with synthetic aperture integral imaging. Then we get a high resolution slice image with computational integral imaging and bilinear interpolation algorithm. Finally, the depth information of objects is extracted by the merit function. To confirm the feasibility of the proposed method, the experiments are carried out and the results are presented. Also, we analyze different evaluation functions in the case of effectiveness and computation.

2. Proposed method

2.1. Slice reconstructed images with computational integral imaging reconstruction

In general, integral imaging is composed of two primary stages, optical pickup stage and reconstruction stage. In the optical pickup stage, the 3D scene is recorded by a micro-lens array and an image sensor, which can capture the elemental images array (EIA)

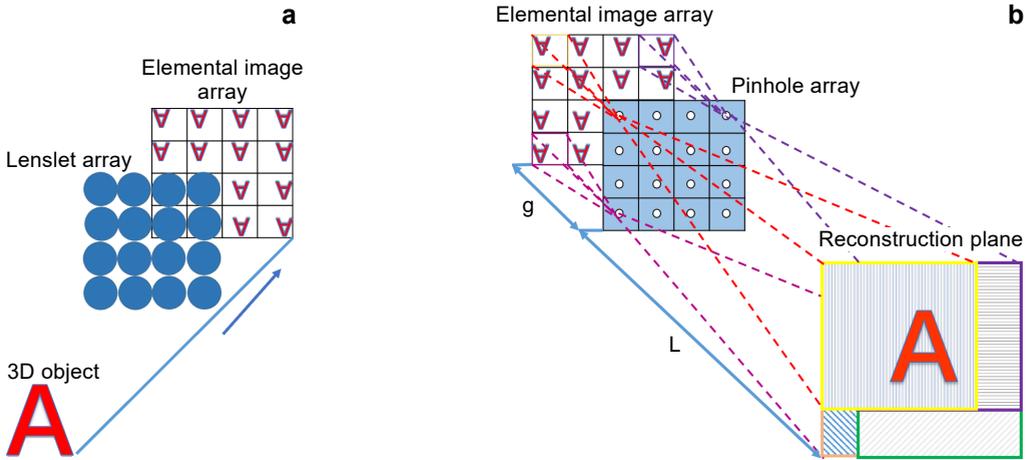


Fig. 1. The concept of integral imaging; the image pickup stage (a), and the reconstruction stage (b).

as shown in Fig. 1a. In the computational reconstruction stage as shown in Fig. 1b, elemental images are projected inversely through a virtual pinhole array to make the 3D scene reconstruct at any reconstruction output plane. There are also two methods in computational reconstruction stage, view-based integral imaging and depth-based integral imaging [13]. View-based integral imaging is via periodically extracting the pixel from the collected elemental images from a certain observation view, and combines the extracted pixels into one view. The resolution of the reconstruction image is equal to the number of the micro-lenses and the resolution is usually low. Depth-based integral imaging is by changing the reconstruction distance to achieve the characteristic of focus and defocus. In this article, we mainly discuss the method of depth-based integral imaging. In order to get elemental images with high resolution, some methods such as synthetic aperture integral imaging and computer-synthesized elemental images have been proposed [13].

Computational integral imaging reconstruction is based on rays back-projection [14]. The conventional computational integral imaging reconstruction is as follows. First, according to the reversibility principle of light rays, every elemental image is projected inversely through its corresponding micro-lens or a visual pinhole with a magnification factor $Lg = L/g$. The magnification factor represents the ratio of the distance $z = L$ between the reconstruction plane and the pinhole array to the distance g between the pinhole array and the elemental image plane. Second, the enlarged elemental images are overlapped and summed at the corresponding pixels of the reconstruction plane. By changing the distance z , we can get a series of reconstruction images.

However, the magnification factor increases with the increase of the reconstructed distance z , so it requires large computation. What is more, the pixel of the image is also magnified by Lg , and the size of the pixel is larger and the resolution of the image is not improved. So we take advantage of pixel-to-pixel mapping proposed by SHIN and Yoo [15] and integrate it with bilinear interpolation algorithm to overcome the

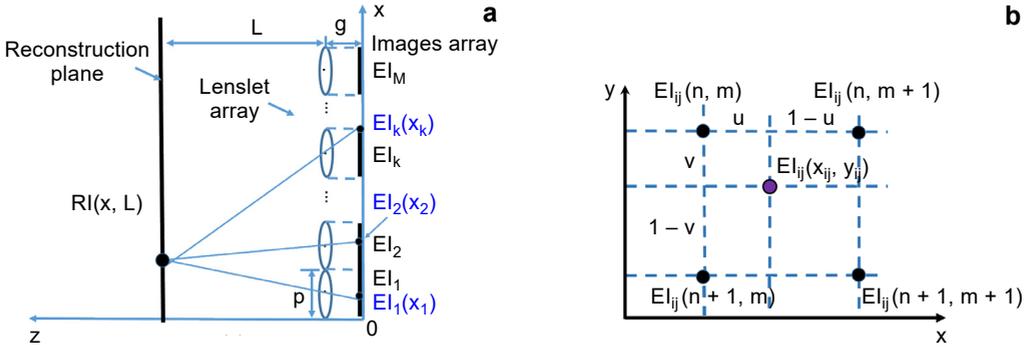


Fig. 2. The principle of our proposed method; the calculation principle of the computational integral imaging reconstruction stage (a), and the diagram of bilinear interpolation ($0 < u < 1$, $0 < v < 1$) (b).

problem of low resolution and huge computation time. Bilinear interpolation algorithm is an effective way to increase the image's resolution.

The proposed computational integral imaging reconstruction method is composed of two primary procedures: the inverse mapping and the interpolation. Suppose that the elemental image array consists of $M \times N$ images and the pitch between the lenslet is P . For the inverse mapping procedure, Fig. 2a illustrates the depth-based method based on the computational integral imaging reconstruction. Let $RI(x, L)$ be the intensity value of x at the reconstructed distance L , and i be the index of the corresponding elemental image. $RI(x, L)$ is evaluated from a set of pixels. The coordinate of each pin-hole in x axis can be set as

$$\text{Pin } X_i = \left(i - \frac{1}{2}\right)p \quad (i = 1, 2, 3, \dots, M) \quad (1)$$

According to geometrical optics, the range of every elemental image in the reconstruction plane can be obtained by Eq. (2). Via the maximum function in $RI_{x, \max i}$ and minimum function in $RI_{x, \min i}$, the ultimate size of the reconstruction image can be given by

$$\begin{cases} RI_{x, \min i} = \text{Pin } X_i - Lg \frac{S_x}{2} \\ RI_{x, \max i} = \text{Pin } X_i + Lg \frac{S_x}{2} \\ RI_x = \max(RI_{x, \max i}) - \min(RI_{x, \min i}) \end{cases} \quad (2)$$

where RI_x , S_x is the size of reconstruction image and the size of each elemental image in x direction, respectively. M and N are the number of elemental images in x , y direction, respectively. This can be extended to vertical y axis as well.

$RI(x, L)$ is attained from a series of pixels and it is determined by averaging the intensity values of all corresponding rays through the ray mapping as shown in Fig. 2a.

This means that the index may not be an integer index. In the reconstruction stage, we use the bilinear interpolation algorithm to calculate the intensity value in each elemental image. The computational integral imaging reconstruction process is given by

$$RI(x, L) = \frac{1}{O(x)} \sum_{i=1}^M EI_i(x_i) = \frac{1}{O(x)} \sum_{i=1}^M EI_i \left(\text{Pin } X_i + \frac{\text{Pin } X_i - x}{Lg} \right) \quad (3)$$

where x_i needs to be limited in a range, that is $RI_{x, \min i} \leq x_i \leq RI_{x, \max i}$. $O(x, y)$ is the total number of pixels that contributes to the wanted intensity value. If the size of square pixel on CCD is l_p , the intensity value of $EI_i(x_i)$ is given by

$$n = \text{floor} \left[\frac{x_i - \text{Pin } X_i + S_x/2}{l_p} \right] \quad (4a)$$

$$u = \left[\frac{x_i - \text{Pin } X_i + S_x/2}{l_p} \right] - n \quad (4b)$$

The index of m and v can be also calculated in y axis. So the final intensity value $EI_{ij}(x_i, y_{ij})$ is determined by the following equation:

$$EI_{ij}(x_{ij}, y_{ij}) = (1-u)(1-v)EI_{ij}(n, m) + u(1-v)EI_{ij}(n+1, m) \\ + (1-u)vEI_{ij}(n, m+1) + uvEI_{ij}(n+1, m+1) \quad (5)$$

After calculating the intensity value, we can get the final reconstructed slice image. The reconstructed image at $RI(x, y, L)$ is the summation of all the inversely mapped pixels. The reconstructed images are a series of slice images.

2.2. Depth extraction with focus merit function

A series of image sequences along z axis direction can be attained via changing the reconstruction distance L from the pinhole array to the reconstruction plane. If the real depth of the object is equal to the reconstructed distance L , the object on the slice image is focused, otherwise the object on the slice image is defocused. As shown in Fig. 3, when the real depth of point A is equal to the reconstructed distance L_1 , the point A on the reconstruction plane 1 is a clear point A' , otherwise the image A' is blurred in reconstruction plane 2. This is fundamental of focus and defocus in reconstruction slice images. Therefore, we can use this method to extract the depth information of objects effectively.

There exist some methods to evaluate the degree of focus, which can be divided into two categories: spatial domain and frequency domain function. The spatial focus evaluation function includes a gray-scale difference method, image gray-entropy and so on [12]. The frequency domain function requires to convert the image into the corresponding frequency domain. The two basic transformations are Fourier transform and wavelet transform. To get the depth more accurately, we use the evaluation function

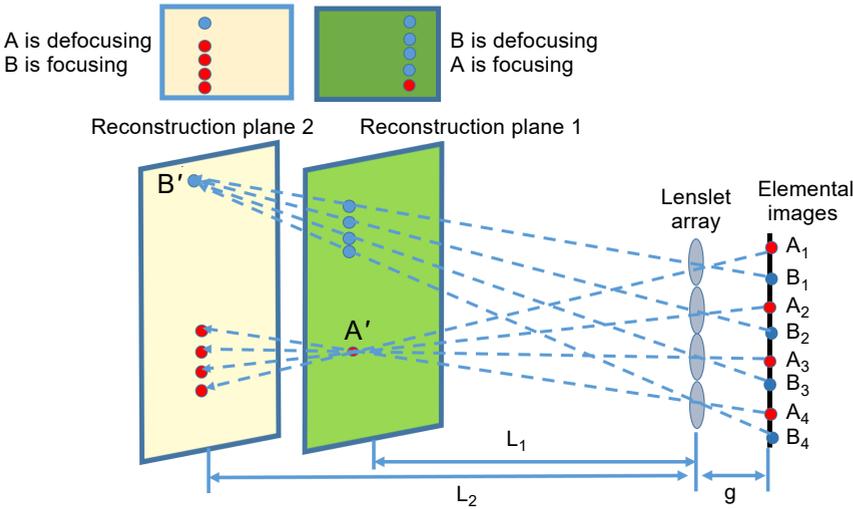


Fig. 3. The principle of focus and defocus of slice reconstruction images.

SMD2, which is the morph of summation of the absolute difference, for a series of reconstructed slice images. SMD2 is an evaluation function that is based on a gray scale difference [12]. It uses the product of an adjacent pixels difference in the horizontal and vertical directions as a core function. It is an effective and simple evaluation function to check up whether the image is in focus or defocus. In general, people take the maximum value of the function as the criterion of judgement. Compared with conventional mostly used evaluation functions such as energy function, peak signal to noise ratio and so on, it could maintain good computational performance and improve steepness of the function, which can speed up the convergence of the search focus and improve the focus accuracy. It is widely used for focus evaluation of sequence images [11]. The evaluation value of SMD2 function can be calculated as [12]

$$\begin{cases} \text{SMD2} = \frac{1}{K} \sum_x \sum_y I(x, y) \\ I(x, y) = |\text{RI}(x, y) - \text{RI}(x + 1, y)| |\text{RI}(x, y) - \text{RI}(x, y + 1)| \end{cases} \quad (6)$$

where $I(x, y)$ is the product of an adjacent pixels difference in the horizontal and vertical directions; K is the total number of pixels for every reconstruction image. By traversing the entire reconstructed image, we can get the evaluation value SMD2. Every slice image has a SMD2 value, the maximum value is corresponding to the focus image. The larger the value, the clearer the reconstructed image.

3. Experiments and results

To show the feasibility of our proposed method, we carried out preliminary experiments. Figure 4a shows the experimental setup where a CCD camera is moving on a two-axis

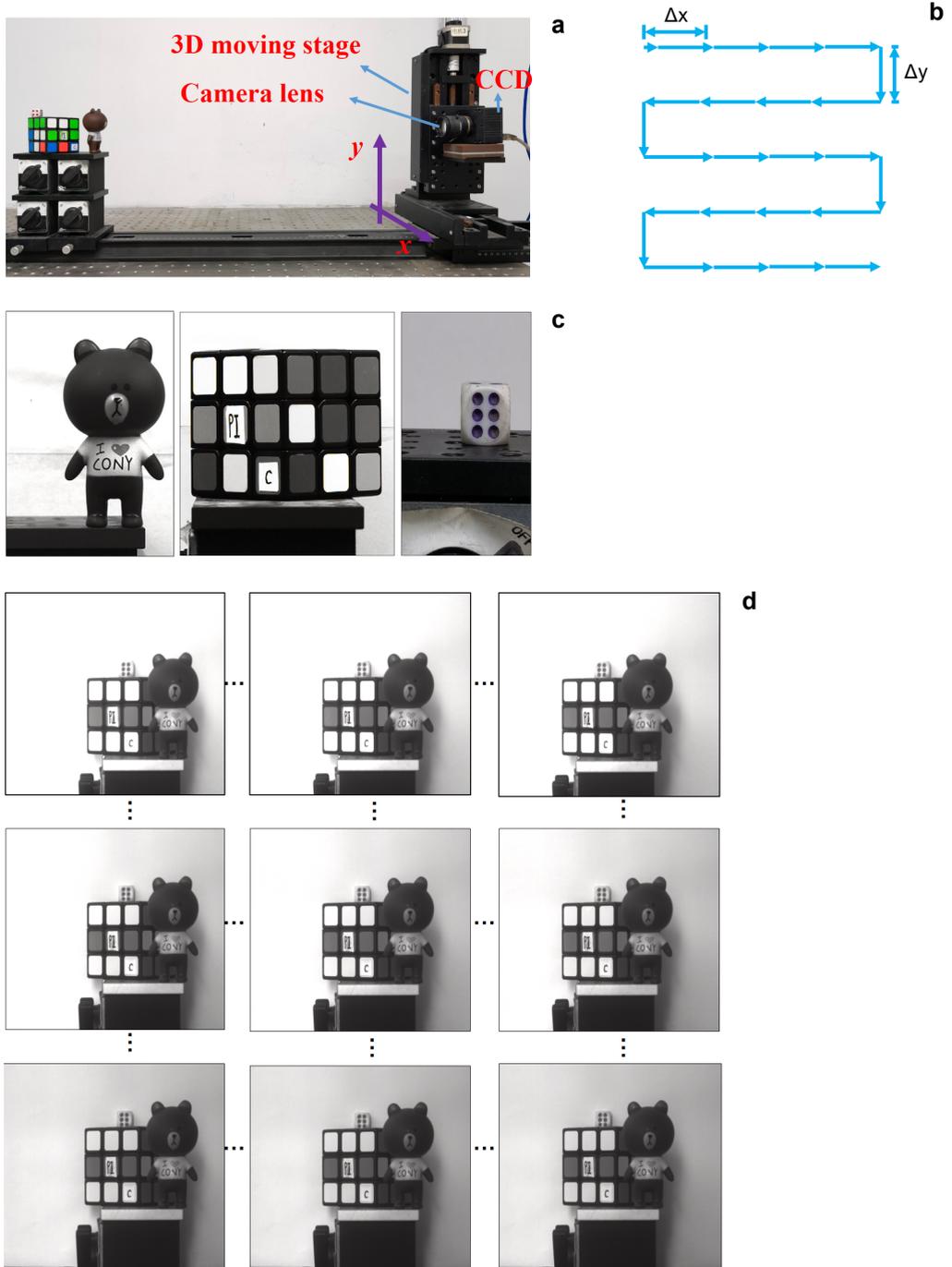


Fig. 4. Experimental setup for capturing the EIA of the scene composed of three objects. Three objects of bear, cube and dice, and a CCD camera on a 3D moving translation stage (a); the moving path of the CCD camera (b); the clear images of the three objects (c); the captured elemental images array (d).

translation stage. The movement of the translation stage is controlled by a software and a three-dimensional controller. The 3D scene is composed of a bear, a cube and a dice which are located at appropriately 900, 958 and 987 mm away from the CCD camera, respectively. On the cube, there are two labels: PI and C. The focal length of the image sensor is 35 mm. Elemental images array with high resolution are captured with synthetic aperture integral imaging system as shown in Fig. 4d. The moving path with the pitch $\Delta x = \Delta y = 4$ mm of the camera is shown in Fig. 4b. The resolution of the CCD is 1024×1280 . The clear images of the bear, the cube and the dice are shown in Fig. 4c, which can be regarded as the reference images.

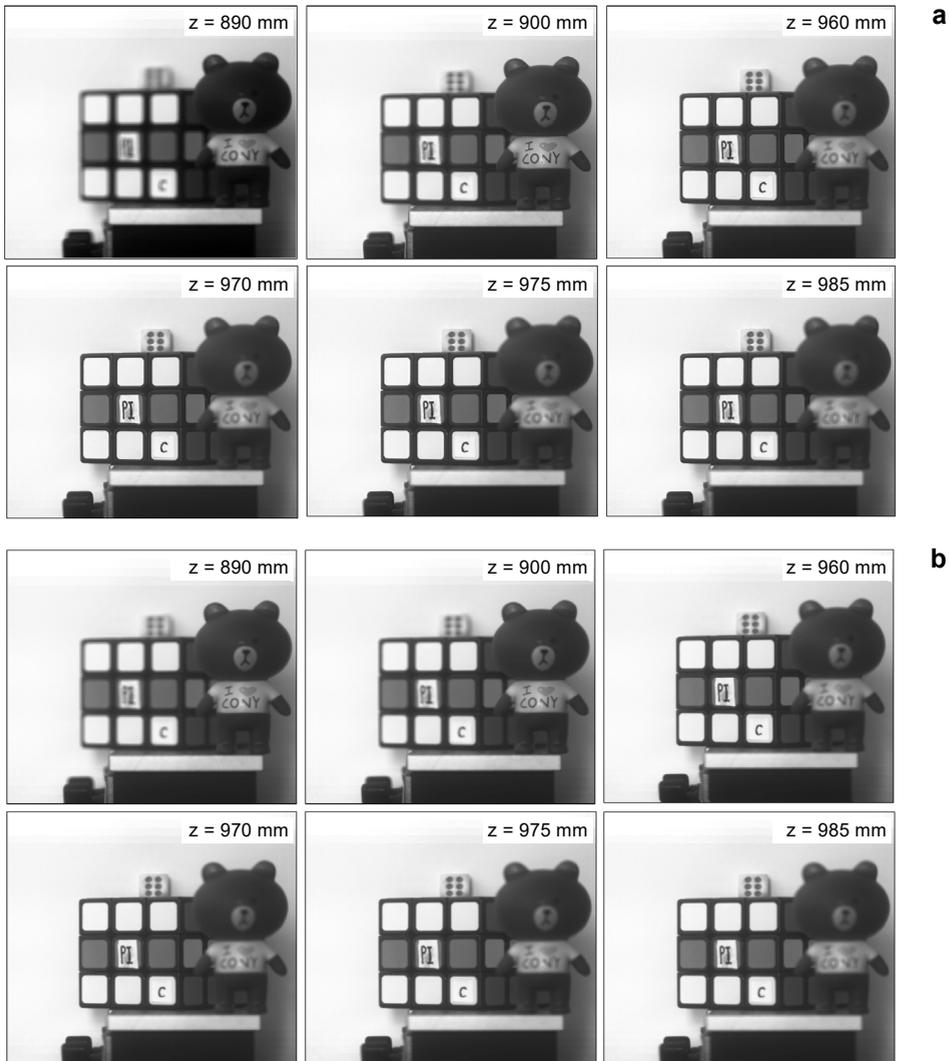


Fig. 5. Comparison of images reconstructed by different methods at the distance of 890, 900, 960, 970, 975, and 985 mm. Conventional method (a); proposed method (b).

All objects images of bear, cube, dice and the two labels on the cube have been simultaneously reconstructed from the captured EIA. We get slice reconstructed images with different depth in the range from 880 to 1030 mm. Slice reconstructed images with different distance are shown in Fig. 5. Figure 5b shows the reconstructed images using the proposed method at the distance of 900, 960, 975, and 985 mm, where the bear, the letter C, the label PI and the dice are in focus, respectively. The comparison between the conventional integral without interpolation and proposed method is also implemented shown in Fig. 5. We can see that the bear in reconstructed images using the proposed method is clearer than the image using the conventional method.

There are several metrics to evaluate image quality, such as contrast, sharpness [16], PSNR and variance. The absolute value of gradient is used to measure the sharpness. Table 1 compares sharpness, contrast, PSNR and variance values of the images reconstructed by the conventional method and the proposed method with six different distances. Higher sharpness, contrast, variance and PSNR are obtained using the proposed method, which indicates that the proposed method outperform the conventional approach.

To show that our method is better for a real-time application, we calculate the time requirement for one time processing. Table 2 compares the computation time of a slice image reconstructed by the conventional method without interpolation, the proposed method with bilinear interpolation and the proposed method with cubic interpolation. It can be seen that the reconstruction without interpolation is the fastest, but in terms of the sharpness and contrast (see Table 1), the quality of the reconstructed image in Fig. 5a is worse. The PSNR and the variance values of the images reconstructed by the conventional method without interpolation are lower than in the proposed method.

T a b l e 1. The comparison of sharpness, contrast, variance and PSNR values between the conventional method and the proposed method for different distances in Fig. 5.

Method		Distance [mm]					
		890	900	960	970	975	985
Conventional	Sharpness	540747	567935	596874	591187	586399	572580
	Contrast	0.789	0.796	0.795	0.796	0.796	0.796
	Variance	79.717	80.226	80.933	80.862	80.764	80.555
	PSNR	28.6651	28.6219	28.5519	28.5582	28.6191	28.6167
Proposed	Sharpness	2496795	3652216	3521428	2762004	2448556	2479165
	Contrast	0.932	1	0.998	0.998	1	1
	Variance	80.140	80.959	81.748	81.620	81.493	81.226
	PSNR	29.0926	29.0675	30.0476	30.0484	30.0509	30.1392

T a b l e 2. The comparison of three methods in computation time.

Method	Without interpolation	Bilinear interpolation	Cubic interpolation
Time required for a slice image	1.427372	55.485942	94.465645

The computation time of reconstruction with bilinear interpolation is reduced by half compared to the cubic interpolation. Considering the computation time, the proposed method with bilinear interpolation is suitable and effective.

To extract the depth of objects, we use the SMD2 function on every slice image. If the reconstruction distance of the object is equal to the original depth of object, the object would be in focus and naturally the SMD2 value will be larger. Figure 6a shows the SMD2 values of slice images. From Fig. 6a, the points of A, B, E, F represent the depth of 900, 960, 975 and 985 mm, which is near equal to the original depth of the toy, letter C, label PI and the dice. The comparison between SMD2 and other merit functions such as SMD, variance, energy function and so on are also shown in Figs. 6b–6e. In terms of sharpness, SMD2 is steeper near the focus than SMD. The SMD function in Fig. 6b can be used in slice image in a certain extent, but it is too flat near the extreme point and the unimodality is not good. For example, the original depth of the cube is

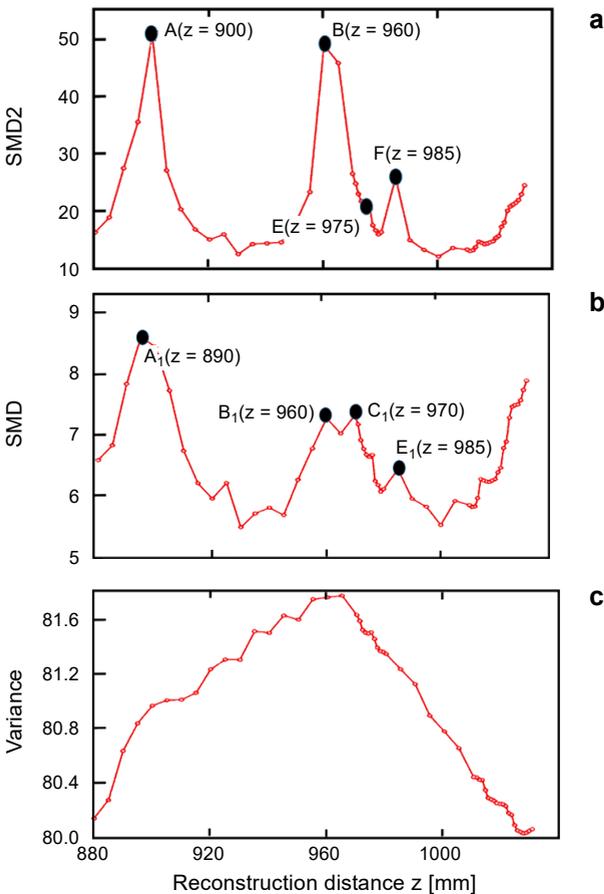


Fig. 6. A comparison of five different merit function values: SMD2 (a), SMD (b), variance (c), Vollath (d), and energy (e).

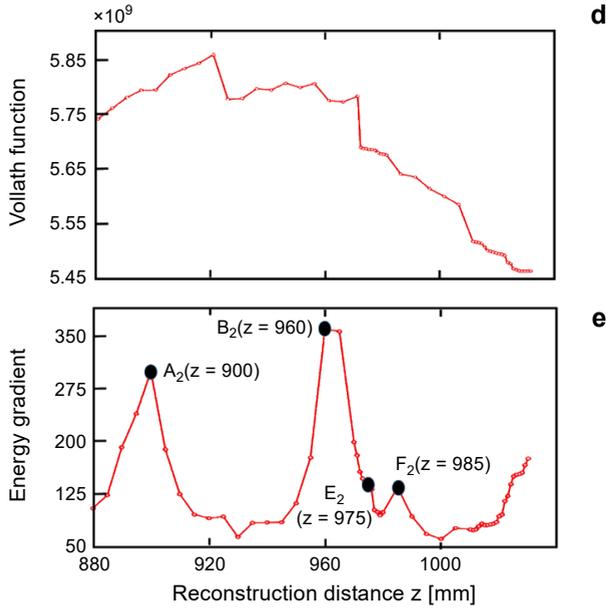


Fig. 6. Continued.

Table 3. The comparison of different merit functions in computation time.

Merit function	SMD2 function	SMD function	Variance	Vollath function	Energy gradient function
Computation time [s]	0.0328775	0.037065	0.027374	0.026448	0.04507

958 mm, and considering the error of measurement, the peak of SMD function in the range of 940–965 mm ought to appear once, but two peaks appear in the same range, which means that the depth extraction is wrong and the precision is not sufficient. The sharpness and unimodality of energy function are similar with SMD2 function, which both have one peak in every target depth range and can determine the target depth uniquely. As for other two functions, we can see the functions are too flat near the peak. Therefore, the SMD2 and energy function are better choices for depth extraction. However, taking the computation time into consideration, the energy function is slower than the SMD2 function as shown in Table 3, which is vulnerable to real-time application. With comprehensive consideration, SMD2 outperform other evaluation functions.

4. Conclusion

In conclusion, we have presented a depth extraction method using computational integral imaging. In the proposed method, elemental images array with high resolution is attained by moving a digital camera in horizontal and vertical direction with the set interval. To improve the resolution of reconstructed slice images, we make use of computational

integral imaging reconstruction with bilinear interpolation algorithm. We compare the sharpness, contrast and PSNR between the conventional method and the proposed method. Besides, the calculation time of different reconstruction methods is compared, which indicates that the reconstruction with bilinear interpolation algorithm is a better choice. In addition, we use the merit function SMD2 to determine the depth of objects and compare different merit functions, which shows that SMD2 function is more accurate and suitable for real-time application.

Acknowledgements – This work was supported by the Natural Science Foundation (6180030017), the Key Technologies R&D Program of Guangzhou City (201704020038, 201704020182, 201803030008).

References

- [1] WANG T.C., EFROS A.A., RAMAMOORTHY R., *Occlusion-aware depth estimation using light-field cameras*, 2015 IEEE International Conference on Computer Vision (ICCV), December 7–13, 2015, Santiago, Chile, pp. 3487–3495, DOI: [10.1109/ICCV.2015.398](https://doi.org/10.1109/ICCV.2015.398).
- [2] WANG C., SAHIN E., SUOMINEN O., GOTCHEV A., *Depth estimation by combining stereo matching and coded aperture*, 2014 IEEE Visual Communications and Image Processing Conference, December 7–10, 2014, Valletta, Malta, pp. 291–294, DOI: [10.1109/VCIP.2014.7051561](https://doi.org/10.1109/VCIP.2014.7051561).
- [3] WANG T.C., EFROS A.A., RAMAMOORTHY R., *Depth estimation with occlusion modeling using light-field cameras*, IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(11), 2016, pp. 2170–2181, DOI: [10.1109/TPAMI.2016.2515615](https://doi.org/10.1109/TPAMI.2016.2515615).
- [4] JANG J.Y., SER J.I., CHA S., SHIN S.H., *Depth extraction by using the correlation of the periodic function with an elemental image in integral imaging*, Applied Optics **51**(16), 2012, pp. 3279–3286, DOI: [10.1364/AO.51.003279](https://doi.org/10.1364/AO.51.003279).
- [5] SAID Z., SUNDARAJ K., WAHAB M.N.A., *Depth estimation for a mobile platform using monocular vision*, Procedia Engineering **41**, 2012, pp. 945–950, DOI: [10.1016/j.proeng.2012.07.267](https://doi.org/10.1016/j.proeng.2012.07.267).
- [6] LONG H.M., GUO H.Y., LIANG F., LIU G.H., *Distance measurement algorithm based on binocular stereo vision*, Applied Mechanics and Materials **635–637**, 2014, pp. 948–952, DOI: [10.4028/www.scientific.net/AMM.635-637.948](https://doi.org/10.4028/www.scientific.net/AMM.635-637.948).
- [7] LIU W., WANG Z., NING X., ZHANG X., *An object recognition and location based on binocular stereo vision*, 2009 International Forum on Computer Science-Technology and Applications, December 25–27, 2009, Chongqing, China, pp. 490–492, DOI: [10.1109/IFCSTA.2009.262](https://doi.org/10.1109/IFCSTA.2009.262).
- [8] LIPPMANN G., *Épreuves réversibles donnant la sensation du relief*, Journal de Physique Théorique et Appliquée **7**(1), 1908, pp. 821–825, DOI: [10.1051/jphys:019080070082100](https://doi.org/10.1051/jphys:019080070082100).
- [9] LEE J.J., LEE B.G., YOO H., *Depth extraction of three-dimensional objects using block matching for slice images in synthetic aperture integral imaging*, Applied Optics **50**(29), 2011, pp. 5624–5629, DOI: [10.1364/AO.50.005624](https://doi.org/10.1364/AO.50.005624).
- [10] YOO H., *Depth extraction for 3D objects via windowing technique in computational integral imaging with a lenslet array*, Optics and Lasers in Engineering **51**(7), 2013, pp. 912–915, DOI: [10.1016/j.optlaseng.2013.02.009](https://doi.org/10.1016/j.optlaseng.2013.02.009).
- [11] ZHAO X., WANG Y., SONG L., ZHANG B., ZHAO X., *Underwater target imaging based on computational integral imaging*, Chinese Journal of Lasers **45**(1), 2018, article 0109001 (in Chinese), DOI: [10.3788/CJL201845.0109001](https://doi.org/10.3788/CJL201845.0109001).
- [12] LI Y., CHEN N., ZHANG J., *Fast and high sensitivity focusing evaluation function*, Application Research of Computers **27**(4), 2010, pp. 1534–1536 (in Chinese).
- [13] XIAO X., JAVIDI B., MARTINEZ-CORRAL M., STERN A., *Advances in three-dimensional integral imaging: sensing, display, and applications [Invited]*, Applied Optics **52**(4), 2013, pp. 546–560, DOI: [10.1364/AO.52.000546](https://doi.org/10.1364/AO.52.000546).

- [14] HONG S.H., JANG J.S., JAVIDI B., *Three-dimensional volumetric object reconstruction using computational integral imaging*, Optics Express **12**(3), 2004, pp. 483–491, DOI: [10.1364/OPEX.12.000483](https://doi.org/10.1364/OPEX.12.000483).
- [15] SHIN D.H., YOO H., *Computational integral imaging reconstruction method of 3D images using pixel-to-pixel mapping and image interpolation*, Optics Communications **282**(14), 2009, pp. 2760–2767, DOI: [10.1016/j.optcom.2009.04.008](https://doi.org/10.1016/j.optcom.2009.04.008).
- [16] DO C.M., JAVIDI B., *3D integral imaging reconstruction of occluded objects using independent component analysis-based k-means clustering*, Journal of Display Technology **6**(7), 2010, pp. 257–262.

*Received July 4, 2019
in revised form November 12, 2019*