




# Automatic objects' depth estimation based on integral imaging

Fatemeh Kargar Barzi<sup>1</sup> · Hossein Nezamabadi-pour<sup>1</sup> 

Received: 23 October 2020 / Revised: 10 March 2022 / Accepted: 12 May 2022 /  
Published online: 25 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

A new approach for depth estimation based on integral imaging is proposed. In this method, multiple viewpoint images captured from a three-dimensional scene are used to extract the range information of the scene. These elemental images are captured through an array of lenses over a high-resolution camera or an array of cameras. Then the scene is computationally reconstructed in different depths using integral imaging reconstruction algorithm. Finally, by processing the reconstructed images and finding objects of the scene in these images using a matching technique with speeded-up robust features (SURF), the depth information of the objects will be acquired. Experimental results show that the proposed method has high accuracy and does not have many limitations of standard image processing, including sensitivity to the surface type and size of the objects.

**Keywords** Depth estimation · Image processing · Integral imaging · 3D reconstruction · SURF features

## 1 Introduction

Measuring the distance between an object and a specific origin is a common problem in human life. There are many methods to measure this distance. It can be calculated by processing the images captured by a camera located at that origin. In this case, the distance between the object and the camera lens is equal to the object's depth. Depth estimation of an object from its 2-D images is one of the primary problems in machine vision [3, 19, 50] and has a wide range of

---

✉ Hossein Nezamabadi-pour  
nezam@uk.ac.ir

Fatemeh Kargar Barzi  
f.kargar@eng.uk.ac.ir

<sup>1</sup> Intelligent Data Processing Laboratory (IDPL), Department of Electrical Engineering, Shahid Bahonar University of Kerman, Kerman, Iran

applications, such as 3-D reconstruction, speed estimation, object detection, and background extraction [20, 32, 44].

There has been plenty of methods for depth estimation, which can be categorized into two main groups: active methods and passive sensor-based methods. Active methods include laser scanning [21] and structured light methods [14, 15]. These methods utilize physical sensors to derive depth information [47]. Usually, these methods are more effective and efficient in acquiring depth information. However, their required devices are expensive [4, 5, 49]. In addition, because depth estimation is based on backlight information, it does not have high accuracy in low-reflection environments such as black areas. Also, estimated depth values at the boundaries of moving objects will not be reliable. Passive sensor-based depth estimation methods use cameras to evaluate depth information [13, 18, 28, 46]. Therefore, these methods are low-cost but slow [7]. The stereo technique [2, 22, 26, 40, 48] is one of the most popular passive methods in which images from two cameras are used to estimate distances. The two images are taken from a pair of cameras at two different positions. Depth estimation in this method is based on the disparity between these two images. Although this system works well in many environments, we should consider that in parts of the image in which illumination or color is uniform, finding the pixels corresponding to the same points is complex. Furthermore, in stereo imaging, all aspects of the environment don't exist in all images. In other words, some features will be hidden because of occlusion [41].

To estimate the depth of objects in obstructive environments, we need to increase the number of cameras; therefore, we can see all the details of the environment. In [12], a depth estimation method is presented based on the processing of images recorded by an array of micro-lenses. The basis of this method is to compare the images in terms of the similarity of pixels. Although this algorithm is highly accurate. Depth estimation accuracy will degrade when object surfaces do not satisfy the Lambertian assumption or when concave surfaces exist. At a Lambertian surface, the reflection of a beam from that surface has the same intensity in all directions. Still, on non-Lambertian surfaces, from different angles, the intensity of the light is different. A method for estimating depth based on integral imaging was proposed in [36, 45]. In this method, a soft-voting system that considers the level of similarity among elemental images recorded from different views is proposed. The accuracy of this method is reported between 83.98% and 93.71% for simulated environments with 3DMAX software. Our purpose is to increase the precision of the depth estimation of objects in real and occluded environments. In addition, in most of the methods mentioned, the test objects are small and close to the lenslet array. In this paper, we intend to introduce a method for estimating the depth of large objects at a far distance from the lenslet array.

Integral imaging system is one of the technologies that is based on processing a large number of images for 3D reconstruction [6, 10, 23, 25, 31, 33, 35]. In this system, first, several images are captured using an array of micro-lenses or cameras. These images are called elemental images. Then using these pictures, the environment will be reconstructed, in 3D. The reconstruction process can be done computationally. In this case, in the reconstructed image related to each specific depth, the objects in that depth are apparent, and the other objects are a blur. Therefore, the object's depth can be realized by detecting on which depth, that object was clearly reconstructed.

In this paper, a new method for depth estimation based on integral imaging system is proposed. In the proposed method, depth information is estimated by processing several images, which are captured in a specific manner which will be explained in Section 2. Increasing the number of elemental images will improve the quality of 3D reconstruction and hence the precision of depth estimation. However, the system speed will be reduced due to computation increment. In the proposed method, we estimate the depth information using five primary images to achieve good

accuracy for depth estimation and acceptable computational speed. Furthermore, the results show that because of the linear camera arrangement, the proposed method has acceptable performance in estimating the depth of occluded objects. The validity of our approach has been demonstrated by experimental results in different scenarios.

The following is a detailed description of this method. Section 2 introduces integral imaging and two approaches have been proposed so far based on integral imaging. Section 3 describes the new method presented in this paper. Experimental setup, an example to explain more about the proposed method, and the results of reconstruction and depth estimation are presented in Section 4. Finally, the paper concludes in Section 5.

## 2 Principles of integral imaging

The integral imaging technique was first introduced by Lippman [34] in 1908. With progressing digital cameras and digital displays and using computers for processing techniques, the attention to integral imaging has been increased. Today, there are many researches in integral imaging using the most advanced technologies and imaging methods. Some of the achievements of this researches are: Three-dimensional displays and TVs [42], automatic target detection, [37], depth detection [1], three-dimensional imaging for photon counting [1], 3D imaging and 3D reconstruction for occluded objects [38], 3D imaging underwater [30], medical imaging [27] and so on.

The method of integral imaging has two main stages: pickup and reconstruction. In the pickup process, the elemental image array is captured through a lenslet array by an imaging device. The pickup step of integral imaging uses a camera array, which is shown in Fig. 1. This camera array contains several cameras in the form of 1-D line or 2-D plate. Determining array characteristics such as the number of lenses, their arrangement (circular, linear, two-dimensional), distance, angles, and size, strongly depends on the application [24].

The reconstruction step is the reverse process of the pickup and it can be performed either optically or computationally. In integral imaging, the reconstruction is based on the principle of reversibility of light rays. In the optical reconstruction, as it is the reverse of the pickup process, elemental images are shown on a display panel such as a liquid crystal display (LCD) as shown in Fig. 2. By placing the lens array used in the pickup step in front of the display panel, a 3-D scene image will be reconstructed. The output of the computational reconstruction in a specific depth  $Z_0$  is a 2-D image in which the objects originally located at  $Z_0$  are apparent and objects located at other depths are blurred. The level of blurring of an object in the reconstructed image is proportional to its axial distance from  $Z_0$ .

To discuss the mathematical formulation, consider Fig. 3. Each voxel of object at position  $(x, y, z)$  maps to several pixels, each one on an image plane behind the lens [23]. The pixels intensity value of an elemental image  $i$  is inversely proportional to the square of the distance  $d$  between that elemental image and the reconstruction plane. Therefore, for the reconstructed point at  $(x, y, z)$  as shown in (1) and (2) [17, 23]:

$$O_{pq}(x, y, z) = \frac{I_{pq}(x_i, y_i)}{d^2(x, y, z)} \quad (1)$$

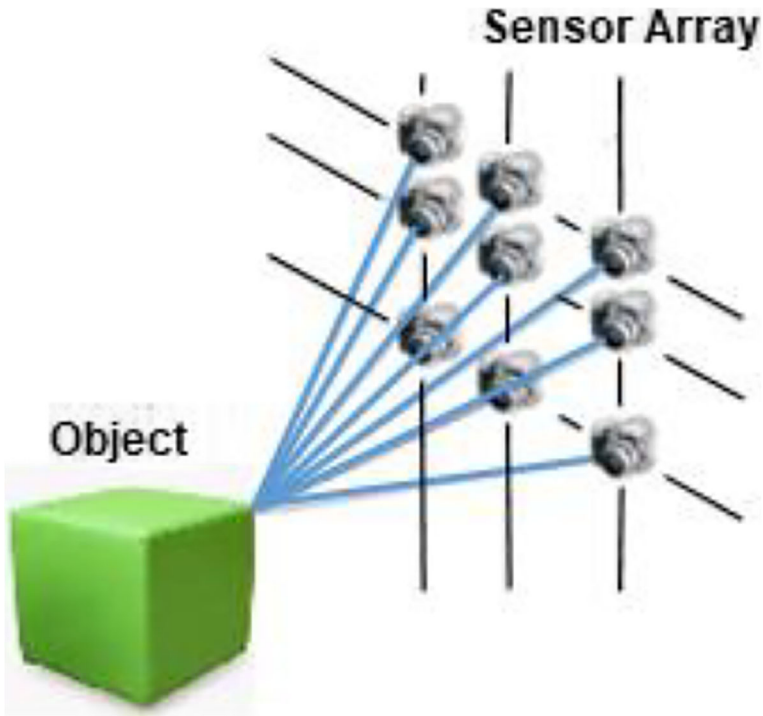


Fig. 1 A typical 2-D camera array used in the pickup step

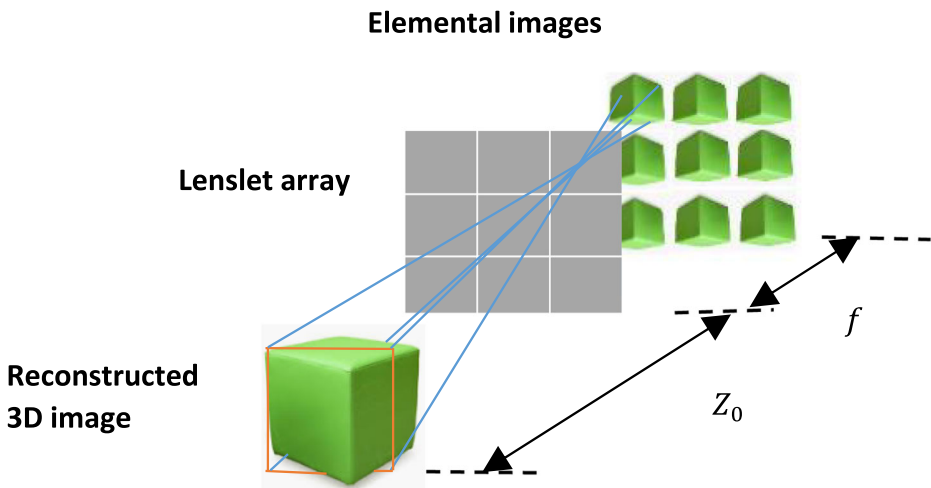


Fig. 2 The 3D optical reconstruction principle based on the back projection of the elemental images through an array of lenses with the focal length  $f$ . All elemental images are back-projected through their viewpoints and the superposition of the ray cones projected from them will reconstruct the 3D scene

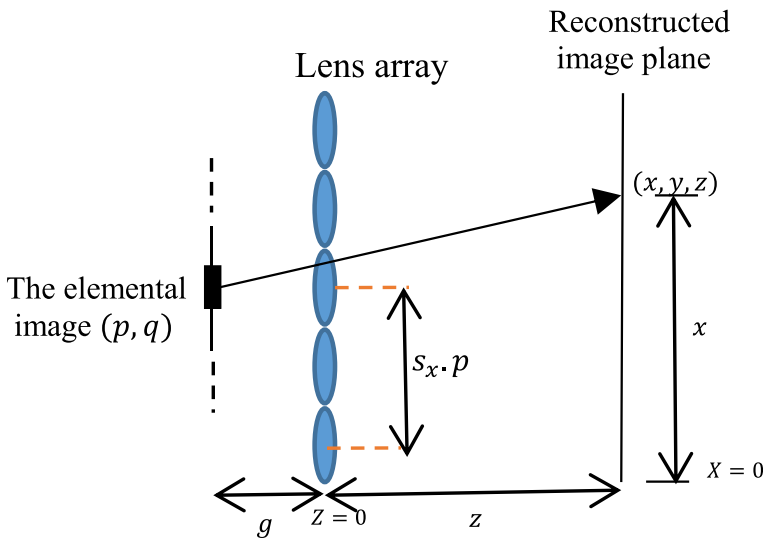


Fig. 3 The computational reconstruction of a point source (x, y, z) on the object

$$x_i = s_x p - \frac{x - s_x p}{M}, \quad y_i = s_y q - \frac{y - s_y q}{M} \tag{2}$$

Where  $O(x, y, z)$  is the intensity value of the reconstructed voxel  $(x, y, z)$ ,  $I(x_i, y_i)$  is the intensity value of the corresponding image point and  $d$  is the distance between the voxel and image point. Magnification factor  $M$  is the ratio of the distance between the lenslet array and the reconstruction image plane at  $z = L$ , to the distance between lenslet array and the elemental image plane ( $g$ ), that is  $M = L/g$ . Finally,  $s_x$  and  $s_y$  are the elemental image sizes in  $x$  and  $y$  directions, respectively.

Considering Fig. 3, distance  $d$  can be written as shown in (3) [23].

$$d^2(x, y, z) = (z + g)^2 + \left[ (x - s_x p)^2 + (y - s_y q)^2 \right] \left( 1 + \frac{1}{M} \right)^2 \tag{3}$$

The 3-D reconstructed image at  $(x, y, z)$  is equal to the summation of effect of all elemental images at  $(x, y, z)$ , as shown in (4) [23].

$$O(x, y, z) = \sum_{p=0}^{m-1} \sum_{q=0}^{n-1} O_{pq}(x, y, z) \tag{4}$$

Where  $m$  and  $n$  are the number of the lenses in  $x$  and  $y$  directions, respectively. In this way, the scene is reconstructed at any desired depth. Authors in [12, 36] proposed a methodology to estimate the depth of objects in a scene using integral imaging-based approaches. The following are the principles of these methods.

### 2.1 Depth estimation using minimum variance

In [12], a methodology is proposed to obtain the depth map of a scene using a minimum variance criterion. In this method, using the integral imaging technique, objects are

reconstructed in the 3D space. For each volumetric pixel (voxel) in space, spectral radiation pattern (SRP) function is the radiation intensity. It is defined as  $L(\theta, \varphi, \lambda)$  where  $-\pi/2 \leq \theta < \pi/2$  and  $0 \leq \varphi < 2\pi$  determine the radiation angle and  $\lambda$  denotes wavelength. Considering Fig. 4, the effect of the  $i$ -th elemental image at  $(x, y, z)$  is defined as (5):

$$L(\theta_i, \varphi_i, \lambda) = \gamma_i \times I_i(\xi, \eta, \lambda) \tag{5}$$

Where

$$(\xi, \eta) = -g_i \tan \theta_i \times (\cos \varphi_i, \sin \theta_i)$$

and  $g$  is the distance between the lens and its image plane. Also  $\gamma_i$  is  $R_i^2 / \min(R_1^2, R_2^2, \dots, R_N^2)$  in which  $R_i$  is the distance between the  $i$ -th lens and the point  $(x, y, z)$ .

For each voxel, by summing up the effect of  $N$  samples created by  $N$  sensors, SRP is calculated. The variance of this function  $L$  can lead us to the depth information of the environment. The SRP samples are correlated with each other in voxels which belong to an object surface. Therefore, the variance of  $L$  has a minimum along with a range,  $z \in [z_{min}, z_{max}]$ . The depth of object surface is estimated by finding  $z$  such that the variance is minimized over the range of interest. Although this method has yielded promising results, it has some limitations in estimating the depth of non-Lambertian and concave surfaces.

### 2.2 Photo-consistency based maximum voting approach

In this approach [36], the depth map of the scene is generated using a voting strategy. This voting strategy is based on the local information processing for each pixel. In this way, a precise depth map with sharp boundaries is created by examining the similarity between

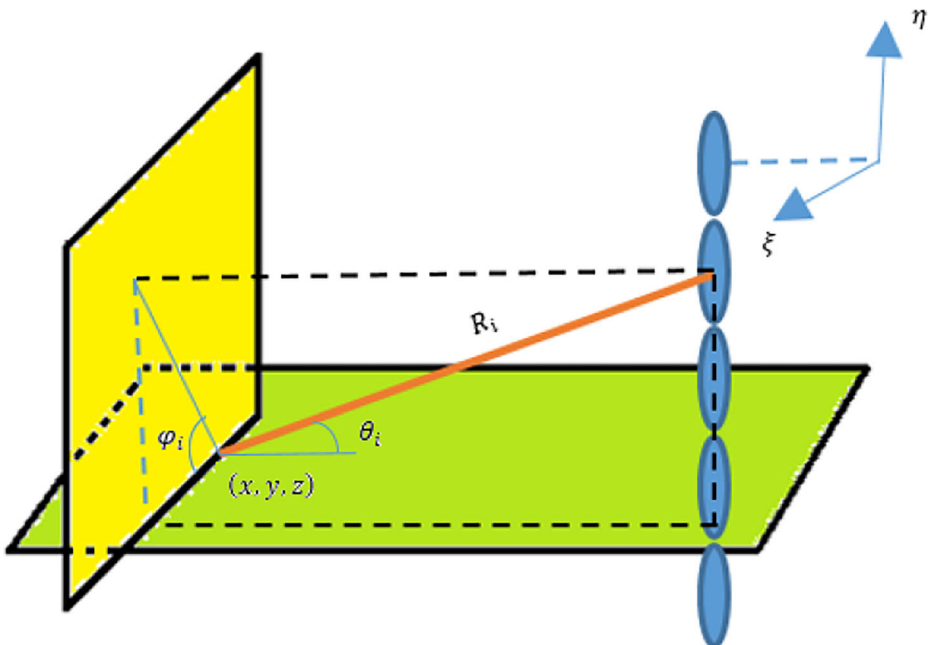


Fig. 4 Parameter definition for SRP

different camera views. In the reconstructed image  $I$  at the depth level  $z$ , let us consider a pixel at position  $(i, j)$  in the image and its square surrounding window  $W$  defined as (6) [36]:

$$W_{ij} = \{I(i+x, j+y) : -\tau \leq x, y \leq \tau\} \quad (6)$$

where  $(2\tau + 1)$  is the size of the window. For a pixel at the position  $(i, j)$  of the image  $I$ , each camera votes to the advantage of the pixel  $(i, j)$  at depth level  $z$  depending on the similarity between the intensities of pixels of each camera and the pixels of reference camera  $R$ . Each camera determines how much its pixel  $(i, j)$  brightness is similar to the pixel  $(i, j)$  image of the reference camera. The similarity criterion, as seen in (7) is the Euclidean distance among the pixel  $(i, j)$  from the reference camera and the pixels  $(x, y)$  from each camera [36].

$$d_{ij}(x, y)^2 = \sum_k^N (I_k(x, y) - I_R(i, j))^2 \quad (7)$$

where  $N$  is the number of cameras. By summing up this distance for all pixels of the window according to Eq. 4 and applying a weighting coefficient, the vote of cameras to the reconstructed pixel at the depth level  $z$  is obtained (8) [36].

$$V(i, j, z) = M \sum_{(x,y)}^{W_{ij}} \exp(-d_{ij}^2(x, y) / THR) \quad (8)$$

In this way, the image of the environment in the range  $z \in [z_{min}, z_{max}]$  with the desired steps is reconstructed and the  $V(i, j, z)$  is calculated for all pixels of the reconstructed images. The proposed Maximum Voting approach estimates the related depth of each pixel  $(i, j)$  based on the maximum value of votes obtained when this process is applied at each depth level  $z$ . Therefore, for all the pixels in the reference elemental image  $R$ , the  $z$  value, where the  $V(i, j, z)$  is maximum, is selected as the range of pixel  $(i, j)$ . Thus, a depth map of the scene is created. This process is done in  $L^*a^*b^*$  color space and the Euclidean distance  $d$  is calculated in this space because of the property of distances between two colors and the human perceptual difference between them.

### 3 Proposed method

Most of the methods, which estimate the depth with processing several images of the scene are inspired by the stereo method and compare images to identify similar points [12, 36]. The same method has been used in the proposed method. The proposed depth extraction system is shown in Fig. 5. In the proposed method, we compare the reconstructed images of different depths with the original image of the environment and choose an image in which the object of interest has the highest sharpness. For automatic depth estimation, the sharpest image is found using speeded-up robust features (SURF) detector and descriptor. Therefore, the depth related to the sharpest reconstructed image of an object, is presented as the object's depth.

The number of elemental images has a significant impact on the quality of reconstruction. As the number of these images increases, the sharpness of the reconstructed image at depth  $Z$  for the object located at depth  $Z$  increase. Therefore, in this case, in the reconstructed image only the object of depth  $Z$  is in-focus and the rest of the scene is blurred. To describe the proposed method, we first look at how to capture elemental images. Then we explain how to find the objects of interest in those photos. In the next section, we will explain how to reconstruct the environment at different depths. Section 3.4 also describes how to automatically estimate and present the depth of objects.

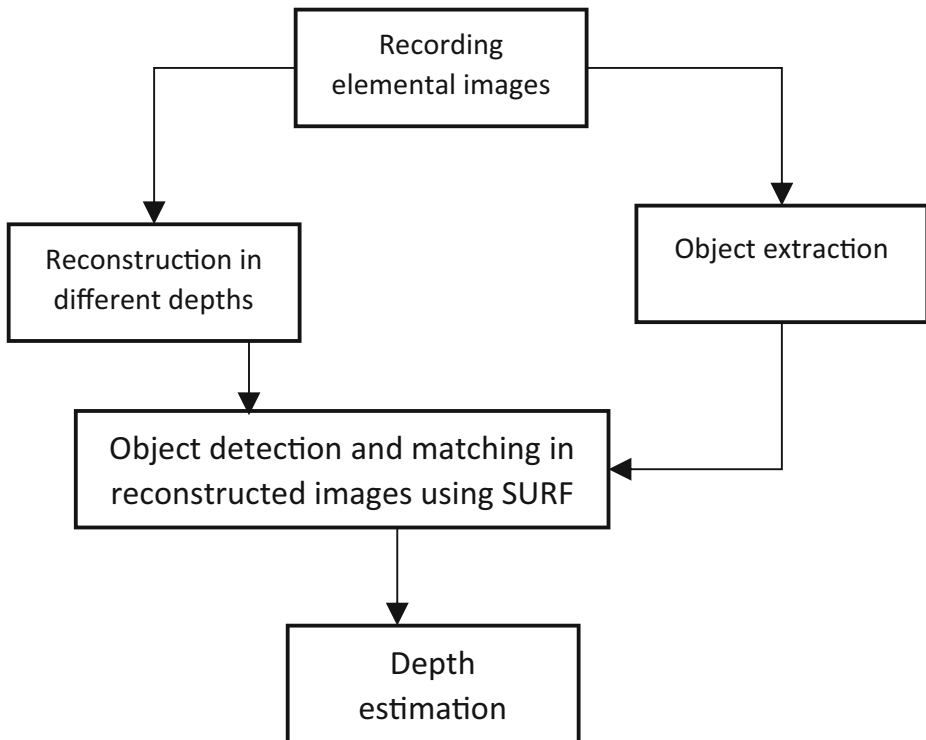


Fig. 5 An overview of the proposed method

### 3.1 Recording elemental images

To capture small objects, an array of lenses, which is in front of a camera can be used. In this way, with a single shot, we have several images of the object from different angles. If the targets are large, instead of using small lenses, several cameras are used. The number of cameras in the imaging step depends on the parameters such as the size of the object and its depth. To arrange cameras, various arrays can be used. They can be placed in a horizontal line parallel to the object, in a line perpendicular to the object, square array, or other structures. In this paper, cameras are arranged linearly. This array is a suitable choice for crowd environments, because it allows us to see all the details of the objects in the environment which may be occluded by the other.

The calculation of this depth estimation method is related to the number of elemental images. The more increase in the number of elemental images, the more time it takes to depth estimation. In similar methods, reconstruction has been done using at least seven elemental images [16]. To reduce the calculations, we have reduced the number of cameras and consequently the number of elemental images. We need at least two cameras to estimate the depth, but exact information about the depth will not be available with this number. As a result, we have gradually increased the number of cameras, and by examining their reconstructed images according to our algorithm, we have arrived at the conclusion that we can achieve acceptable accuracy with five elemental images and still have proper computational time. In arranging the cameras, we must notice that the



distance between the cameras should be such that all the objects we want to estimate their depth are in their viewing angle. We've connected the cameras to ensure the distance between them is the same in different imaging scenarios. In fact, with 5 cameras, we are able to capture all parts of the environment. Therefore, in the proposed method, five cameras with linear arrangement capture elemental images as shown in Fig. 6.

### 3.2 Object extraction

To automatically detect the depth of the objects in the environment, those objects must first be identified in the captured images. Therefore, our goal is to extract the objects in the images and separate them from the background. Background separation is the basic idea in many object extraction algorithms. In the usual background separation method, the image of the environment containing the objects is subtracted from the background image. Then, using a simple threshold on the resulting image, the objects will be recognized [39]. In this paper, we identify the objects in the environment using this method.



**Fig. 6** Image recording in the proposed method. In this paper, linear camera arrangement is used

### 3.3 Reconstruction in different depths

After recording elemental images, a computational reconstruction algorithm, as discussed in Section 2, generates a set of 3D slice images in different depths. The depth of the image with the highest sharpness of the object is considered as the depth of the object. In other words, consider an infinitesimal point source at  $(x, y, z_0)$ . In the pickup stage, this voxel of 3D object maps into several pixels in the display plan of the lenslet array. And in the reconstruction process, to reconstruct the depth image at an arbitrary depth distance of  $z$ , we find the corresponding voxels for all of these pixels. Then, we assign the intensity value of each pixel to the corresponding voxel. If we consider the rays emanating from these pixels as shown in Fig. 7, they will converge at the point  $(x, y, z_0)$  and then continue their path as a diverging beam. Therefore, at reconstruction planes  $z < z_0$  or  $z > z_0$  the computed voxels will not coincide in the 3D scene, leading to a blurred image of the voxel  $(x, y, z_0)$ . However, they will coincide at the reconstruction plane  $z = z_0$  and produce a sharp image of the original voxel. The accuracy of depth estimation is equal to the step between the distances selected for reconstruction.

### 3.4 Automatic depth estimation

Our purpose is to find the objects of the environment in each reconstructed image automatically. To this end, we have benefited from SURF matching technique [8, 9]. This approach uses a fast Hessian detector and gradient-based descriptor. The Hessian-based BLOB detector, which is used to find the points of interest, approximates the Difference of Gaussian (DoG) with box filters. What's more, for orientation assignment, wavelet responses in both horizontal and vertical directions are used by applying adequate Gaussian weights. Also, for feature description, SURF uses wavelet responses. A neighborhood around the vital point is selected and divided into subregions. Then, to get the SURF feature descriptor, for each subregion, the wavelet responses are taken and represented. The sign of Laplacian computed in the detection

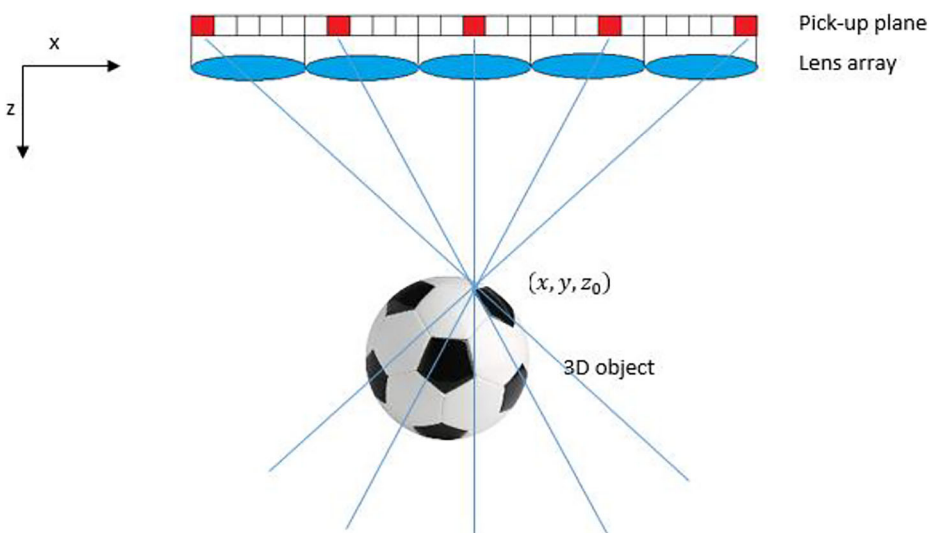


Fig. 7 The 3D reconstruction of an infinitesimal point source in integral imaging

step is used for detecting interest points. To speed-up matching process, the features are compared only if they have the same type of contrast (based on the sign) [29]. Therefore, the SURF features of the objects in the scene are detected and matched with the SURF features of reconstructed images. In this way, the reconstructed image, which has the highest clarity of object, can be found automatically. Finally, the depth of the object will be the depth in which the image with the highest clarity is reconstructed. Comparing the match rate of three popular image matching techniques, i.e., Scale-Invariant Feature Transform (SIFT) [11], SURF and Oriented FAST and rotated BRIEF (ORB) [43], shows the best performance in detection is in most cases related to the SIFT feature, but this method is much slower than SURF and ORB [43]. Therefore, we used SURF features to accelerate the depth estimation process and reduce the computation time. The validity of our approach in finding reconstructed images, in which the objects have the highest sharpness has been demonstrated by experimental results.

## 4 Experimental setup

To estimate the depth based on the integral imaging method, we captured the environment using five cameras. Since the 3D image disparity occurs mostly along the horizontal rather than the vertical direction for the objects considered in this work, we capture the environment using a linear array of cameras. Instead of using multiple cameras, different viewpoint images can be captured by shifting a single camera to different positions. However, an array of five cameras have been selected such that the third camera is oriented at line-of-sight to the first object with two cameras at its right side and two at its left side. As depicted in Fig. 3, the x-axis is chosen to be the camera line, the z-axis to be the optical axis of the cameras, and the y-axis to be perpendicular to the table's surface. The specification of the cameras is given in Table 1. Then computational reconstruction was performed at various depths. In the following, we recognized the objects of the environment in the reconstructed images using the initial images. All programs have been implemented in MATLAB software.

### 4.1 Image database

To evaluate the proposed method, we employed it to estimate the depth of objects in several different scenarios. Two of these scenarios are shown in Fig. 8. In the first scenario, five objects with different surfaces and shapes are placed at a distance of 50 to 285 cm from the camera. In this environment, a small cube covers the part of the white bottle, and a ball covers the red bottle. Therefore, this environment is occluded. In addition, the red bottle surface is non-Lambertian. In the second environment, six letters are written on cards and placed at 60, 105, 160, 205, 248, and 395 cm from the camera array, respectively. The card with the letter “B” is fully occluded by the card of the letter “F”. This scenario has been designed to assess

**Table 1** Camera parameters

Camera specifications	Nikon D7200
Focal length	18 mm
Lens size	23.4 mm × 15.7 mm
Image size	540 × 810



**Fig. 8** Two scenarios used to evaluate the proposed method. **a** In the first scenario five different objects are placed in front of an array of cameras. **b** In the second scenario we estimate the depth of 6 different cards

the accuracy of the proposed method in reconstructing corners, curves, and lines in scenes with different levels of occlusion.

## 4.2 An example

In this section, we will explain the proposed method with an example. Figure 9 shows elemental images of six cards within 60 to 395 cm of the camera array.

Figure 10 shows examples of reconstructed images in different depths of the environment shown in Fig. 9. Consider reconstructed image in-depth 60 cm in Fig. 9c. As shown, this image contains an in-focus image of the card of “F” letter and a blurred effect of all other objects.

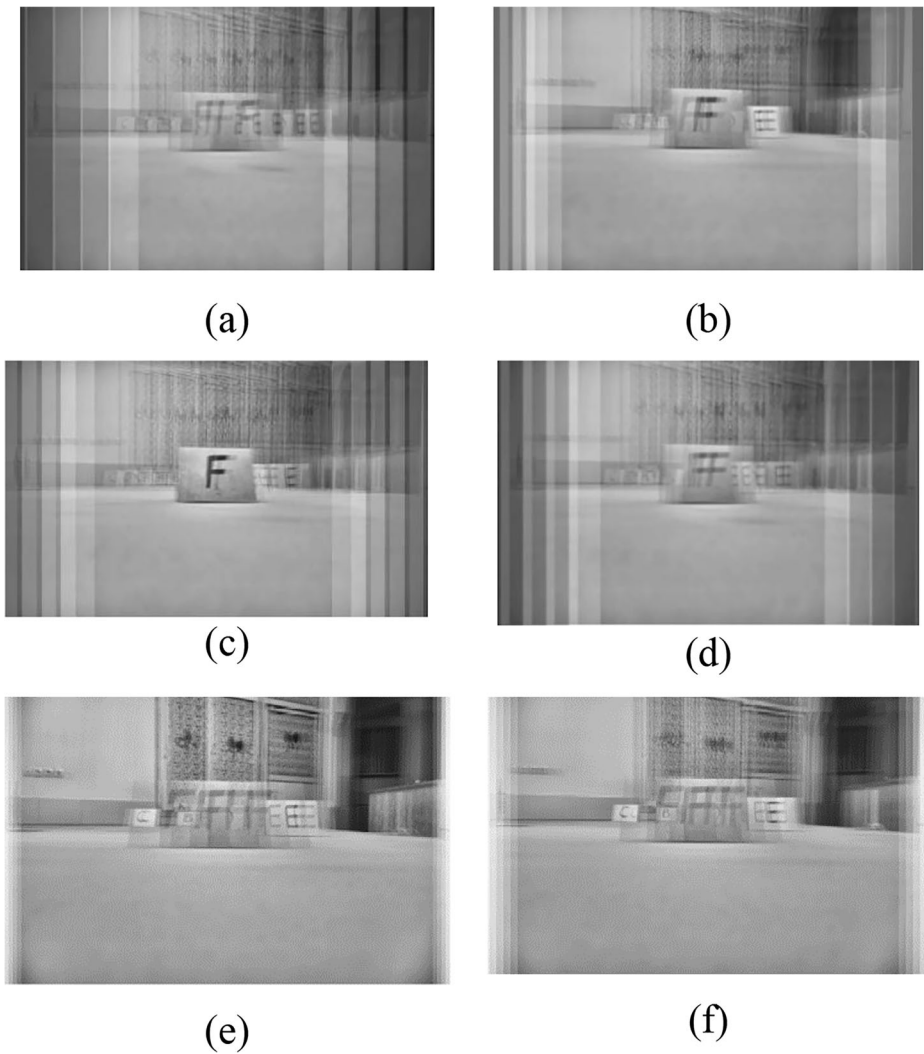
Now we extract the cards from the reconstructed images as described in Section 3.2. The cards containing letters “A”, “B”, “C”, “D”, and “F” are identified in Fig. 11d by subtracting the background image from one of the elemental images.

In this step, we examine reconstructed images to find the in-focus image of the detected objects. Figure 10 shows a set of reconstructed images of different depths in our second scenario. To find the depth of the F letter, the image which has an in-focus F letter in the image set of Fig. 10 should be found. To achieve this goal, SURF features on the image of F letter are detected then the SURF features are also searched in the collection of images in Fig. 10. In the following, the most similarity between the features of image 12(a) with the features found in the image collection is searched. As shown in Fig. 12c, the most similarity was found in a picture reconstructed at a depth of 65 cm. As a result, the depth of the card on which the letter F is inserted is equal to 65 cm.

In this way, we identify the images in which the objects have the highest sharpness and estimate the objects depths. We also follow the algorithm steps outlined in this section for our



**Fig. 9** Five elemental images captured to extract depth information of six cards



**Fig. 10** The reconstruction results of elemental images of Fig. 8, **a** to **f**: in six depths of 35, 55, 60, 70, 95, and 110 cm

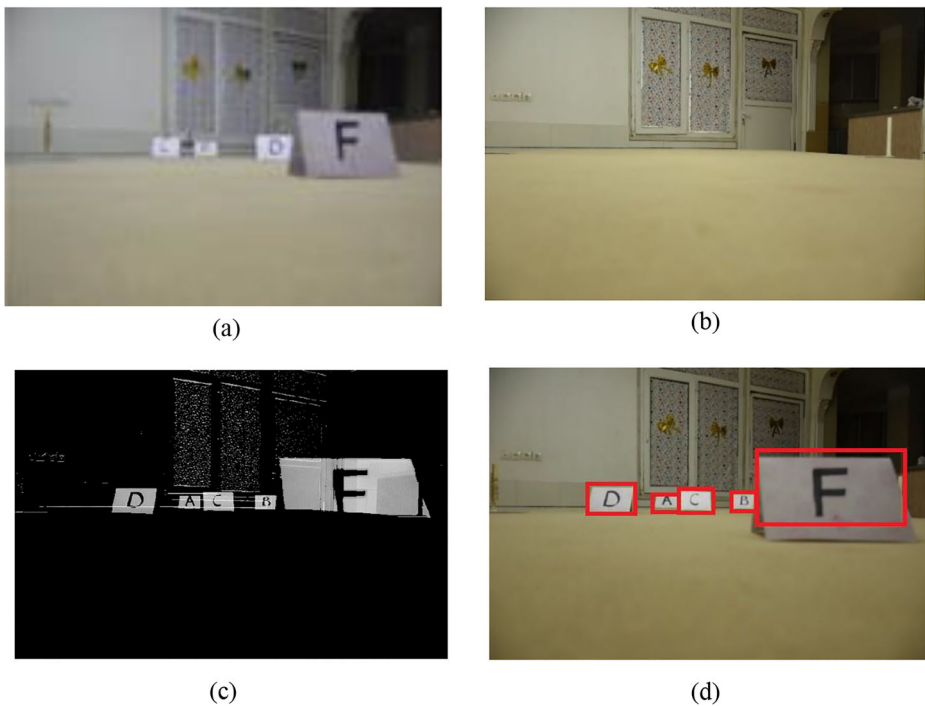
second scenario. Figure 13 shows the reconstructed images in which each object is in-focus. These images are selected through matching using SURF features.

### 4.3 Evaluation method

To evaluate the proposed method, we calculate the percentage of the average relative error (9).

$$E_r = (|D_{Pr} - D_{Re}| / D_{Re}) * 100 \quad (9)$$

where  $E_r$  is the relative error percentage,  $D_{Pr}$  is the depth estimated by the proposed method and  $D_{Re}$  is the actual depth value.



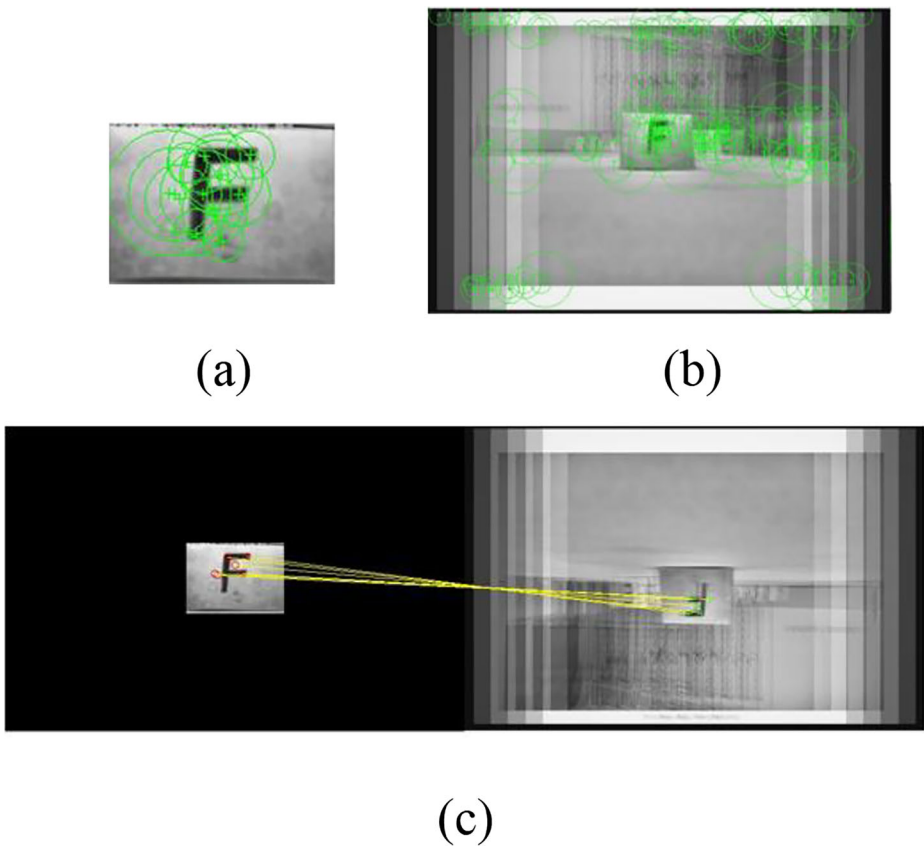
**Fig. 11** Object extraction details. **a** One of the elemental images, **b** background image, **c** after subtracting the background image from the elemental image, **d** detected objects

#### 4.4 The effect of camera numbers on the depth estimation

In this section, we will investigate the effect of changes in the number of cameras on the accuracy of the proposed method. To this end, we change the number of cameras in each of the scenarios of the database from 2 to 10 and calculate the depth of the objects with the proposed method. In each scenario, we calculate the average estimation error of the proposed method for all of the objects and as a result the average accuracy of the proposed method. Figure 14 shows the average accuracy of the proposed method versus the number of cameras in both scenarios. As expected, the accuracy of depth estimation falls with a decrease in the number of viewpoint images. Still, even in the case of heavy occlusion of our environment, the depth estimation accuracy is more than 80% for the number of cameras, being as low as 5. On the other side, there is a tradeoff between the number of elemental images and the computational time, therefore, we can achieve acceptable accuracy with 5 elemental images and still have acceptable computational time.

#### 4.5 Results

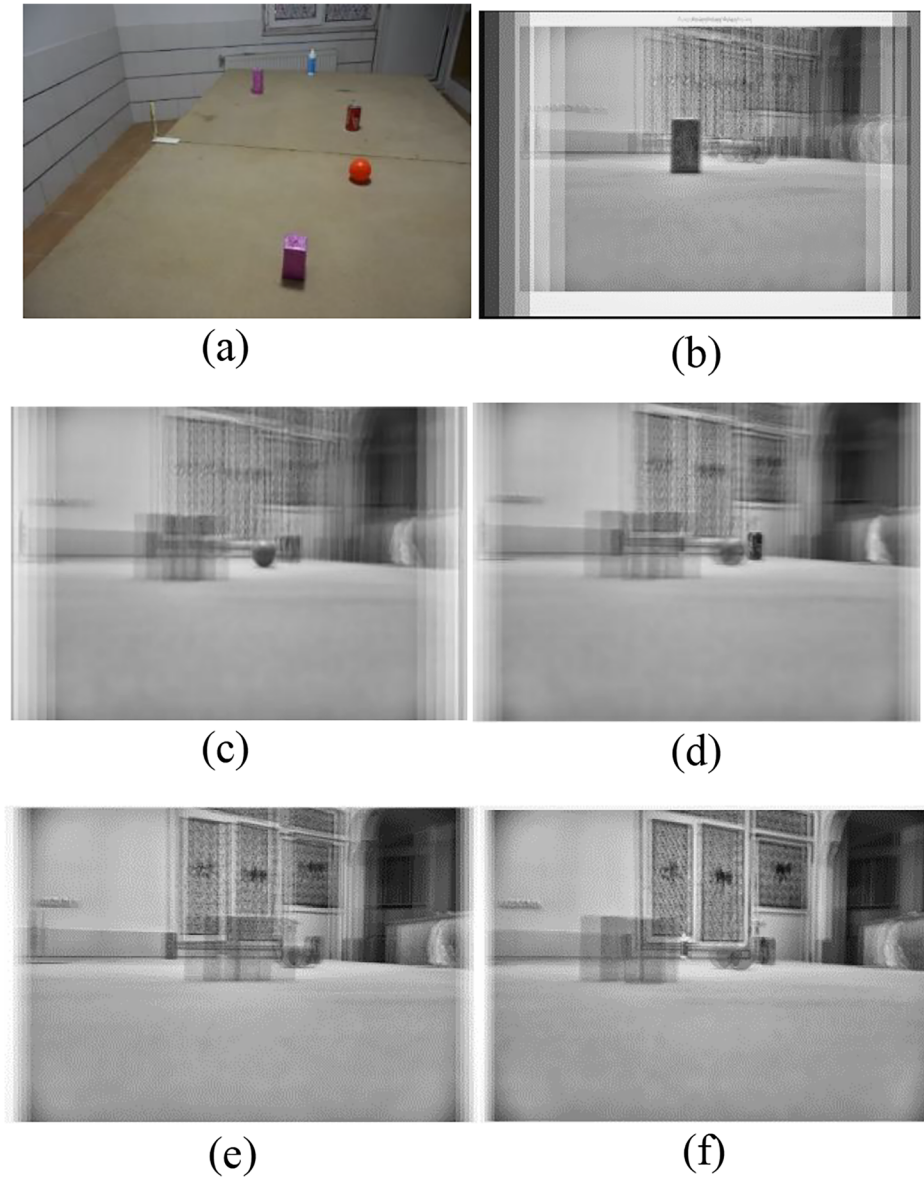
Our query set includes 11 objects of various types. This is composed of six 2D alphabetical patterns of “A,” “B,” “C,” “D,” “E,” “F,” and “E” with  $540 \times 810$  pixels each located at 60, 105, 160, 205, 248 and 395 cm from the origin of a pinhole array, respectively. In addition, our query set includes five 3D objects with different shapes. The dimension of the pinhole array



**Fig. 12** Matching the SURF features in the reconstructed images and the image of “F” letter. **a**, **b** the SURF features on the image of F letter and reconstructed image, respectively. **c** Detecting F letter in reconstructed image of depth 65 cm

used is given by  $1 \times 5$  pinholes, and the distance between the pickup plane and the pinhole array is set at 18 mm. In each environment, a total number of 351 discrete frames of the scene are reconstructed at the output plane ranging from  $z = 50$  cm to  $z = 400$  cm with the increment of  $z = 1$  cm between each frame. Therefore, estimated depth is a natural number between 50 and 400. Some of the final reconstructed plane images are shown in Fig. 13. If we decrease the reconstruction step, the accuracy of the depth estimation will increase, but this will increase the computation and the depth estimation time too.

The experimental results are summarized in Table 2, where the estimated depth of 11 objects is reported. The results shows that the proposed method yields an error rate of 5.53% to 10.35% for depth estimation in real environments. In the article [36], which has the best performance in estimating depth using an integral imaging based technique, the accuracy is reported to be 83.98% to 93.71% but we have achieved an accuracy of 80% to 97.47%. This indicates that the maximum accuracy has been improved. Table 3 shows the high error and mean error values associated with the depth estimation methods. Results show that the high error of the proposed method is better than the Min-variance method and close to the max-voting while we have a lower mean error. Therefore, the proposed method has been able to provide better accuracy while it doesn't have the limitations of previous methods.

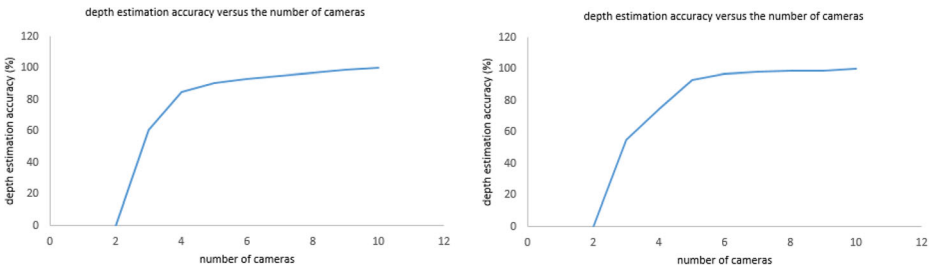


**Fig. 13** Reconstructed images in which desired objects are in-focus. **a** Image of occluded objects and reconstructed slice images at **b** 50 **c** 100 **d** 150 **e** 200, and **f** 285 cm

## 5 Conclusions

A novel depth extraction method for 3D objects, behind occlusions, based on computational integral imaging has been presented. Unlike most previous methods, which are proposed to estimate the depth of small objects that are located near the camera array, in this method, we estimate the depth of large objects at a far distance from the camera array. For depth estimation, we first record several images from different views using an array of cameras.





**Fig. 14** Depth estimation accuracy versus the number of cameras in the **a** first scenario and **b** second scenario of the database

**Table 2** Experimental results in two different scenarios

First environment			Second environment		
Real depth (cm)	Estimated depth(cm)	Error (%)	Real depth(cm)	Estimated depth(cm)	Error (%)
50	43	14	60	65	8.33
100	80	20	105	115	9.52
150	135	10	160	165	3.12
200	195	2.5	205	215	4.87
285	270	5.26	248	260	4.83
			395	405	2.53
Average error		10.35			5.53

Then we use the integral imaging algorithm to reconstruct the environment at different depths. The necessary information for estimating the range of an object of interest can be found in the 3D reconstructed images. A reconstructed image at depth  $z_0$  contains an in-focus image of the object initially located at  $z_0$  and a blurred effect of all other objects. To evaluate the performance of the proposed method, various scenarios have been investigated. Two groups of images acquired with a  $1 \times 5$  camera array, were used to test the performance of the depth estimation method. In addition, a measure was used to assess the quality of the depth estimation results. The experimental results demonstrate that using this method, the precision of depth estimation in occluded scenes increases. At the same time, there are no limitations such as sensitivity to gender and type of the object surface and its size. Future research includes using this method to estimate the speed of vehicles. The vehicle ‘s depth (car-to-camera distance) is calculated over a time interval. The distance traveled by the vehicle is equal to the difference of estimated depths at the beginning and end of this time interval. The vehicle speed will be calculated as the distance traveled divided by the time elapsed.

**Table 3** High error and mean error of the 3 methods

	High error (%)	Mean error (%)
Min–variance [12]	54.13	32.3
Max–voting [36]	16.12	11.17
Proposed method	20	7.94

## Declarations

**Competing interests** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Aloni D, Stern A, Javidi B (2011) Three-dimensional photon counting integral imaging reconstruction using penalized maximum likelihood expectation maximization. *Opt Express* 19:19681–19687
2. Ashutosh S, Schulte J, Ng AY (2007) Depth estimation using monocular and stereo cues. *IJCAI* 7:2197–2203
3. Aslantas V (2007) A depth estimation algorithm with a single image. *Opt Express* 15(8):5024–5029
4. Axelsson P (1999) Processing of laser scanner data—algorithms and applications. *ISPRS J Photogramm Remote Sens* 54:138–147
5. Badenko V, Fedotov A, Vinogradov K (2018) Algorithms of laser scanner data processing for ground surface reconstruction. In: International conference on computational science and its applications, pp 397–411
6. Bae J, Yoo H (2019) Review and comparison of computational integral imaging reconstruction. *Int J Appl Eng Res* 14(1):250–253
7. Baek E, Ho Y (2017) Real-time depth estimation method using hybrid camera system. *Electron Imaging* 5: 118–123
8. Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. *European conference on computer vision*. Springer, Berlin, pp 404–417
9. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). *Comput Vis Image Underst* 110(3):346–359
10. Cho M, Javidi B (2009) Computational reconstruction of three-dimensional integral imaging by rearrangement of elemental image pixels. *J Disp Technol* 5(2):61–65
11. Cruz-Mota J, Bogdanova L, Paquier B, Bierlaire M, Thiran JP (2012) Scale invariant feature transform on the sphere: theory and applications. *Int J Comput Vis* 98(2):217–241
12. Daneshpanah M, Javidi B (2009) Profilometry and optical slicing by passive three-dimensional imaging. *Opt Lett* 34(7):1105–1107
13. Diebel J, Thrun S (2005) An application of Markov random fields to range sensing. *Proc. of Advances in Neural Information Processing systems*, pp 291–298
14. Ens J, Lawrence P (1993) An investigation of methods for determining depth from focus. *IEEE Trans Pattern Anal Mach Intell* 15(2):97–108
15. Furukawa Y, Hernández C (2015) Multi-view stereo: a tutorial. *Found Trends Comput Graph Vis* 9(1–2):1–148
16. Ghaneizad M, Aghajan H, Kavehvash Z (2016) Three-dimensional reconstruction of heavily occluded pedestrians using integral imaging. *Proceedings of the 10th international conference on distributed smart camera*. ACM, pp 1–7
17. Ghaneizad M, Kavehvash Z, Mehrany KH, Tayaranian Hosseini SM (2017) A fast bottom-up approach toward three-dimensional human pose estimation using an array of cameras. *Opt Lasers Eng* 95:69–77
18. Godard C, Mac Aodha O, Brostow G J (2017) Unsupervised monocular depth estimation with left-right consistency. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 270–279
19. Greengard A, Schechner Y, Piestun R (2006) Depth from diffracted rotation. *Opt Lett* 31:181–183
20. Gur S, Wolf L (2019) Single image depth estimation trained via depth from defocus cues. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 7683–7692
21. Hazirbas C, Soyer S G, Staab M C, Leal-Taixé L, Cremers D (2018) Deep depth from focus. In: *Asian conference on computer vision*, pp 525–541
22. He W, Hu J, Zhou C, Li C, Li W, Han Y (2018) A hierarchical model for embedded real-time stereo imaging. In: *High-Performance Computing in Geoscience and Remote Sensing VIII* 10792: 1079209. International Society for Optics and Photonics
23. Hong S-H, Jang J-S, Javidi B (2004) Three-dimensional volumetric object reconstruction using computational integral imaging. *Opt Express* 12(3):483–491
24. Hong S, Shin D, Lee B, Kim E (2012) Depth extraction of 3D objects using axially distributed image sensing. *Opt Express* 20(21):23044–23052

25. Inoue K, Cho M (2019) Enhanced depth estimation of integral imaging using pixel blink rate. *Opt Lasers Eng* 115:1–6
26. Izzat I, Li F (2015) Stereo-image quality and disparity/depth indications. U.S. Patent 9: 030-530
27. Javidi B, Moon I, Yeom S (2006) Three-dimensional identification of biological microorganism using integral imaging. *Opt Express* 14:461–468
28. Jeon H-G, Park J, Choe G, Park J, Bok Y, Tai Y-W, Kweon IS (2015) Accurate depth map estimation from a lenslet light field camera. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pp 1547–1555
29. Karami E, Prasad S, Shehata M (2017) Image matching using SIFT, SURF, BRIEF and ORB performance comparison for distorted images. *CoRR arXiv: 1710.02726*
30. Komatsu S, Markman A, Javidi B (2018) Optical sensing and detection in turbid water using multidimensional integral imaging. *Opt Lett* 43:3261–3264
31. Lee M-C, Inoue K, Kim C-S, Cho M (2016) Regeneration of elemental images in integral imaging for occluded objects using a preoptic camera. *Chin Opt Lett* 14(12):101–121
32. Levin A, Fergus R, Durand F, Freeman WT (2007) Image and depth from a conventional camera with a coded aperture. *ACM Trans Graph* 26(3):70-es
33. Li X, Wang Y, Wang QH, Liu Y, Zhou X (2019) Modified integral imaging reconstruction and encryption using an improved SR reconstruction algorithm. *Opt Lasers Eng* 112:162–169
34. Lippmann G (1908) Épreuves réversibles donnant la sensation du relief. *J Phys Theor Appl* 7(1):821–825
35. Martínez-Corral M, Javidi B (2018) Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems. *Adv Opt Photon* 10(3):512–566
36. Martínez-Usó A, Latorre-Carmona P, Sotoca JM, Pla F, Javidi B (2016) Depth estimation in integral imaging based on a maximum voting strategy. *J Disp Technol* 12(12):1716–1723
37. Matoba O, Tajahuerce E, Javidi B (2001) Real-time three-dimensional object recognition with multiple perspectives imaging. *Appl Opt* 40:3318–3325
38. Moon L, Javidi B (2008) Three-dimensional visualization of objects in scattering medium by use of computational integral imaging. *Opt Express* 16:13080–13089
39. Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9: 62–66
40. Poozesh P, Sarrafi A, Mao Z, Avitabile P, Niezrecki C (2017) Feasibility of extracting operating shapes using phase-based motion magnification technique and stereo-photogrammetry. *J Sound Vib* 407:350–366
41. Qu H, Piao Y, Xing L, Zhang M (2017) Segmentation-based occlusion removal technique for partially occluded 3D objects in integral imaging system. *Appl Opt* 56(9):151
42. Ren H, Ni LX, Li HF, Sang XZ, Gao X, Wang QH (2019) Review on tabletop true 3D display. *J Soc Inf Disp* 28(1):75–91
43. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. 2011 international conference on computer vision, pp 2564–2571
44. Saadat A (1998) Formulating depth information in an image. *Sci Iran* 4(4):183–189
45. Son JY, Son WH, Kim SK, Lee KH, Javidi B (2013) Threedimensional imaging for creating real-world-like environments. *Proc IEEE* 101(1):190–205
46. Torralba A, Oliva A (2002) Depth estimation from image structure. *IEEE Trans Pattern Anal Mach Intell* 24(9):1226–1238
47. Um G, Kim K, Ahn C, Lee L (2005) Three-dimensional scene reconstruction using multiview images and depth camera. *Proc. of 3D digital imaging and modeling*, pp 271–280
48. Venkataraman K, Gallagher P, Jain A, Nisenzon S (2016) Pelican imaging Corp. Systems and methods for stereo imaging with camera arrays. U.S. Patent 9: 438,888
49. Zhang L, Nayar S (2006) Projection defocus analysis for scene capture and image display. *ACM Trans Graph* 25(3):907–917
50. Zhou C, Lin S, Nayar SK (2011) Coded aperture pairs for depth from defocus and defocus deblurring. *Int J Comput Vis* 93(1):53–72