applied optics

Effects of elemental images' quantity on three-dimensional segmentation using computational integral imaging

Doron Aloni and Yitzhak Yitzhaky*

Electro Optics Engineering Department, Ben Gurion University of the Negev, Beer-Sheva 84105, Israel
**Corresponding author: ytshak@bgu.ac.il*

Received 17 October 2016; revised 25 December 2016; accepted 8 February 2017; posted 9 February 2017 (Doc. ID 278916); published 6 March 2017

3D object detection and isolation can be achieved algorithmically using computational integral-imaging data. The 3D scene is acquired by a multi-channel system, where each channel (elemental image) captures the scene from a shifted perspective angle. The number of these channels affects the weight, the cost, and the computational load of the segmentation process, while a lower number of channels may reduce the performance of the objects' separation in the 3D scene. This research examines the effect of the elemental images' quantity on the 3D object detection and segmentation, under both regular and noisy conditions. Moreover, based on our previous works, we perform an improvement of the 3D object segmentation quality using an adapted active-contour method.

OCIS codes: (100.6890) Three-dimensional image processing; (200.4560) Optical data processing; (100.3008) Image recognition, algorithms and filters.

https://doi.org/10.1364/AO.56.002132

1. INTRODUCTION

3D object segmentation can be useful and effective in various fields of computer vision, such as robot navigation, medicine, and surveillance. A more specific application is described in Ref. [1]. In this case, a clearer scene presentation in a prosthetic vision device is achieved with 3D object segmentation.

Integral imaging [2], which was first proposed by Lippmann [3], has drawn vast attention and developments recently due to the availability of high-performance capturing and displaying systems. 3D images can be visualized by recording and processing 2D multi-view images, which are different angular perspectives of the 3D scene under incoherent illumination conditions; each image is referred to as an elemental image (EI). A computational process reconstructs confocal images of different depth planes of the scene, as described in the next section. While, as the number of EIs (channels) is higher, the 3D information is richer, the quantity of the EIs also defines the weight of the system, its cost, and its computational load. A previous work [4] performed a comparison between the integral-imaging system and the traditional stereo vision method, which uses only two channels. This comparison defines the synthetic aperture, the distance between the two farthest EIs, and the total number of pixels in all the acquired data to be equal in both methods. According to this work, the integral-imaging system has much better results in the case of 3D computational reconstruction of planes with occluded objects. However, this work used only one integral-imaging configuration, and did not suggest the optimal amount of EIs. Here, we test the effect of the EIs' quantity in the case of 3D image segmentation, under both regular and noisy conditions. Our previous works [5,6] developed a method that finds the depth locations of 3D objects directly from the gradient properties of the computationally reconstructed depth planes without an additional procedure of comparison with the recorded images, in contrast to previous works [7–11]. Then, the method segments the detected objects and produces isolated edge-type regions of them. As will be shown in the next section, the gradient value in the reconstructed depth planes depends on the blur severity, and, when we have more perspective angles of the scene, the blur is denser and smoother. On other hand, more channels yield a more expensive and heavy optical system. This method is robust in noisy imaging conditions because it uses only the summation-based reconstructed images. However, the effect of this summation depends on the number of summed perspectives. As detailed in the next section, this method provides 3D segmentation based on the assumption that the focused regions consist of the highest frequencies along the depth plane.

This paper examines the influence of the amount of EIs on the ability to detect and segment objects in the 3D scene using

this algorithm. The experiment evaluates the use of various numbers of EIs toward the lowest number of two (as in stereo imaging). In addition to this examination, we present an improvement of the accuracy of the 3D object segmentation using an active-contour method [12], with which a closed contour of the isolated objects is achieved.

The rest of the paper is organized as follows: Section 2 presents an overview of the 3D segmentation algorithm from Refs. [5,6]. In Subsection 2.A, we propose to use accurate regions at which false edges may appear in reconstructed planes, in addition to the correct edges at the most focused plane. Section 3 examines the effect of the quantity of the EIs on the object detection and segmentation quality. This effect was examined under both regular and very noisy conditions (Subsections 3.A and 3.B, respectively). Moreover, we examined the influence of the EIs' quantity on the computational load, and the effect of the EI resolution on the object localization quality, in Subsections 3.C and 3.D, respectively. In Section 4, we present another improvement to the final segmentation using the active-contour method with adaptation to our algorithm. Finally, conclusions and a summary are given in Section 5.

2. OVERVIEW OF THE 3D OBJECT SEGMENTATION

The 3D object detection and segmentation algorithm is as follows:

First, computational integral imaging [13] is performed using the matrix of the recorded EIs. The whole depth range available for reconstruction of depth planes is defined according to the depth of field (DOF) of a single EI, and the resolution of the reconstructed image is limited by the pixel size [13]. The reconstructed image of the integral-imaging system at z_j depth is [13]

$$f^{BP}(x,y,z_j)=\frac{1}{KL}\sum_{k=0}^{K-1}\sum_{l=0}^{L-1}g_{kl}\left(x+\left(\frac{1}{M_j}\right)S_x k, y+\left(\frac{1}{M_j}\right)S_y l\right), \tag{1}$$

where g_{kl} is the $K \times L$ EI array, k, l are indices for the particular EI, M_j is the magnification factor that depends on the distance between the camera and the reconstructed plane z_j, S_x, S_y are the distances between the cameras at x and y directions, respectively, and $f^{BP}(x,y,z_0)$ is 2D reconstructed image at distance z_j from the camera. The effect of the values of K and L is examined in this work.

According to [6], the 3D localization and isolation process performs high-pass filtering in order to detect the focused sharp object regions. This process is very sensitive to noise in the image. Therefore, we use only the reconstructed planes for this process, without employing the EIs (captured) that may be noisy. These planes are formed in a noise-reducing summation process as shown in Eq. (1). Furthermore, we apply the adaptive Wiener filter [14] to the reconstructed planes for smoothing the remaining noise while preserving the object's sharp structure, as follows:

$$\hat{f}^{z_j}=m_f(x,y,z_j)+\frac{\sigma_f^2(x,y,z_j)}{\sigma_f^2(x,y,z_j)+\sigma_v^{z_j 2}}\times[f^{BP}(x,y,z_j)-m_f(x,y,z_j)], \tag{2}$$

where $\hat{f}^{z_j}$ is a de-noised reconstructed image, in the distance z_j, $m_f(x,y,z_j)$ is the local mean value in a [3 × 3] sliding window, $\sigma_f^2(x,y,z_j)$ is local variance there, and $\sigma_v^{z_j 2}$ is the variance of white Gaussian noise with zero mean, assumed in the image. The noise power is estimated from regions where the standard deviation (STD) of the original image is very small (non-significant activity in the original image). Areas with high activity are detected by thresholding the image gradient magnitude. The noise power is estimated by dividing the image into overlapping square regions contained in the areas where the absolute gradient is below a threshold. According to Ref. [15], a threshold value of 4 (out of 256 levels) and a square region size of 10 pixels perform well with regular images. Each square region size yields a local noise variance, $\sigma_v^{z_j 2}$, and the noise variance in the image is estimated as

$$\sigma_v^{z_j 2}=\frac{1}{W}\sum_{w=1}^{W}(\sigma_{vw}^{z_j 2}), \tag{3}$$

where w is the index of the local noise variance in such a single square region, and W is the total number of square regions.

For the purpose of producing accurate 3D object segmentation, we propose several stages. In the first stage, we find 3D depth locations of objects in the 3D space by looking for the local higher gradient values of the reconstructed image along the depth axis [6]. The average gradient magnitude of the reconstructed plane (AGMR) is then calculated [6]:

$$\mathrm{AGMR}(z_j)=\frac{1}{N_x\cdot N_y}\sum_y\sum_x|\nabla(\hat{f}^{z_j})|. \tag{4}$$

The results of Eq. (4) yield a "depth graph." Each local maximum in a curve in this graph has shown convergence to a significant object location along the z axis, at the imaging capture instance. In the next phase, an object segmentation process for each depth location is performed. This process initially applies an adaptive threshold to the wavelet coefficients of each reconstructed plane where an object was detected [6]. These coefficients are high-pass filtered versions of the image in three directions. The adaptive threshold value, r_{T_i}, is found according to the probability density function (PDF) of each wavelet sub-band:

$$r_{T_i}=r{:}fp_{T_i}=0.995;\quad \text{where } fp_{T_i}=\frac{\int_{r_{0_i}}^{r_{T_i}}|\mathrm{PDF}_i(r)|\mathrm{d}r}{\int_{r_{0_i}}^{r_{f_i}}|\mathrm{PDF}_i(r)|\mathrm{d}r}, \tag{5}$$

where r is the sub-band gray level's index that is limited by r_{0_i} and r_{f_i}. The PDF is determined as the normalized histogram of sub-band i. In addition to the initial object isolation, the method in [6] removes false edges according to the highest gradient value of edges detected at more than a single depth plane. In order to eliminate such false edges, we search for the highest

gradient value among the different reconstructed edge-detected images that contain edges of the same object.

A. Removing False Edges in Multiple Reconstructed Planes

Because of the different blur levels along the depth, the object's edge can be found in different locations in different reconstructed images. The point spread function (PSF) region from the focused area to the blurred area depends on the depth distance between them:

$$\mathrm{KER}_{\mathrm{blr}} = \left[\left|\frac{S_x(K-1)}{M_j} - \frac{S_x(K-1)}{M_{j+1}}\right| \times \left|\frac{S_y(L-1)}{M_j} - \frac{S_y(L-1)}{M_{j+1}}\right|\right], \tag{6}$$

where M_j and M_{j+1} are the magnification factors according to depth planes z_j and z_{j+1}, respectively. This PSF region gives the area in which a detected edge may shift from one depth plane to the other. The following procedure eliminates the false edges. This approach is an improvement to [6]. In this case we used the depth-adaptive PSF region as calculated in Eq. (6):

$$\begin{aligned}&\text{Phase 1: find}(x_t, y_t) \rightarrow \quad \text{for } \{\cap(F_p * \mathrm{KER}_{\mathrm{blr}}, F_{p+1})\},\\ &\text{Phase 2: } Tp = \left\{\sum_{i=1\rightarrow 3} |r_{T_i}(x_t, y_t)|\right\},\\ &\text{Phase 3: } F_p(x_t, y_t) = 1 \quad \text{if } \max_p(Tp); \text{else } F_p(x_t, y_t) = 0,\end{aligned} \tag{7}$$

where x_t, y_t are the coordinates that include common edges along the depth axis at planes associated with all the local peaks p is the peak index.

3. CHANNEL QUANTITY'S EFFECT

This section presents the effect of the amount of multi-perspective recording channels (the EIs) on the 3D segmentation. The synthetic aperture in all the cases is equal, in order to preserve the depth resolution. The effect of the channels' amount [the values of K and L in Eq. (1)] is tested at both regular and noisy conditions. For both cases we checked the results of the depth graph and the final segmentation. In order to examine the robustness of the algorithm, we present two very different scenes. The first one, as presented in Fig. 1(a), includes two toy cars with uniform background, where the first toy is located 54 [cm] from the camera, and the second stands at 62 [cm] from it. The second scene, as presented in Fig. 1(b), includes two cups that are located 60 [cm] from the camera, in the front plane, and a very rich background that includes various objects near the wall. In both cases, the effective focal length was 20 [mm], and the synthetic aperture size was 30 mm × 30 mm.

(a) (b)

Fig. 1. Elemental images of two different examined scenes.

The examined amounts of EIs were: 5 × 5, 4 × 4, 3 × 3, 2 × 2, and 1 × 2. Note that a higher number of EIs produces results that are quite similar to the 5 × 5 case. In Subsection 3.A, we examine the channel quantity's effect under regular illumination conditions, and in Subsection 3.B the results are examined for images with simulated severe noise conditions.

A. Operations under Regular Conditions

As mentioned above, in this section we perform 3D segmentation according to Section 2. The 3D scenes are shown in Figs. 1(a) and 1(b).

In the first phase, the depth graphs according to Eq. (4) are yielded. Figures 2(a) and 2(b) show plots of the depth graphs for the locations of the 3D objects, for different numbers of EIs, where Fig. 2(a) is obtained from the scene of Fig. 1(a) and Fig. 2(b) is from the scene of Fig. 1(b). We define the STD of the depth graph as a measure of the robustness of the depth localization procedure, because it gives higher values as the peaks of the graph (object locations) are more prominent from their surroundings. Figure 3 shows the resulting STD values for both depth graphs in Fig. 2. Each STD graph in this figure is normalized to a maximum value of 1. It can be seen that as the amount of EIs is larger, the STD is increased; thus the algorithm is more robust and less sensitive to noises. It can be seen that the STDs have sharp drops from the case of 3 × 3 EIs toward lower

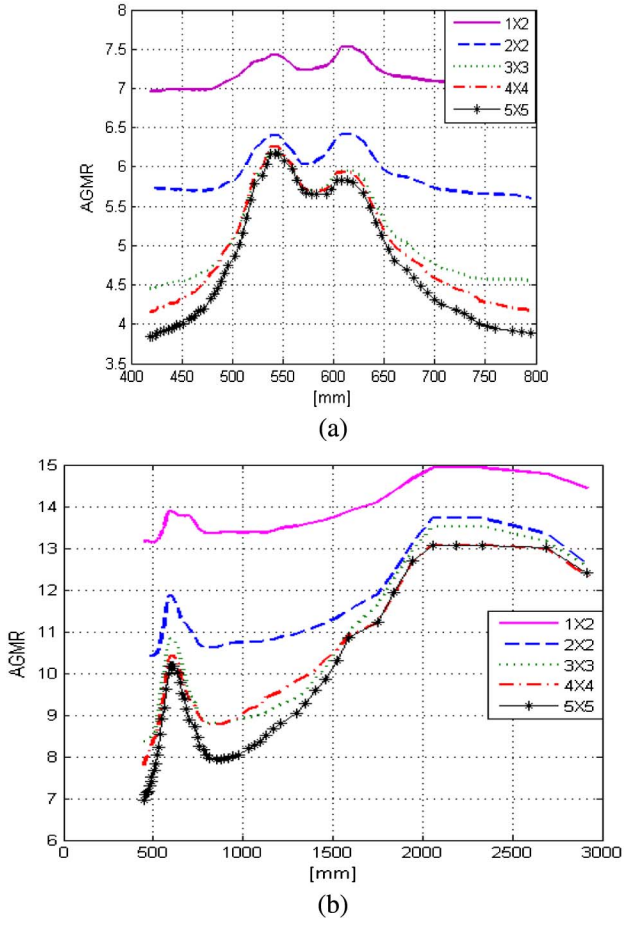


Fig. 2. Depth graphs for different numbers of elemental images. (a) for the scene in Fig. 1(a), and (b) for the scene in Fig. 1(b).

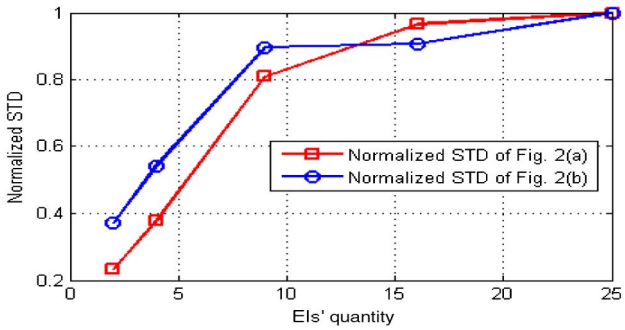


Fig. 3. STD values of the depth graphs as functions of the EIs' quantity. Each STD graph is normalized to a maximum value of 1. It can be seen that when the amount of EIs is higher, the STD is higher; thus the object depth detection is more robust.

numbers of EIs, while for higher numbers of EIs until 5 × 5 the values do not change much. Therefore, according to both scenes, the best configuration is 3 × 3 EIs. In this case, the algorithm is relatively robust and less sensitive to noises, while the amount of channels is relatively low. It can be seen that both cases yielded a similar tendency. The depth of the objects was detected in all the cases at the same locations. However, as the amount of EIs is higher, the distinction between the focused and the defocused regions is somewhat clearer.

Figure 4 presents the number of detected depth planes where objects exist as a function of the EIs' quantity for both scenes shown in Fig. 1. The number of such detected depth planes is according to the number of peaks in each depth graph in Fig. 2. It can be seen that in all the cases aside from the 2 × 1 case, two depth planes are detected (in the same depth locations). In the case of just two EIs, wrong depth planes are detected in addition to the correct ones. This can be explained according to the only one direction of 3D orientation instead of two in the other cases. Next, we present the segmentation edge results according to Eqs. (5)–(7). In Figs. 5 and 6 the segmentation results using different numbers of EIs are represented for the first and second scenes, respectively. From Fig. 5 it can be seen that the segmentations in the cases of 5 × 5 to 2 × 2 are very similar, but in the bottom case of 1 × 2 EIs, the segmentation includes false edges. Note that in this case, which

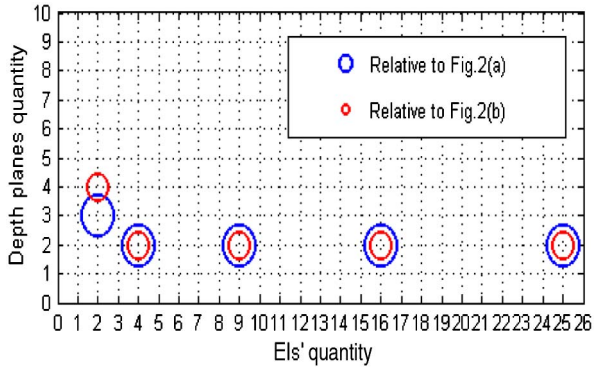


Fig. 4. Quantity of detected sharpest depth planes (depth locations of objects) as a function of the EIs' quantity according to the results from the depth graphs in Fig. 2.

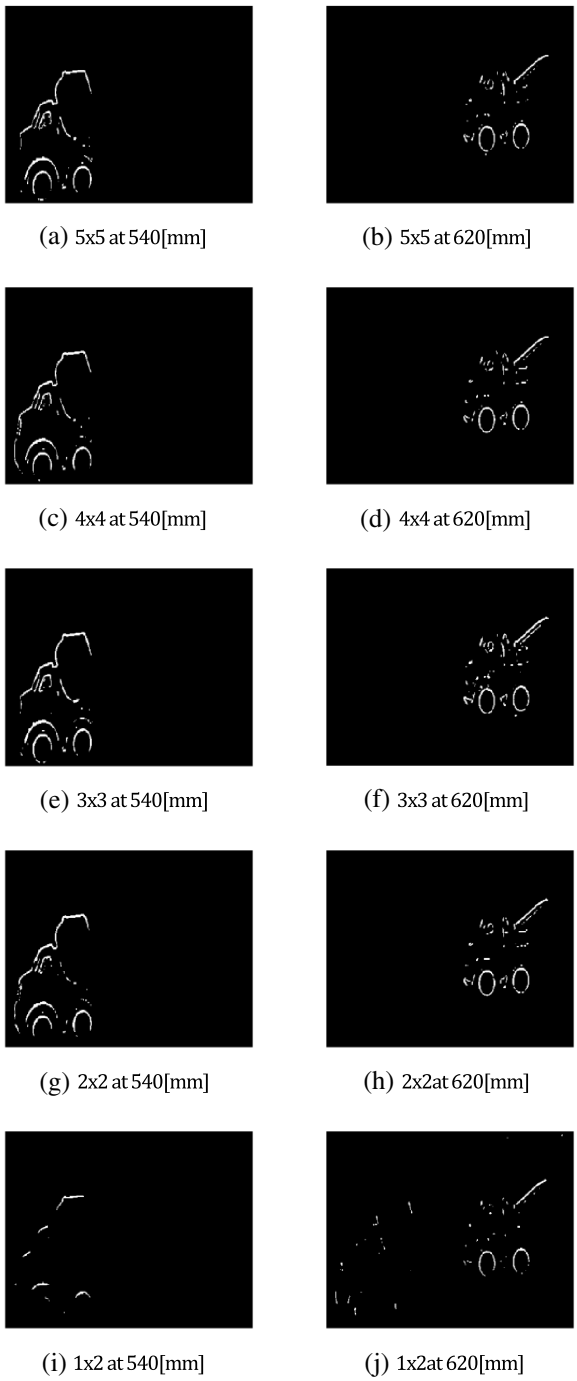
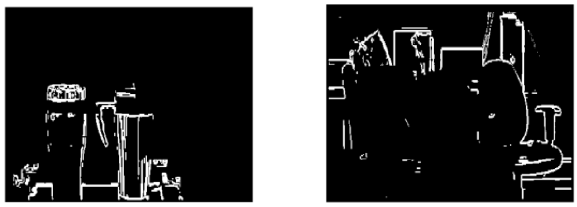


Fig. 5. Edge-based segmentations of the reconstructed images that include focused regions (the peaks in the graphs of Fig. 2) for different amounts of EIs used, for the scene shown in Fig. 1(a). (a) 5 × 5 at 540 [mm], (b) 5 × 5 at 620 [mm], (c) 4 × 4 at 540 [mm], (d) 4 × 4 at 620 [mm], (e) 3 × 3 at 540 [mm], (f) 3 × 3 at 620 [mm], (g) 2 × 2 at 540 [mm], (h) 2 × 2 at 620 [mm], (i) 1 × 2 at 540 [mm], and (j) 1 × 2 at 620 [mm].

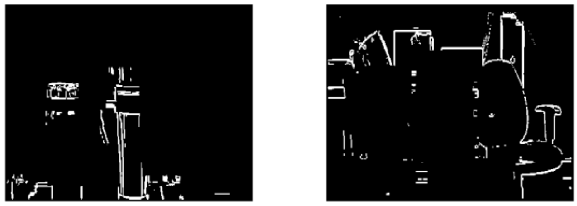
is a kind of stereo imaging, the 3D data are obtained only in the horizontal direction and false edges may be more likely to appear in the other direction. The results in Fig. 6 are quite similar, although some decrease in segmentation quality can also be seen in the 2 × 2 case.

(a) 5x5 at 600[mm] (b) 5x5 at 2000[mm]



(c) 4x4 at 600[mm] (d) 4x4 at 2000[mm]

(e) 3x3 at 600[mm] (f) 3x3 at 2000[mm]



(g) 2x2 at 600[mm] (h) 2x2 at 2000[mm]

(i) 1x2 at 600[mm] (j) 1x2 at 2000[mm]

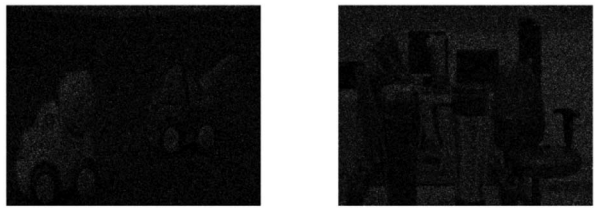
Fig. 6. Same as Fig. 5, but for the scene shown in Fig. 1(b). (a) 5 × 5 at 600 [mm], (b) 5 × 5 at 2000 [mm], (c) 4 × 4 at 600 [mm], (d) 4 × 4 at 2000 [mm], (e) 3 × 3 at 600 [mm], (f) 3 × 3 at 2000 [mm], (g) 2 × 2 at 600 [mm], (h) 2 × 2 at 2000 [mm], (i) 1 × 2 at 600 [mm], and (j) 1 × 2 at 2000 [mm].

B. 3D Segmentation under Noisy Conditions

In this section, we perform the same experiment as in the previous section, but under noisy conditions. The signal-to-noise ratio (SNR) of a single EI is calculated as

$$\mathrm{SNR}^{\mathrm{EI}} = \sigma_s^2/\sigma_n^2, \tag{8}$$

where σ_s^2 is the variance of the signal and σ_n^2 is the variance of noise, which approximates the mean square error (MSE)



(a) (b)

Fig. 7. Simulated noisy elemental images of the scenes in Fig. 1.

between the normalized EI and its noisy version (NEI) at each case, assuming additive noise [16]:

$$\mathrm{MSE} = \frac{1}{N \cdot M} \sum |\mathrm{EI} - \mathrm{NEI}|^2, \tag{9}$$

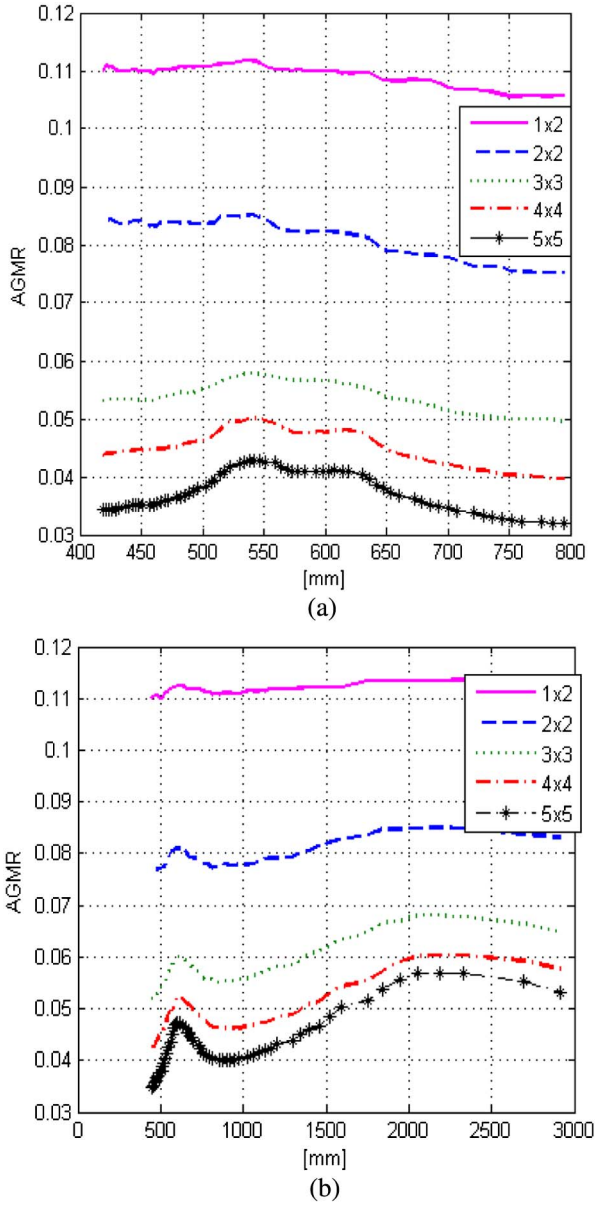


Fig. 8. Depth graphs as functions of the elemental images' quantity. (a) for the scene in Fig. 7(a), and (b) for the scene in Fig. 7(b).

where N and M are the numbers of pixels along the x and y directions.

As shown in Section 2, the 3D segmentation process is applied to the reconstructed image. In order to evaluate the SNR of the reconstructed image as a function of the EIs' quantity, we approximate the SNR in a reconstructed image, SNR^{z_j}, which is a sum of $K \times L$ noisy EIs by

$$\mathrm{SNR}^{z_j} = \sum_K \sum_L \mathrm{SNR} = KL\sigma_s^2/\sigma_n^2 = KL \cdot \mathrm{SNR}^{\mathrm{EI}}. \quad (10)$$

This approximation is quite accurate only in the focused regions at which similar locations of the $K \times L$ deterministic EIs components are summed with the uncorrelated additive noise.

Figures 7(a) and 7(b) present the simulated noisy elemental images shown in Figs. 1(a) and 1(b), respectively. In those cases, the SNR values of the EIs are 0.32 and 0.33, accordingly.

Next, the depth graphs according to Eq. (4) are presented in Figs. 8(a) and 8(b) for the noisy scenes in Figs. 7(a) and 7(b), respectively. In both cases, the algorithm detects the 3D object locations; however, as the number of EIs is decreased, more false alarms appear in the depth graphs, as will be quantified later. Moreover, as the number of EIs is higher, the detection of object depths is clearer. This is also evident from Fig. 9, where similarly to the non-noisy case, the depth detection robustness (quantified by the STD) is increased when the amount of EIs is increased.

Figure 10 presents the number of detected depth planes where objects exist as a function of the EIs' quantity for both noisy scenes shown in Fig. 7, according to the number of peaks in each depth graph. In this noisy case, for the image in Fig. 7(b), results are accurate (two detected planes) from

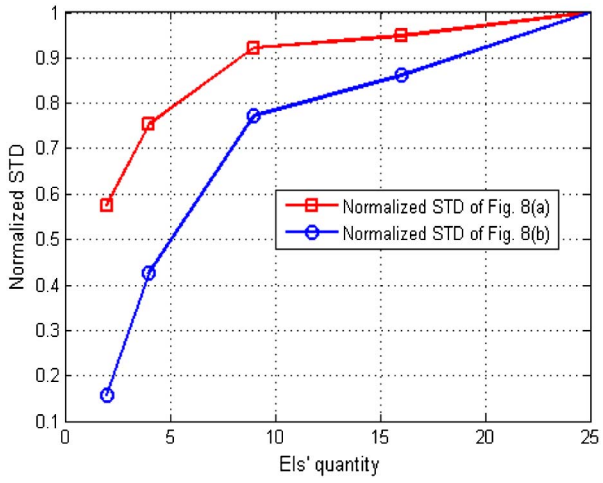


Fig. 9. Same as Fig. 3, but for the case of severely noisy images.

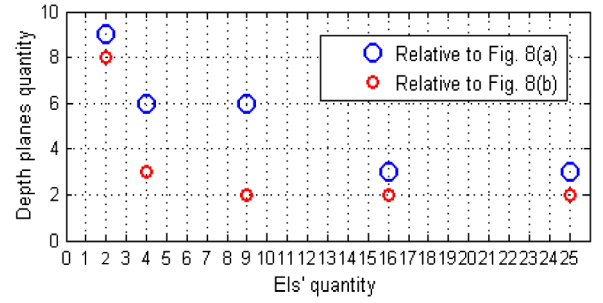


Fig. 10. Quantity of detected sharpest depth planes (depth locations of objects) as a function of the EIs' quantity according to the results from the depth graphs in Fig. 8.

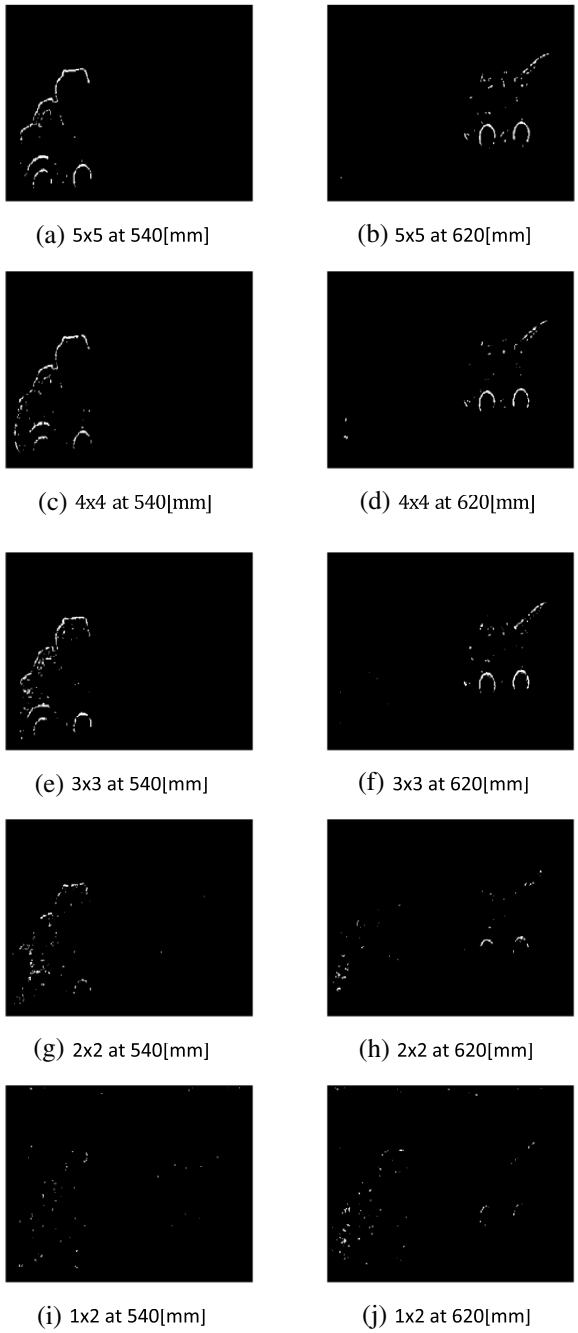


Fig. 11. Edge-based segmentations of the reconstructed images that include focused regions (according to the graphs of Fig. 8) for different amounts of EIs used, for the scene shown in Fig. 7(a). (a) 5×5 at 540 [mm], (b) 5×5 at 620 [mm], (c) 4×4 at 540 [mm], (d) 4×4 at 620 [mm], (e) 3×3 at 540 [mm], (f) 3×3 at 620 [mm], (g) 2×2 at 540 [mm], (h) 2×2 at 620 [mm], (i) 1×2 at 540 [mm], and (j) 1×2 at 620 [mm].

3×3 EIs and higher, while for the image in Fig. 7(a) false alarms always exist, but decrease as the number of EIs increases, resulting from the increase of the SNR in the reconstructed image. In the next phase, the 3D edge-based segmentation according to Eqs. (5)–(7) is presented. Figure 11 presents the results for the case of Fig. 7(a), and Fig. 12 shows the results for Fig. 7(b). In both cases, it can be seen that the quality of the segmentation is clearly lower when the amounts of the EIs are 2×2 and 1×2. In these cases, the segmentation is less clear and includes false edges. This fits the decreased SNR in the reconstructed planes as the number of EIs is decreased, according to Eq. (10).

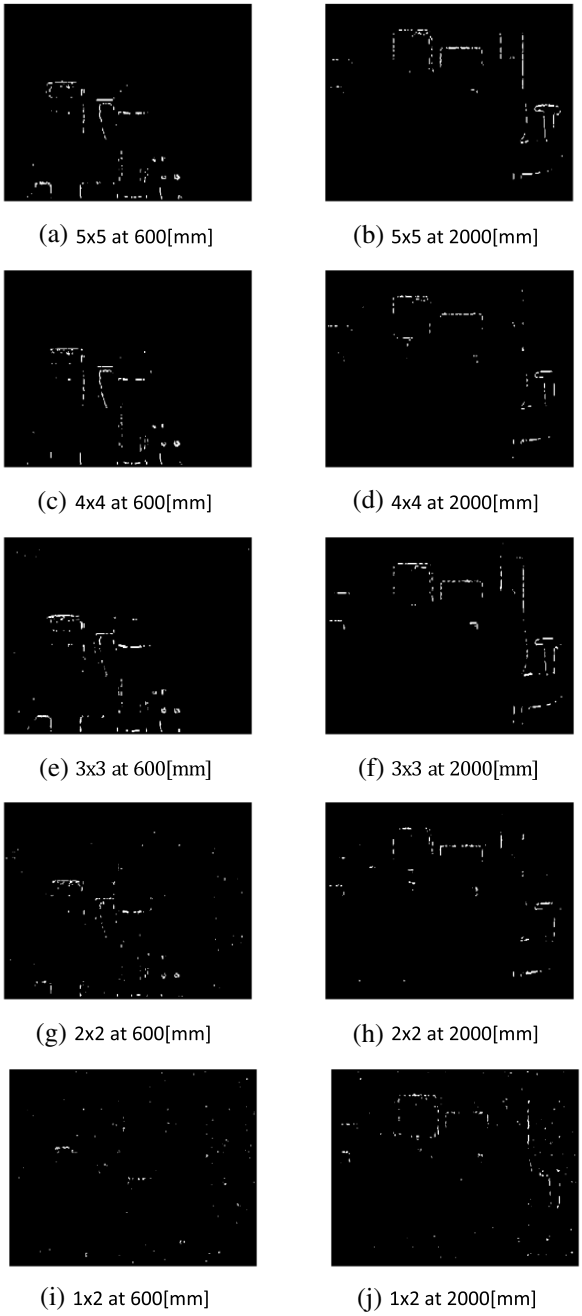


Fig. 12. Same as Fig. 11, but for the scene shown in Fig. 7(b). (a) 5×5 at 600 [mm], (b) 5×5 at 2000 [mm], (c) 4×4 at 600 [mm], (d) 4×4 at 2000 [mm], (e) 3×3 at 600 [mm], (f) 3×3 at 2000 [mm], (g) 2×2 at 600 [mm], (h) 2×2 at 2000 [mm], (i) 1×2 at 600 [mm], and (j) 1×2 at 2000 [mm].

C. Effect on the Computational Load

The effect of the EIs' quantity on the computational load depends on the amount of operations in the reconstructed image creation as presented in Eq. (1). This procedure is applied during the depth localization formation, according to Eq. (4). The other stages of the proposed algorithm do not depend on the EI's quantity, because they are applied to the reconstructed images. In Fig. 13 we present the changes of the computational load of the entire algorithm (in percentages) as a function of the EIs' quantity, where 100% is defined as the computational load without the part of the formation of the reconstructed images. It can be seen that the computational load increases approximately linearly as a function of the EIs' quantity, as expected from Eq. (1).

D. EI's Resolution Effect

In this section, we evaluate the effect of the EI's spatial resolution (number of pixels) on the segmentation operation as a function of the EIs' quantity. In Ref. [1] we have dealt with the segmentation quality as a function of the confocal image resolution, for low image resolutions.

There, the segmentation quality has been measured through object recognition by observers. It was found that the visual recognition rate is rapidly increased with increasing resolution and saturates at about 10,000 pixels per object region. In this paper, we examine the effect of the EI's resolution on the depth localization operation as presented in Eq. (4). Figure 14 shows the depth graphs for various EI resolutions.

In Figs. 14(a) and 14(b) we test the cases of 500×500, 250×250, 200×200, 150×150, and 100×100 pixels in the EI. In Figs. 14(c) and 14(d) the tested cases are: 400×600, 200×300, 160×240, 120×180, and 80×120 pixels in the EI. Moreover, the effect of the EIs' quantity in this case was also examined, by applying the algorithm using both 5×5 EIs [in Figs. 14(b) and 14(d)] and 3×3 EIs [in Figs. 14(a) and 14(c)].

It can be seen from Fig. 14 that the effect of the EIs' quantity is not strong. This supports the preferred use of the 3×3 EIs as concluded previously. Regarding the effect of the EI resolution, It can be seen in all cases that in the lower resolution EIs (upper

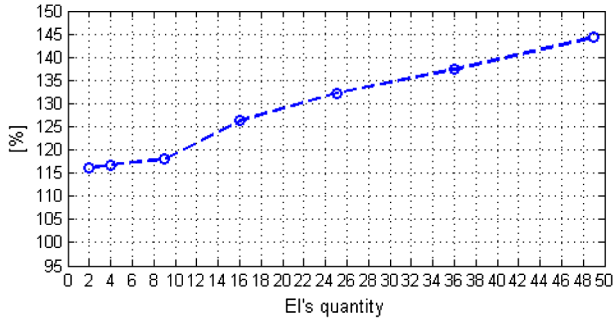


Fig. 13. Changes (in percentages) of the computational load as a function of the EIs' quantity, with regard to 100%, defined as the computational load without the reconstructed images formation part.

curves in the depth graphs) the depth planes are poorly detected comparing to the ones with higher resolution (peaks have been missed).

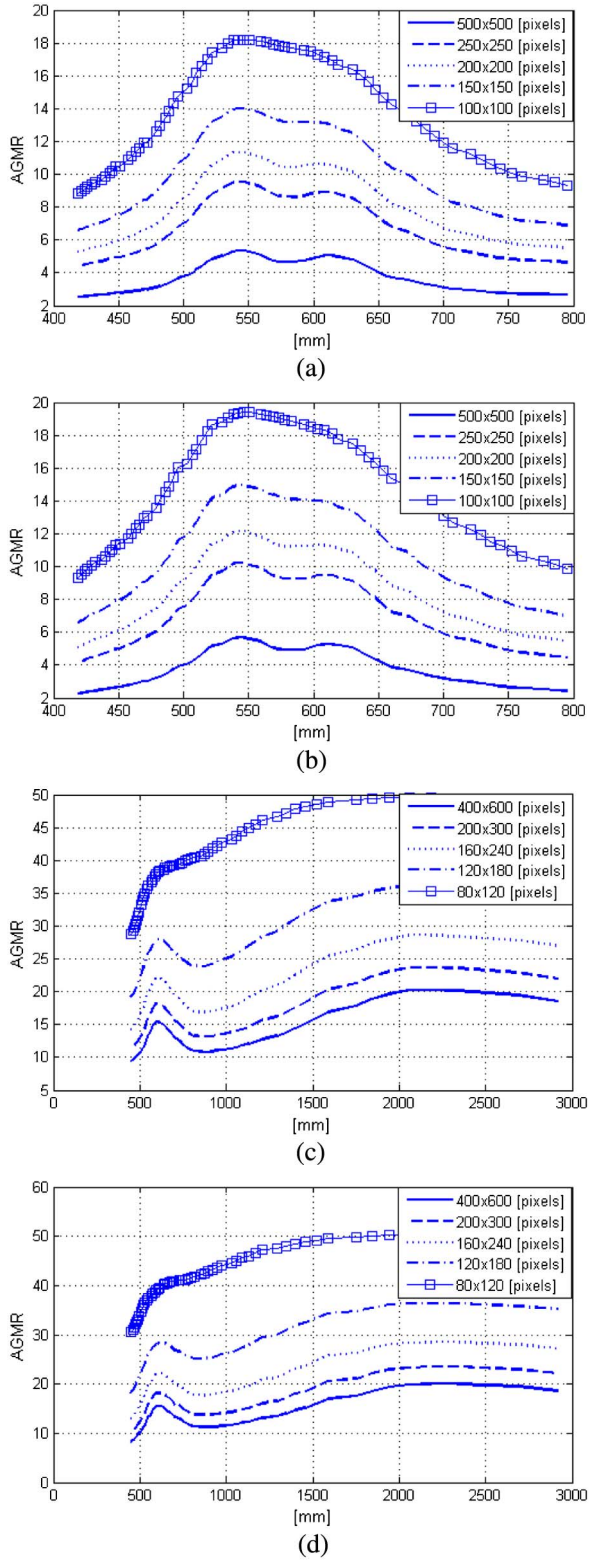


Fig. 14. Effect of the EI resolution and number of EIs on the depth graph as formed by Eq. (4). (a) and (b) are related to the scene from Fig. 1(a) with 3 × 3 and 5 × 5 EIs, respectively. (c) and (d) are the same, for the scene in Fig. 1(b).

4. SEGMENTATION IMPROVEMENT USING ADAPTED ACTIVE CONTOUR

In order to achieve a more accurate region of the detected objects, we applied the active-contour algorithm [12]. The geodesic active-contour (GAC) model is a variational model that consists of finding the contour C that minimizes the following energy functional [12]:

$$E_{\mathrm{GAC}}(C) = \int_0^{L(C)} g_b(|\nabla \hat{f}_{\mathrm{bb}}^{z_j}(C(s))|)\mathrm{d}s, \tag{11}$$

where $\hat{f}_{\mathrm{bb}}^{z_j}$ is the bounding-box's region in $\hat{f}^{z_j}$ at distance z_j, ds is the Euclidean element of length, $L(C)$ is the length of the curve C, and the function g_b is an edge indicator function that vanishes at object boundaries such that [12]

$$g_b(|\nabla \hat{f}_{\mathrm{bb}}^{z_j}|) = \frac{1}{1+\beta|\nabla \hat{f}_{\mathrm{bb}}^{z_j}|^2}, \tag{12}$$

where β is an arbitrary positive constant. Hence, the energy functional, E_{GAC}, is actually a new length obtained by weighting the Euclidean element of length ds by the function g_b, which contains information concerning the image gradient characteristic [12]. The result from this algorithm is a binary filled region. This algorithm is an iterative procedure that requires a stop condition that we define as

$$\max(f_d \cap f_{AC}^t), \tag{13}$$

where f_d is the binary edge resulting from the object detection and isolation process as presented in Eqs. (5)–(7), f_{AC}^t is the contour according to the active-contour algorithm [12] shown in Eqs. (11) and (12), and t is the iteration index.

The result of Eq. (13) is a binary image ("mask") in the region of the detected object (can be termed "mask"). Figure 15 shows the regions in the EIs of Fig. 1 covered by the masks found using the adapted active-contour process applied to the segmented images obtained according to Section 2. It can be seen that in both cases the results can be considered as richer relative to the results in Figs. 5 and 6, as they produce

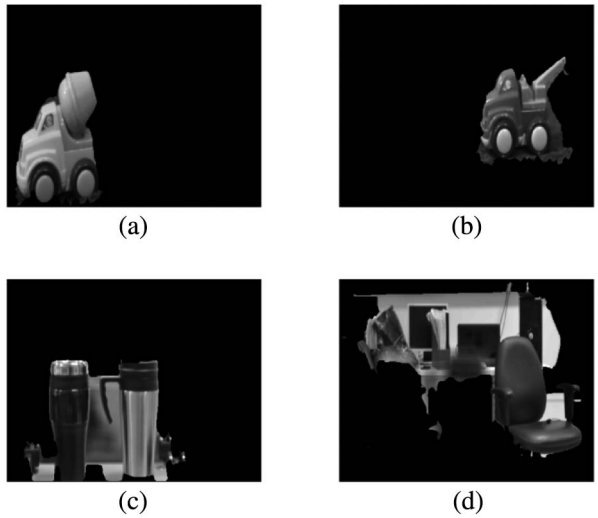


Fig. 15. Improved segmentation results using active-contour method, in addition to the basic binary-edge-based segmentation algorithm.

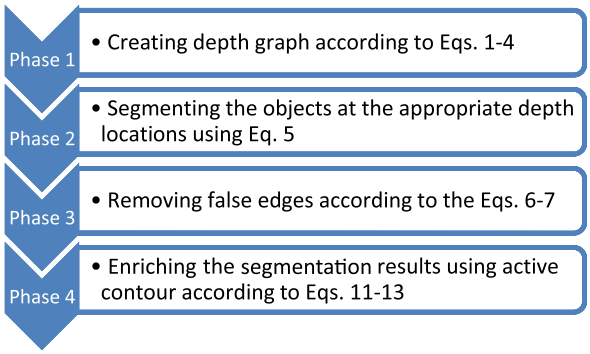


Fig. 16. Chronological flow chart of the 3D segmentation algorithm.

visual images of the detected objects, instead of a binary segmentation.

5. CONCLUSIONS AND SUMMARY

This paper examines the effect of the number of EIs on a 3D object detection and segmentation method, which employs reconstructed (confocal) images obtained by computational integral imaging. In addition, it proposes some improvements of the segmentation process. The following flow chart (Fig. 16) summarizes the segmentation algorithm chronologically.

We found that the segmentation can be achieved with only 2×2 EIs for regular images. Moreover, we showed that when using just two EIs, where the 3D data are obtained only in one direction (horizontally in this case, as in stereo imaging), the segmentation quality is decreased. From the depth graphs (Figs. 2 and 8) and their analyses we can see the effect of the EIs' quantity on the algorithm's object depth detection robustness. It is clear that in general, as the quantity is increased, the robustness is increased. However, considering both the quality of results and the system's cost and weight (increased by the number of EIs), it can be concluded from the results that for the regular case (no significant noise in the images) a number of 3×3 EIs is a good choice. In the case of severe noisy conditions, we have seen that a higher number of EIs should be used, depending on the amount of SNR in the images, because the quantity of EIs is needed not just for the separation between blurred and sharp regions, but also for the reduction of noise in the summation process in Eq. (1).

In addition, we found that using an active-contour method adapted to our case for an improvement of the final segmentation, richer results of the segmented regions can be obtained (with regard to the edge-segmentation done before). This can be very useful for object recognition.

Funding. U.S. Department of Defense (DOD), Office of the Assistant Secretary of Defense for Health Affairs, Clinical and Rehabilitative Medicine Research Program and the Joint Program Committee 8, Neurosensory and Rehabilitative Research Program (W81XWH-16-1-0033).

REFERENCES

1. J. H. Jung, D. Aloni, Y. Yitzhaky, and E. Peli, "Active confocal imaging for visual prostheses using light-field and the impact on object recognition," Vis. Res. **111**, 182–196 (2015).
2. A. Stern and B. Javidi, "Three dimensional sensing, visualization, and processing using integral imaging," Proc. IEEE **94**, 591–607 (2006).
3. G. Lippmann, "La photographie integrale," in *Comptes-Rendus* (Academie des Sciences, 1908).
4. H. Navarro, A. Dorado, G. Saavedra, A. Llavador, and M. Martinez-Coral, "Is it worth using an array of cameras to capture the spatio-angular information of a 3D scene or is it enough with just two?" Proc. SPIE **8384**, 838406 (2012).
5. D. Aloni and Y. Yitzhaky, "Detection of object existence from a single reconstructed plane obtained by integral imaging," IEEE Photon. Technol. Lett. **26**, 726–728 (2014).
6. D. Aloni and Y. Yitzhaky, "Automatic 3D object localization and isolation using computational integral imaging," Appl. Opt. **54**, 6717–6724 (2015).
7. Y. Frauel and B. Javidi, "Digital three-dimensional image correlation by use of computer-reconstructed integral imaging," Appl. Opt. **41**, 5488–5496 (2002).
8. S. Yeom, D. Lee, J.-Y. Son, and S.-H. Kim, "Three dimensional object reconstruction and recognition using computational integral imaging and statistical pattern analysis," Jpn. J. Appl. Phys. **48**, 09LB05 (2009).
9. J. J. Lee, B. G. Lee, and H. Yoo, "Depth extraction of three dimensional objects using block matching for slice images in synthetic aperture integral imaging," Appl. Opt. **50**, 5624–5629 (2011).
10. M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combination defocus and correspondence using light-field cameras," in *IEEE International Conference on Computer Vision (ICCV)* (2013), pp. 673–680.
11. F. Yi, J. Lee, and I. Moon, "Simultaneous reconstruction of multiple depth images without off-focus points in integral imaging using a graphics processing unit," Appl. Opt. **53**, 2777–2786 (2014).
12. V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," Int. J. Comput. Vis. **22**, 61–79 (1997).
13. S. H. Hong, J. S. Jang, and B. Javidi, "Three-dimensional volumetric object reconstruction using computational integral imaging," Opt. Express **12**, 483–491 (2004).
14. J. S. Lim, *Two-Dimensional Signal and Image Processing*, 1st ed. (Prentice Hall, 1990).
15. E. Cohen and Y. Yitzhaky, "No-reference assessment of blur and noise impacts on image quality," Signal Image Video Process. **4**, 289–302 (2010).
16. D. Aloni, A. Stern, and B. Javidi, "Three-dimensional photon counting integral imaging reconstruction using penalized maximum likelihood expectation maximization," Opt. Express **19**, 19681–19687 (2011).