Three-dimensional object visualization and recognition based on computational integral imaging

Bahram Javidi[a] and Yann Frauel[b*]
[a]Electrical and Computer Engineering Dept., Univ. of Connecticut, Storrs, CT, USA
[b]IIMAS, Univ. Nacional Autónoma de México, México, D.F., Mexico

ABSTRACT

We propose to use integral images to reconstruct and recognize three-dimensional (3D) scenes in a computer. A stereo-matching algorithm is applied to integral images in order to extract the depth information. This information is used to digitally reconstruct the 3D scenes. A numerical 3D correlation is then computed between various reconstructed scenes. We demonstrate the reconstruction and correlation results from experimental integral images. We propose to use a nonlinear correlation for better discrimination and we present the successful recognition and 3D localization of an object in a 3D scene. We finally compare the discrimination of two- and three-dimensional correlations.

Keywords: Integral imaging, depth estimation, three-dimensional object recognition, three-dimensional correlation, nonlinear correlation.

1. INTRODUCTION

The bases of integral imaging were proposed as soon as 1908 by G. Lippmann to capture and render three-dimensional (3D) images[1]. This technique has recently regained popularity because of its potential capability to provide 3D auto-stereoscopic displays[2-13]. The original and most common integral imaging scheme uses optical capture and rendering. In this paper we discuss a technique where the 3D input scene is digitally reconstructed in a computer using the information provided by integral images. The multiple perspectives of the scene contained in an integral image are processed using a matching algorithm to retrieve the depth information. Once the digital reconstruction is done, the 3D scenes can easily be visualized and processed as digital objects. We describe here how to use these digital reconstructed scenes for the 3D object recognition problem. We compute the 3D correlation between a reference scene and an unknown scene, which allows us to recognize and locate an object in a 3D scene. We demonstrate that a nonlinear correlation is required to obtain a good discrimination. In section 2, we will describe briefly the principle of integral imaging. In section 3, we will explain how we digitally reconstruct the 3D scene from integral images. In section 4 we will describe the implementation of the nonlinear digital 3D correlation. Finally, in section 5 we will provide experimental results of reconstruction, recognition and localization of a 3D object. We will also experimentally investigate the effect of the nonlinearity strength.

2. DESCRIPTION OF AN INTEGRAL IMAGE

Fig. 1 describes our experimental setup. A hexagonal array of microlenses is placed in front of the 3D scene to be analyzed. The microlenses have a diameter $\varphi = 200\ \mu$m and a focal length f about 2.3 mm. Each microlens generates an elementary image of the scene. All these elementary images are obtained from a slightly different point of view. We assume that the depth of focus of the microlenses is sufficient to consider that the images of all the objects are obtained in the same plane P, independently of their longitudinal position in the 3D scene. We also assume that the elementary images generated by neighbour microlenses do not overlap each other. These conditions can be obtained by placing the objects sufficiently far from the microlenses.

* yann@leibniz.iimas.unam.mx

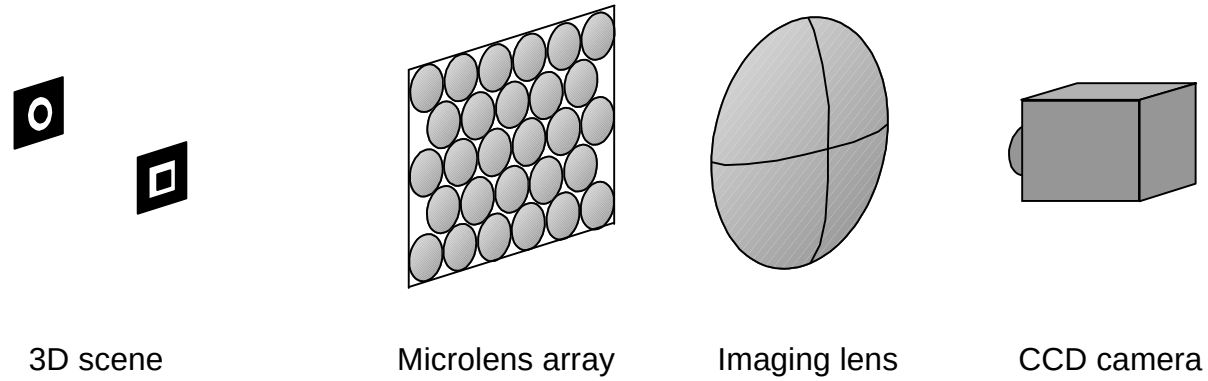


Fig. 1: Experimental setup.

Fig. 2 illustrates the formation of the elementary images by every microlens. The distance between two projections of the same object point depends on the distance of the object point. The depth of a given object point can thus be recovered by comparing the projections through different microlenses. This is a triangulation technique.

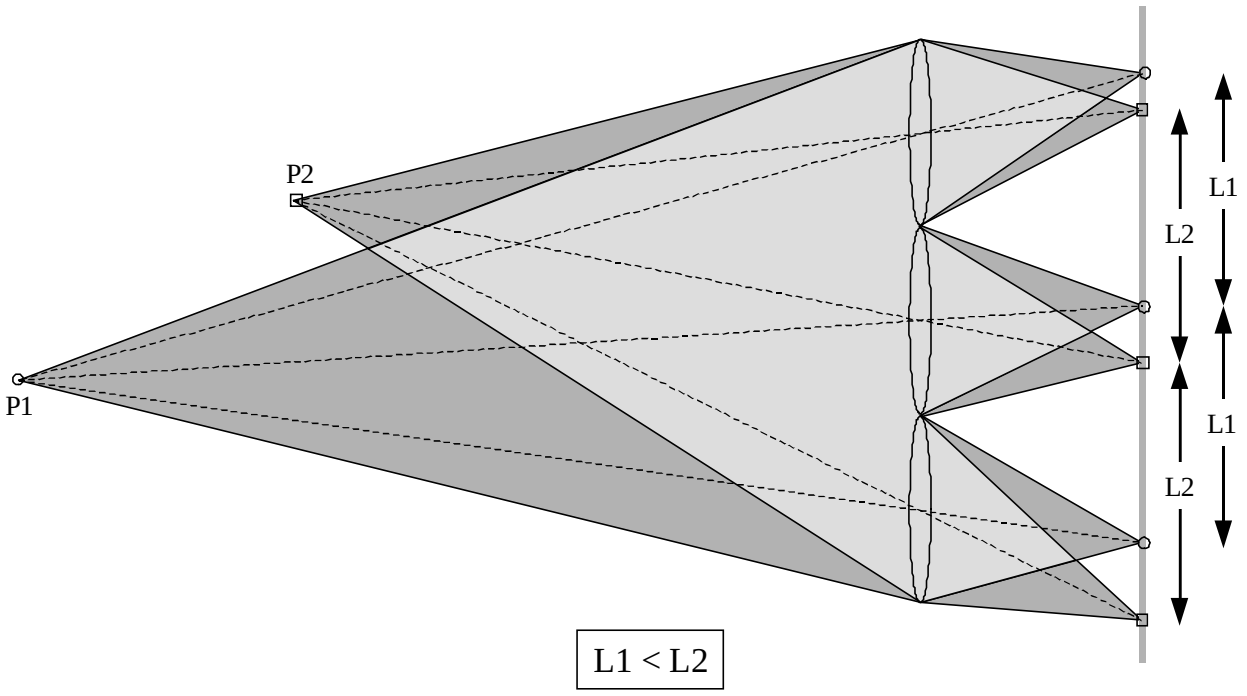


Fig. 2: Image formation through the microlens array.

3. DIGITAL RECONSTRUCTION OF THE THREE-DIMENSIONAL SCENE

Fig. 3 shows an example of our integral images. A heuristic study showed that using only 7×7 of the elementary images (marked in Fig. 3) was a good trade-off between computation time and accuracy of the depth estimation. The depth information of every object point of the scene is found thanks to a stereo-matching algorithm.[7,11] We first consider a particular point of the central elementary image – the one corresponding to microlens (0,0). We choose an arbitrary depth *z* for this point and we consequently determine the corresponding points in the other elementary images. In order to check whether our guess about the depth was correct, we have to check whether all these projected points

are actual projections of the same object point. This is done by computing 2D correlations between pairs of 9×9 pixel windows surrounding these points. We compare each window with all of its immediate neighbours (horizontally and vertically) and we add together all the correlation values. This gives us a matching criterion *M(z)* which value is maximum if all the considered projected points correspond to the same object point. The depth which yields the highest value for $M(z)$ is the actual depth of the point under consideration. This procedure is repeated for every point of the central elementary image in order to obtain the depth of every point in the 3D scene. We can therefore reconstruct a model of the 3D scene in the computer. This reconstructed 3D scene can be used to perform 3D image processing. In particular, we can reconstruct a reference scene and an unknown scene and perform a digital 3D correlation between them.

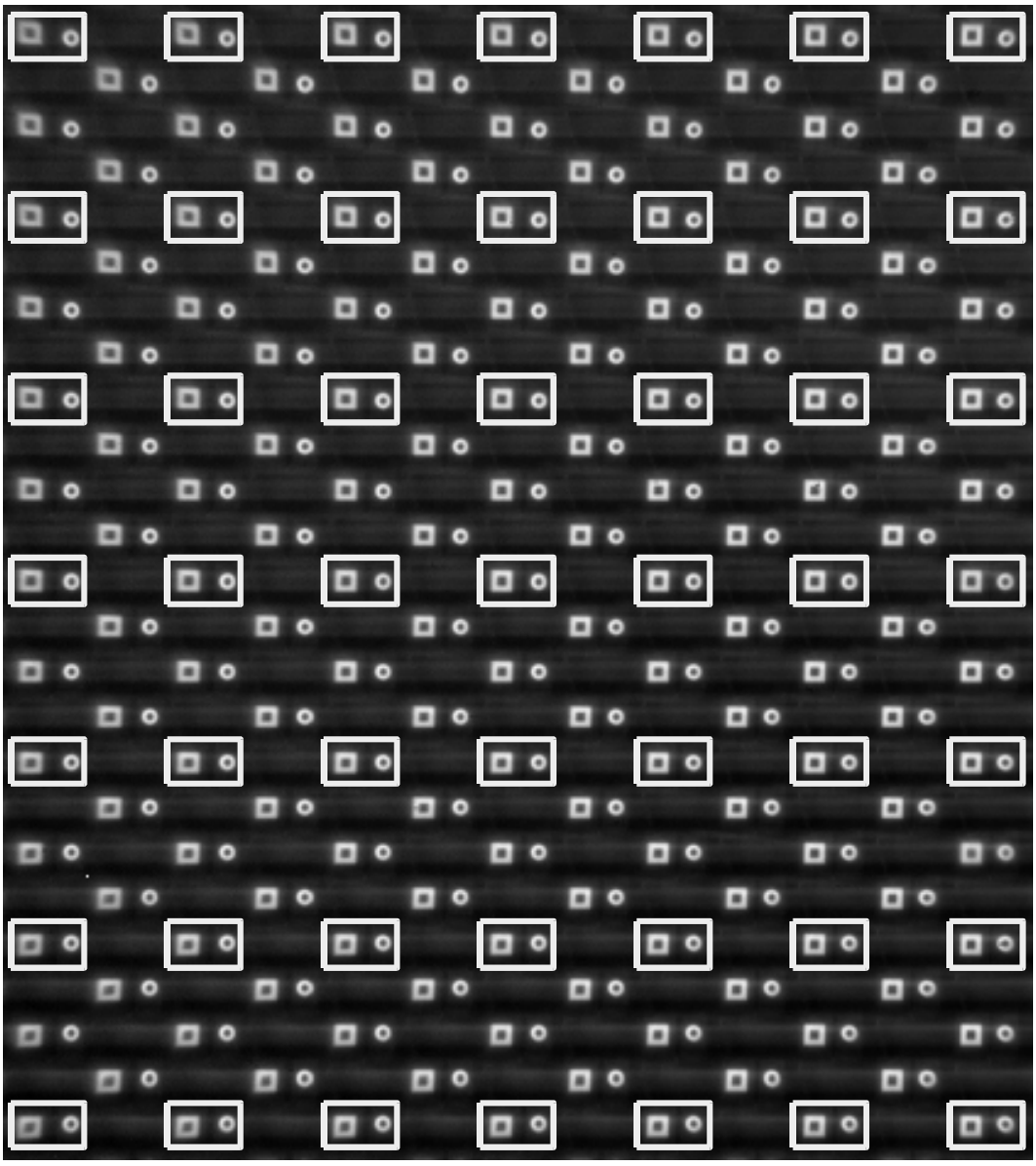


Fig. 3: Example of integral image of a 3D scene.
The marked elementary images are the ones used to determine the depth information.

4. THREE-DIMENSIONAL CORRELATION

Let $A(x,y,z)$ and $B(x,y,z)$ be two 3D scenes. We define their similarity as the square modulus of their mathematical 3D correlation. This similarity can easily be computed in the Fourier domain:

$$S_{AB} = \left|A \otimes B\right|^2 = \left|FT^{-1}\left\{\tilde{A}.\tilde{B}^*\right\}\right|^2, \tag{1}$$

where the symbol $\otimes$ stands for the 3D correlation, $\tilde{A}$ and $\tilde{B}$ are the Fourier transforms of A and B respectively and FT^{-1} is the inverse Fourier transform operation. However, the usual correlation used in Eq. (1) is known to be poorly discriminant. In order to improve the recognition performance, we can use the kth-law nonlinear correlation[14] which provides us with the following similarity criterion:

$$S_{AB}^k = \left|A \underset{k}{\otimes} B\right|^2 = \left|FT^{-1}\left\{\left|\tilde{A}\right|^k \exp\left(i\varphi_{\tilde{A}}\right).\left|\tilde{B}\right|^k \exp\left(-i\varphi_{\tilde{B}}\right)\right\}\right|^2, \tag{2}$$

where $\left|\tilde{A}\right|$ and $\left|\tilde{B}\right|$ are the magnitudes of $\tilde{A}$ and $\tilde{B}$ respectively and $\varphi_{\tilde{A}}$ and $\varphi_{\tilde{B}}$ are their phases. The value of the nonlinear factor k is chosen between 0 and 1. The linear similarity described in Eq. (1) is obtained for $k = 1$. Using a strong nonlinearity – which means k close to 0 – improves the discrimination between similar objects. In the following, we will use the term "correlation" to designate the similarity criteria defined in Eqs. (1) and (2).

5. EXPERIMENTAL RESULTS

In this section we present examples of 3D reconstruction and recognition of objects from experimental integral images. We investigate the effect of the nonlinearity and we recognize and locate a 3D object in the 3D input scene. These experiments demonstrate the greater recognition and discrimination capability of 3D correlation over 2D correlation.

5.1 Reconstruction of the 3D scenes from integral images

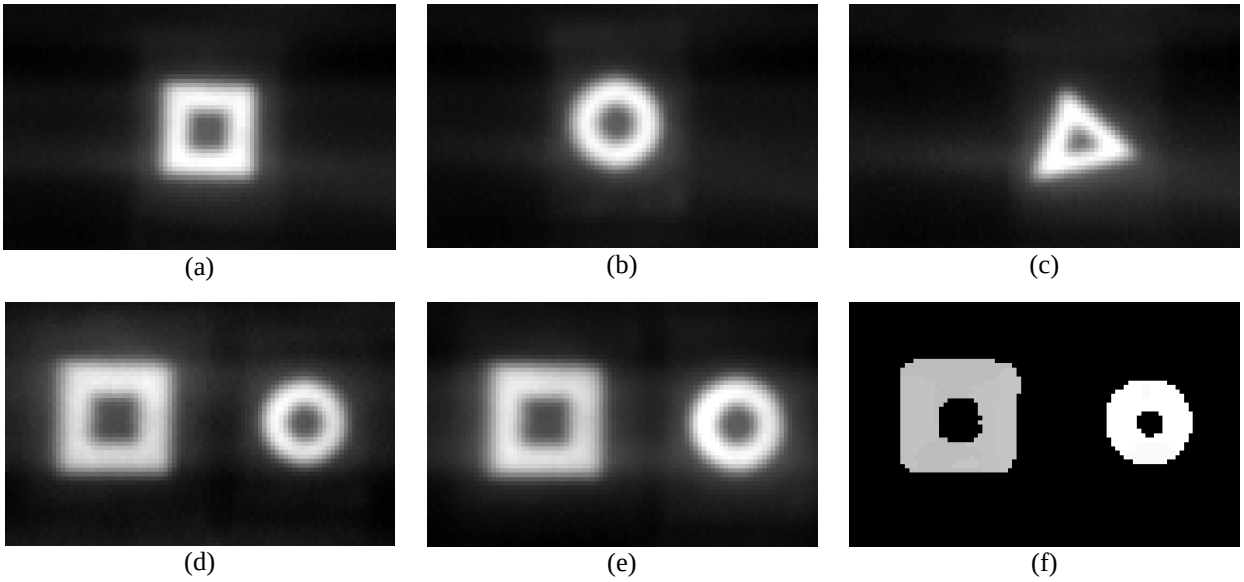


Fig. 4: Views of the 3D objects used in the experiments – (a), (b) and (c) are planar reference objects – (d) and (e) are composite 3D input scenes with the reference objects at various distances – (f) Map of the estimated depths for the 3D scene shown in (d).

In the experiments, we use three elementary planar objects representing three different geometrical shapes, namely a square, a circle and a triangle [Fig. 4(a)–(c)]. We create two composite 3D scenes by placing the square and the circle at various distances from the array [Fig. 4(d)–(e)]. We call Scene 1 the scene in Fig. 4(d) and Scene 2 the scene in Fig. 4(e). Although they look similar, the perspective views in Fig. 5(a)-(b) show the difference of depth Fig. 4(f) provides a map of the distances found by the matching algorithm for Scene 1. The brighter points represent larger distances of *z*. Fig. 6 illustrates the 3D computer reconstructions of Scene 1 and Scene 2.

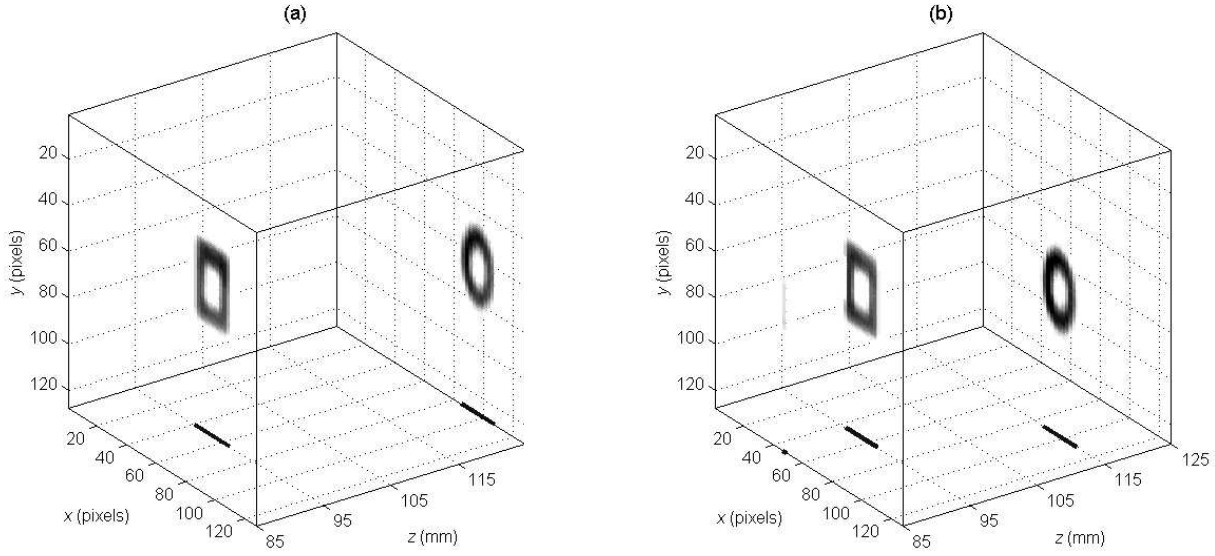
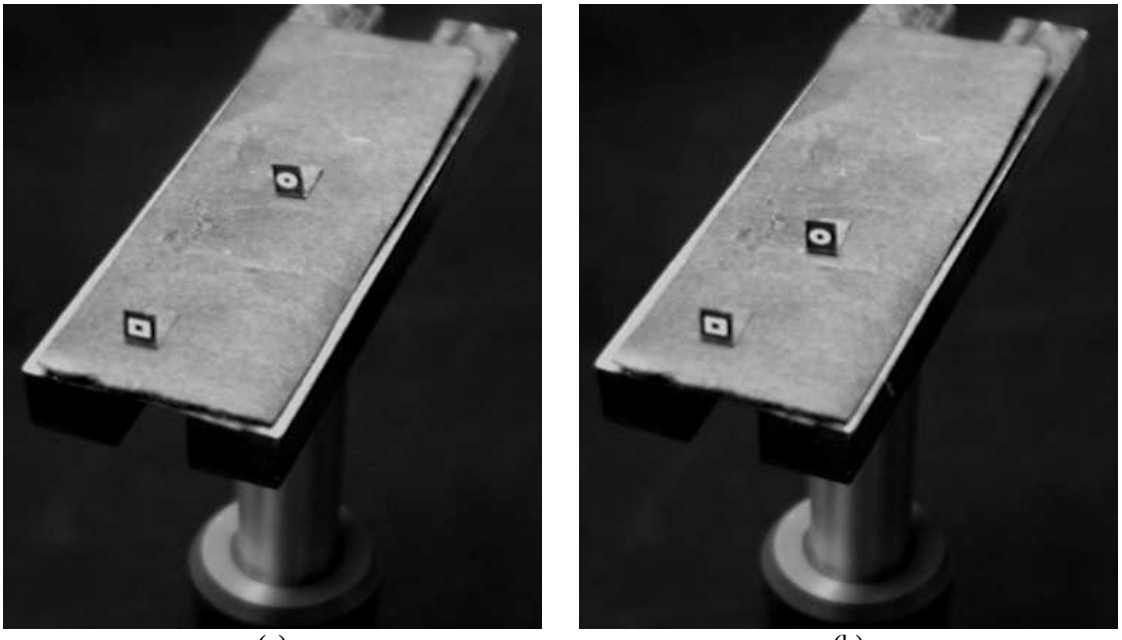


Fig. 6: Three-dimensional computer reconstructions of the scenes – (a) Scene 1 – (b) Scene 2.



(a) (b)

Fig. 5: Perspective views of the composite 3D scenes used in the experiments – (a) Scene 1 – (b) Scene 2.

5.2 Recognition and localization of a 3D object using nonlinear correlation

We use the three single geometrical objects (square, triangle and circle) as 3D reference objects. We want to be able to detect them in the two composite scenes (Scene 1 and Scene 2). We compute the 3D correlations between every input scene and every reference object, which provides $2 \times 3 = 6$ correlations. In each of these correlations, we obtain two 3D peaks that correspond to the two objects present in the input scenes. Thus a total of $2 \times 6 = 12$ peaks are generated among which only four are considered as desired detection peaks: the ones corresponding to the square in both scenes when using the square as a reference, and the ones corresponding to the circle in both scenes when using the circle as a reference. All the other peaks are cross-correlation peaks and are undesirable.

We use a *k*th-law nonlinear correlation[14] and we study the effect of the nonlinearity on the discrimination. We determine the relative values of the different peaks for each particular value of the nonlinearity *k*. A different normalization factor is applied for every *k*, so that one of the four detection peak is always unity. Fig. 7 illustrates the normalized peak values versus *k*. It is evident that it is possible to separate detection peaks from undesirable peaks only if $k \leq 0.5$. It can be seen that the best discrimination is obtained for $k = 0.2$. For this value of the nonlinear factor, it is easy to find a threshold (for instance at 0.5) that will allow us to discriminate between detection peaks and undesirable peaks.

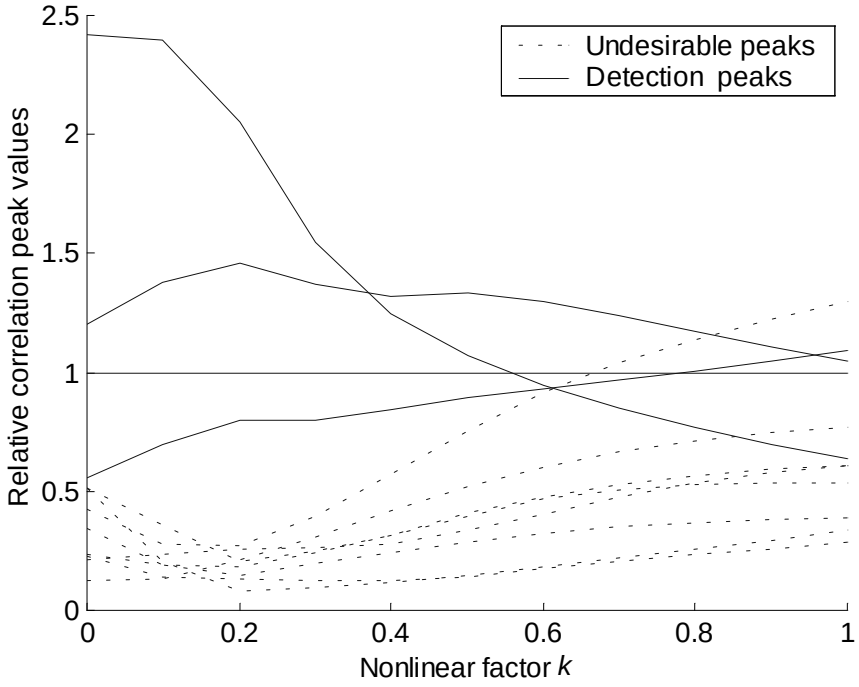


Fig. 7: Normalized values of the correlation peaks versus the *k*th-law nonlinearity. The detection peaks are the ones corresponding to the presented reference object. The other peaks are undesirable (false alarms).

When computing 3D correlations, the recognition peaks are three-dimensional. Namely, they are defined as the points in the 3D correlation space that present the highest values. These peaks indicate the 3D location of the objects in the input scene with respect to the original location of the reference object. For instance, computing the 3D correlation between Scene 1 and the square reference gives us the 3D the location of the square in Scene 1 [Fig. 8(a)]. Similarly, computing the 3D correlation between Scene 1 and the circle reference provides the location of the circle in Scene 1 [Fig. 8(b)].

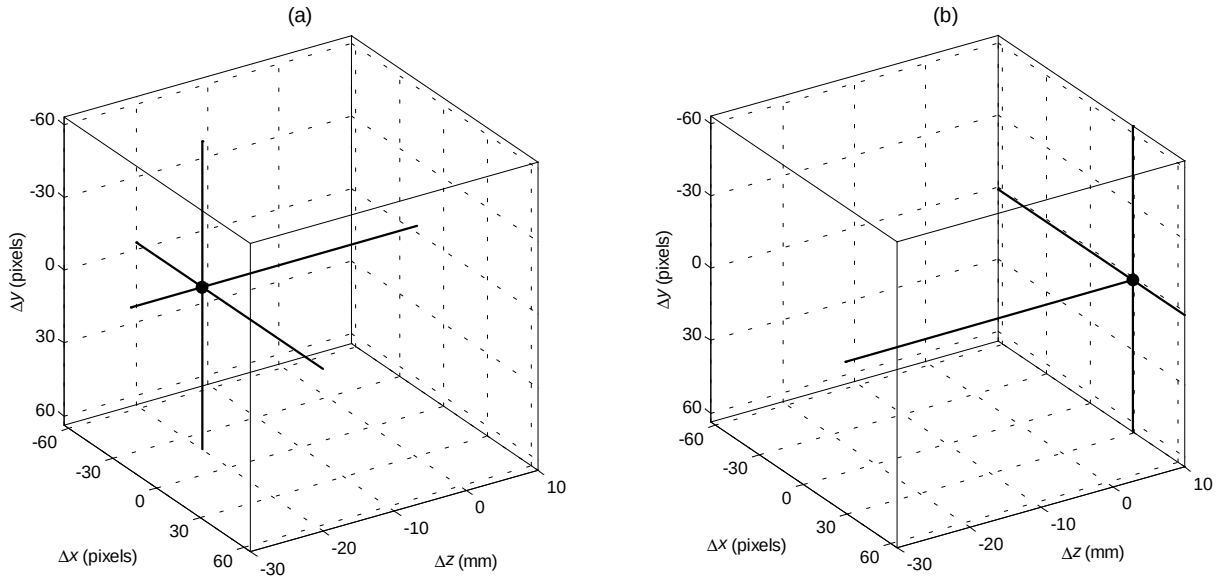


Fig. 8: Three-dimensional localization of the objects – (a) Position of the square in Scene 1 – (b) Position of the circle in Scene 1.

5.3 Comparison between 2D and 3D correlation

In the previous subsections, we detected and localized elementary objects (a square and a circle) in composite 3D scenes (Scene 1 and Scene 2). Now, we want to compare the two composite scenes one to another. Even if they contain the same elementary objects, Scene 1 and Scene 2 have different 3D structures. They therefore constitute two different 3D objects. We compare these two objects either by conventional correlation of 2D images or by the proposed 3D correlation. The 2D correlation is obtained between the images shown in Fig. 4(d) and Fig. 4(e) which are 2D projections of the two scenes. The 3D correlation is obtained between the digitally reconstructed 3D scenes. In both cases (2D or 3D), we use a nonlinear correlation with $k = 0.2$. We define the discrimination ratio as the ratio between the value of the auto-correlation peak ($C_{1\text{-}1}$ or $C_{2\text{-}2}$) and the value of the cross-correlation peak ($C_{1\text{-}2}$). Table 1 provides the values of these ratios for both 2D and 3D correlations. It can be seen that the 3D correlation is roughly 3 times more discriminant than the 2D correlation. This is because it takes into account some additional information concerning the depth structure of the 3D objects.

Discrimination ratio	$\mathbf{C_{1\text{-}1} / C_{1\text{-}2}}$	$\mathbf{C_{2\text{-}2} / C_{1\text{-}2}}$
2D	8.2	8.5
3D	21	27

Table 1: Comparison between 2D and 3D correlations for discriminating between two 3D objects.

6. CONCLUSION

In this paper, we used integral images to digitally reconstruct 3D scenes and to perform recognition of 3D objects. A matching algorithm was applied to the integral images in order reconstruct the 3D scenes in the computer. We then used the reconstructed 3D scenes to perform numerical 3D correlations. We described the use of a 3D *k*th-law nonlinear correlation and we investigated the effect of the nonlinearity strength on the discrimination capability. It was demonstrated – from experimental integral images – that the proposed technique may be used to recognize and locate 3D objects in a 3D scene. Finally, it was shown that 3D correlation provides a better discrimination than 2D correlation since it uses the depth information. The proposed 3D recognition technique is shift-invariant.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. M. R. Taghizadeh, at Heriot-Watt University in Edinburgh (United Kingdom), for providing the microlens array.

REFERENCES

1. G. Lippmann, “La photographie intégrale,” *C.R. Académie des Sciences* **146**, 446-451, 1908.
2. T. Okoshi, *Three Dimensional Imaging Techniques*, Academic Press, New York, 1971.
3. M. Starks, “Stereoscopic imaging technology: a review of patents and the literature,” Int. J. Virtual Reality **1**, 2-4, 1995.
4. F. Okano, J. Arai, H. Hoshino, and I. Yuyama, “Three-dimensional video system based on integral photography,” Opt. Eng. **38**, 1072-1077, 1999.
5. S. Nakajima, K. Masamune, I. Sakuma, and T. Dohi, “Three-dimensional display system for medical imaging with computer-generated integral photography,” Proc. SPIE **3957**, 60-67, 2000.
6. S.-W. Min, S. Jung, J.-H. Park, and B. Lee, “Three-dimensional display system based on computer-generated integral photography,” Proc. SPIE **4297**, 187-195, 2001.
7. J.-H. Park, S.-W. Min, S. Jung, and B. Lee, “New stereovision scheme using a camera and a lens array,” in *Algorithms and Systems for Optical Information Processing* V, B. Javidi and D. Psaltis, eds., Proc. SPIE **4471**, 73-80, 2001.
8. R. Zaharia, A. Aggoun, M. McCormick, “Adaptive 3D-DCT compression algorithm for continuous parallax 3D integral imaging,” Signal Processing: Image Commun. **17**, 231-242, 2002.
9. R.F. Stevens, N. Davies, G. Milnethorpe, “Lens arrays and optical system for orthoscopic three-dimensional imaging,” Imaging Science J. **49**, 151-164, 2001.
10. J.-H. Park, S. Jung, H. Choi, B. Lee, “Viewing-angle-enhanced integral imaging by elemental image resizing and elemental lens switching,” Appl. Opt. **41**, 6875-6883, 2002.
11. Y. Frauel and B. Javidi, “Digital three-dimensional image correlation by use of computer-reconstructed integral imaging, ” *Appl. Opt.* **41**, 5488-5496, 2002.
12. S.-H. Shin, B. Javidi, “Viewing-angle enhancement of speckle-reduced volume holographic three-dimensional display by use of integral imaging,” Appl. Opt. **41**, 5562-5567, 2002.
13. J.-S. Jang, B. Javidi, “Real-time all-optical three-dimensional integral imaging projector,” Appl. Opt. **41**, 4866-4869, 2002.
14. B. Javidi, “Nonlinear joint power spectrum based optical correlation,” *Appl. Opt.* **28**, 2358-2367, 1989.