Digital three-dimensional object reconstruction and correlation based on integral imaging

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

ABSTRACT

Integral images contain multiple views of a scene obtained from slightly different points of view. They therefore include three-dimensional (3D) information – including depth – about the scenes they represent. In this paper, we propose to use this depth information contained in integral images in order to recognize 3D objects. The integral images are first used to estimate the longitudinal distances of the objects composing the 3D scene. Using this information, a 3D model of the scene is reconstructed in the computer. These models are then used to compute digital 3D correlations between various scenes and objects. For a better discrimination we use a nonlinear 3D correlation. We present experimental results for digital 3D reconstruction of real 3D scenes containing several objects at various distances. With these experimental data, we demonstrate the recognition and 3D localization of objects through nonlinear correlation. We investigate the effect of the nonlinearity strength in the correlation. We finally present experiments to show that the three-dimensional correlation is more discriminant than the two-dimensional correlation.

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

1. INTRODUCTION

The two-dimensional pattern recognition problem has been researched for decades[1-3]. However, only recently has this issue been extended to take into account the three dimensions of the objects. First of all, this extension requires an acquisition technique that is able to measure the depth of the objects. Several techniques are available such as holography,[4,5] integral imaging[6-8], range images[9] or tomography[10]. This paper proposes, and demonstrates with experimental data, a 3D object recognition technique based on integral imaging[11]. Namely, we observe our 3D objects through an array of microlenses. Each of these microlenses creates an image with a slightly different point of view depending on its location with respect to the object. The set of 2D views or elementary images is acquired by a high-definition CCD camera. This method allows us to capture various 2D perspectives of the 3D scene in a single step, without having to move neither the camera nor the object. The proposed technique contains two steps: a digital reconstruction of the 3D scene and a digital 3D correlation between the reconstructions of the input and reference objects. For a better discrimination, we use a nonlinear digital correlation[12]. This technique allows us to recognize a 3D object and to locate it in the 3D scene. 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 in the computer 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. ACQUISITION OF AN INTEGRAL IMAGE

Our experimental setup is sketched in Fig. 1. An hexagonal microlens array 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 of about 2.3 mm. Each of these microlenses generates a small image of the scene taken from a different point of view. We assume that the depth of focus of the

[*] yann@leibniz.iimas.unam.mx

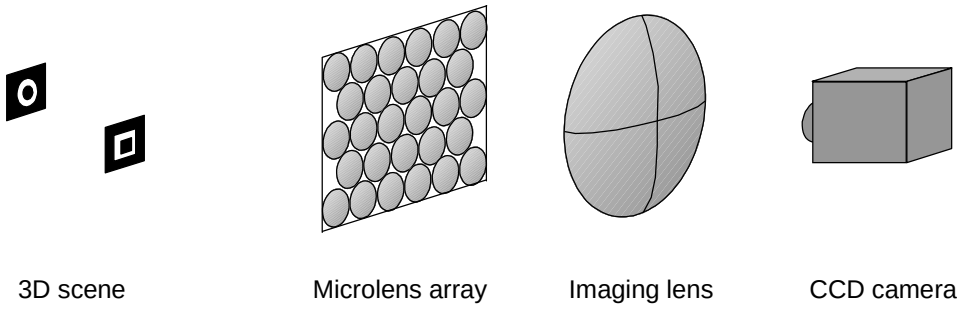


Fig. 1: Experimental setup.

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 every microlens do not overlap each other. These conditions can be obtained by placing the objects sufficiently far from the microlenses. In this case, all the images are obtained in a plane P at distance $d \approx f$ from the array (Fig. 2). This plane is then imaged by an additional lens onto a CCD camera. We will neglect this last imaging step and conduct all the calculations in plane P.

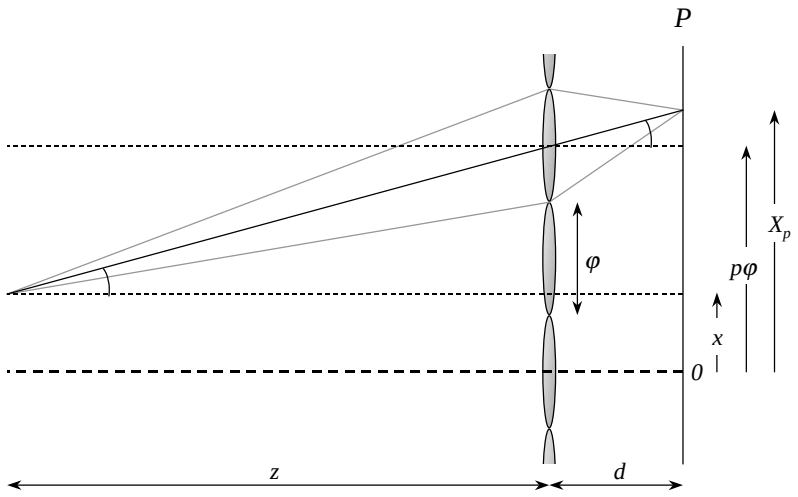


Fig. 2: Image formation through the microlens array.

As shown in Fig. 2, the coordinate X_p of an object point projected in plane P by the microlens number p depends on the original coordinate x of the object point as well as on its depth z, according to the following:

$$\frac{p\varphi - x}{z} = \frac{X_p - p\varphi}{d}, \tag{1}$$

which yields

$$X_p = p\varphi\left(1+\frac{d}{z}\right) - \frac{d}{z}x. \tag{2}$$

The same formula is obtained with the coordinate y. The distance between the projections of a same object point given by two microlenses p and q is:

$$X_q - X_p = (q-p)\varphi\left(1+\frac{d}{z}\right). \tag{3}$$

Therefore the depth of a given object point can be recovered by comparing the projections through different microlenses. We calibrate our images by illuminating the microlens array with a uniform plane wave produced by a He-Ne laser. We thus obtain an image with focused spots which provide the locations of the centers of the microlenses.

3. DIGITAL RECONSTRUCTION OF THE THREE-DIMENSIONAL SCENE

The digital reconstruction of the 3D scene consists in two steps: retrieval of the depth by a stereo matching technique, and correction of the scale according to the distance.

3.1 Estimation of the 3D scene depth

Eq. (3) suggests that two elementary images should be sufficient to determine the depth z of every point. However, it is not always easy to determine the correspondences between the points of different elementary images. By reducing the ambiguities, the use of several perspectives allows us a more accurate identification of the features, and therefore a better determination of the depth. In our experiment, the size and number of elementary images were dictated by the available optical components. Fig. 3 shows an example of our integral images. We found out 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.

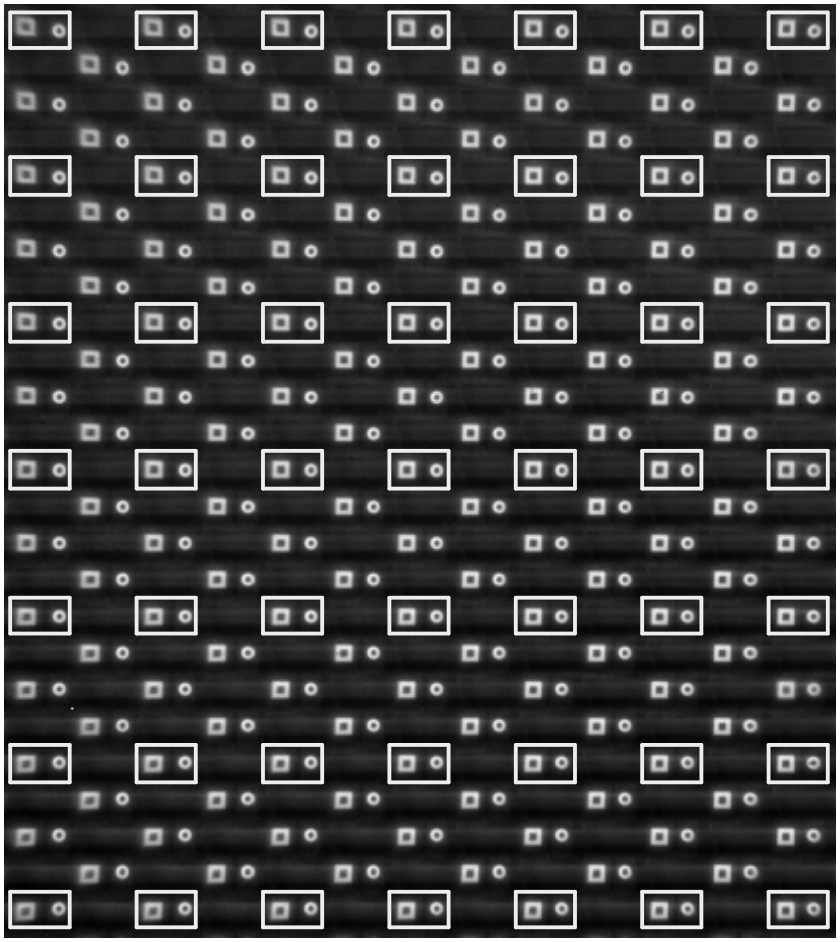


Fig. 3: Example of integral image of a 3D scene.
The marked elementary images are the ones used to determine the scene depth.

In order to find the depth of the object points, we use a stereo matching algorithm.[8,11] Let us consider one particular point of the central elementary image – the one corresponding to microlens (0,0). If we choose its depth z arbitrarily, we can determine the corresponding points in the other elementary images according to Eq. (3). We now need to verify that these points are actual projections of the same object point. This cannot be done by comparing only one point to another. We need to compare the surroundings of each of these points. In order to do that, we compute the normalized 2D cross-correlations between pairs of 9×9 pixel windows centered on the tested points. If I denotes the integral image, the projection of the inspected object point corresponding to the microlens (p,q) is $I(X_p,Y_q)$. The normalized cross-correlation between the window contained in the elementary image (p,q) and the one contained in the elementary image (p',q') is:

$$C[(p,q),(p',q')] = \frac{\sum_{m=-4}^{4}\sum_{n=-4}^{4} I\left(X_p+m,Y_q+n\right) I\left(X_{p'}+m,Y_{q'}+n\right)}{\left[\sum_{m=-4}^{4}\sum_{n=-4}^{4} I^2\left(X_p+m,Y_q+n\right)\sum_{m=-4}^{4}\sum_{n=-4}^{4} I^2\left(X_{p'}+m,Y_{q'}+n\right)\right]^{1/2}}. \quad (4)$$

This similarity criterion has the advantage of being independent of the intensity variations that can occur between two elementary images. 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 that we need to maximize:

$$M(z) = \sum_{p=-3}^{3}\sum_{q=-2}^{3} C[(p,q-1),(p,q)] + \sum_{q=-3}^{3}\sum_{p=-2}^{3} C[(p-1,q),(p,q)]. \quad (5)$$

We can compute the value of this criterion for a range of assumed depths z. The depth which provides 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. A more complete discussion about this depth estimation procedure can be found in Ref. 11.

3.2 Correction of the depth-dependent magnification ratio

In the previous subsection, we have described how to compute the z coordinate of each point of the central elementary image. However, for each of these points, we only know the projected coordinates X_0 and Y_0 and we need to find their actual coordinates x and y in the object space in order to reconstruct the 3D object. Knowing their depth z, this can be done using Eq. (2) which yields:

$$x = -\frac{z}{d} X_0 \quad \text{and} \quad y = -\frac{z}{d} Y_0 . \quad (6)$$

These equations compensate for the well-known fact that distant objects appear smaller in an image plane. By using Eq. (6), we can reconstruct their size independently of their distances from the microlens array. At this stage, we obtain a 3D reconstruction of the object space and this 3D scene can be used to perform 3D image processing such as correlations. We threshold the reconstructed 3D scene to remove the background noise before computing the 3D correlation.

4. THREE-DIMENSIONAL CORRELATION

If $A(x,y,z)$ and $B(x,y,z)$ are two 3D objects, we define their similarity as the square modulus of their 3D correlation. We compute it through the Fourier domain:

$$S_{AB} = \left|A \otimes B\right|^2 = \left|FT^{-1}\left\{\tilde{A}.\tilde{B}^*\right\}\right|^2 , \quad (7)$$

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. Moreover, in order to improve the recognition performance, we can use the kth-law nonlinear correlation[12] which provides us with the following degree of similarity:

$$S_{AB}^{k} = \left| A \underset{k}{\otimes} B \right|^2 = \left| FT^{-1} \left\{ \left| \tilde{A} \right|^k \exp\left(i\varphi_{\tilde{A}} \right) . \left| \tilde{B} \right|^k \exp\left(-i\varphi_{\tilde{B}} \right) \right\} \right|^2 , \tag{8}$$

where $\left|\tilde{A}\right|$ and $\left|\tilde{B}\right|$ are the modulus of $\tilde{A}$ and $\tilde{B}$ respectively and $\varphi_{\tilde{A}}$ and $\varphi_{\tilde{B}}$ are their arguments. The value of the nonlinear factor k is usually chosen between 0 and 1. The linear similarity described in Eq. (7) 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. (7) and (8).

5. EXPERIMENTAL RESULTS

In this section we present experiments for 3D reconstruction and recognition of objects from experimental integral images. Nonlinear correlation techniques are presented to recognize and locate a 3D object in the 3D input scene. The experiments demonstrate the greater recognition and discrimination capability of 3D correlation over 2D correlation.

5.1 Acquisition and reconstruction of the 3D scenes

In the experiments, three planar objects representing three different geometrical shapes are used, namely a square, a circle and a triangle. These shapes are about 2 mm large and are located between 90 mm and 120 mm from the microlens array. The three objects are shown in Fig. 4(a)–(c). These images are the central elementary images generated by the microlens array. We create two 3D scenes by placing the square and the circle at various distances from the array. These scenes can be seen in 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. In Scene 1 the circle is located farther away from the square. This explains its smaller size in Fig. 4(d) compared to Fig. 4(e). Fig. 4(f) provides a map of the distances obtained by the matching algorithm (see Section 3) for Scene 1. The brighter points correspond to larger distances of z. The points whose intensity is below the threshold have been plotted in black.

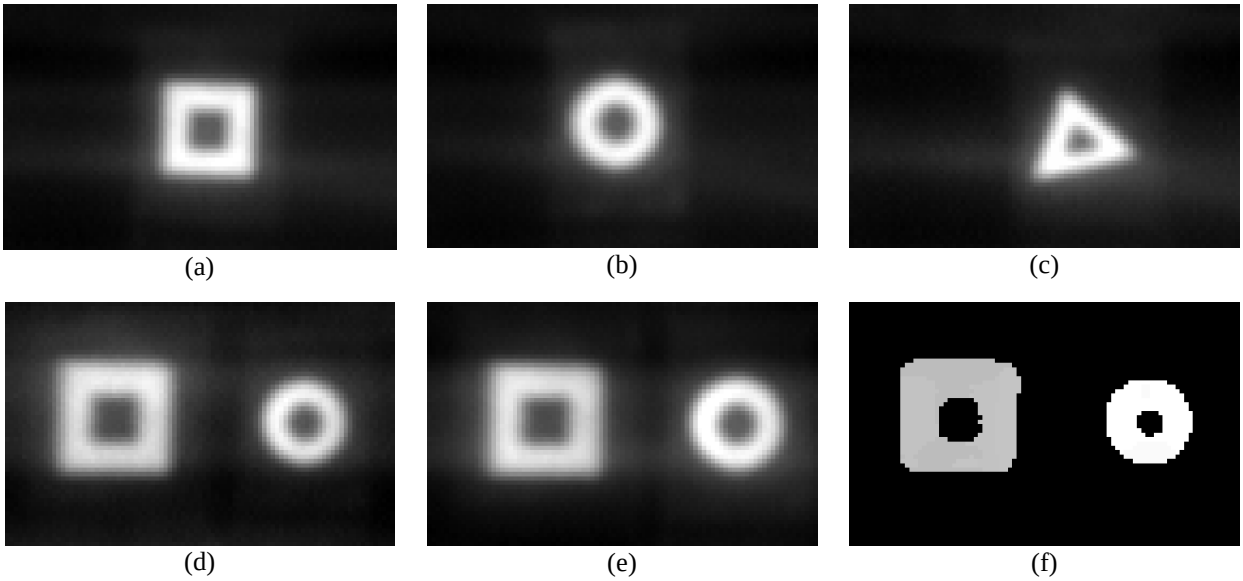
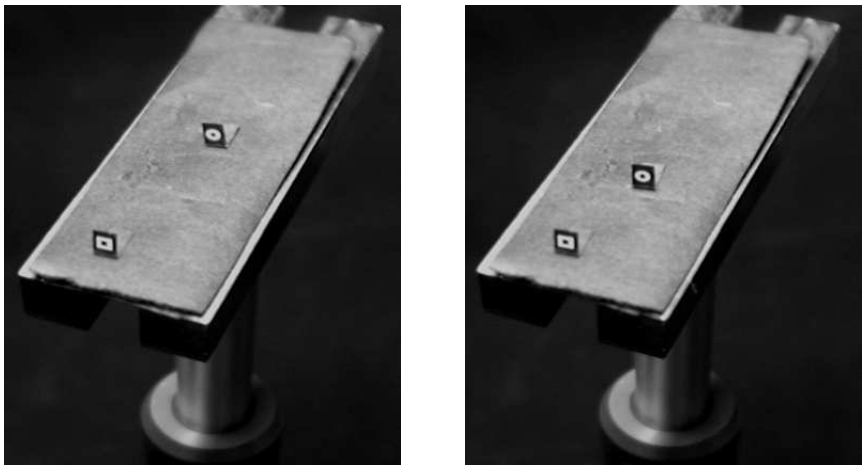


Fig. 4: Views of the 3D objects used in the experiments – (a), (b) and (c) are 3D reference objects – (d) and (e) are 3D input scenes with the reference objects at various distances – (f) Map of the estimated depths for the 3D scene shown in (d).



(a) (b)

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

Fig. 6 illustrates the 3D reconstructions of Scene 1 and Scene 2. The contrast has been inverted for a better visualization. It can be noticed that, due to the use of Eq. (6), the circle is now the same size as the square, which was not the case in Fig. 4(d)–(e). The absolute depth of the objects is estimated with an accuracy better than 10 mm.

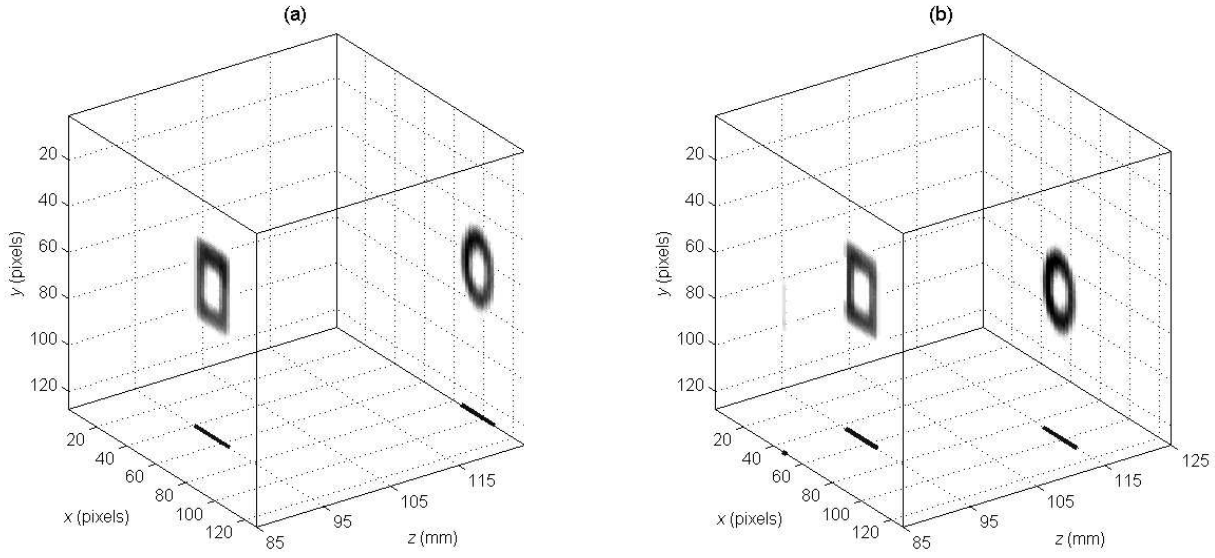


Fig. 6: Three-dimensional representation of the reconstructed scenes – (a) Scene 1 – (b) Scene 2.

5.2 Detection of a 3D object using nonlinear correlation

In order to study the recognition and discrimination capability of the proposed system, we first consider the two composite scenes (Scene 1 and Scene 2) as the 3D inputs to be tested. The three single geometrical objects (square, triangle and circle) are used as the 3D reference objects. The 3D correlations between each input scene and each reference object is computed, which provides $2 \times 3 = 6$ correlations. For each of these correlations, we obtain two main peaks that correspond to the two objects present in both input scenes. Thus $2 \times 6 = 12$ peaks are generated among which only four are considered as 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 undesirable. We illustrate the effect of *k*th-law nonlinear correlation[12] on the values of these peaks as a

function of the nonlinear factor k. We determine the relative values of the different peaks for each particular value of 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 from this graph that it is possible to separate detection peaks from undesirable peaks only if $k \leq 0.5$. A linear correlation is thus excluded. 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.

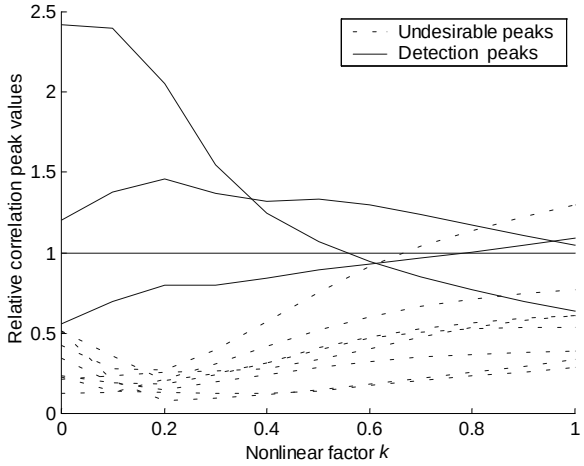


Fig. 7: Normalized values of the correlation peaks versus the kth-law nonlinearity. The detection peaks are the ones corresponding to the presented reference object. The other peaks are undesirable (false alarms).

5.3 Three-dimensional localization of an object

In the experiment described in the previous subsection, a correlation peak was obtained for each object in the 3D input scene. This peak is obviously three-dimensional and indicates the 3D location of the object in the input scene relative to the location of the object in the reference scene. For instance, Fig. 8 presents the maximum correlation values at every depth when correlating Scene 1 with the three 3D reference objects. A nonlinear correlation with $k = 0.2$ is used. As mentioned in the previous subsection, the detection peaks are selected by applying a threshold to the output at 0.5. The relative locations Δz of the remaining peaks indicate the longitudinal depths of the corresponding objects.

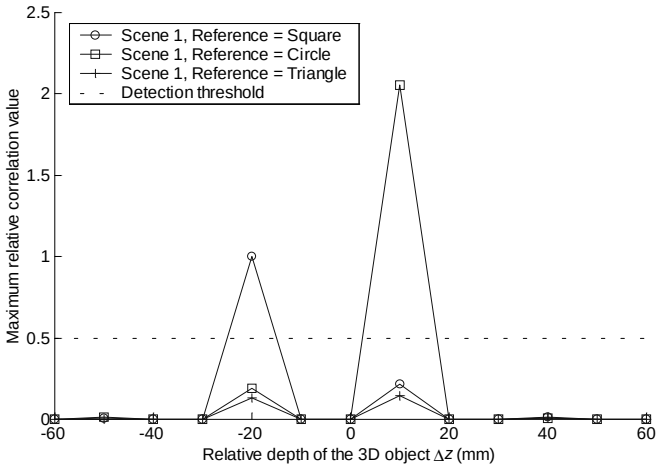


Fig. 8: Segmentation of the depth of the detected 3D objects.

Fig. 9 illustrates the correlation planes with fixed z where the maximum peaks for the square reference are generated. This graph demonstrates that the relative (x,y,z) locations of the objects can be found by the positions of the peaks. In this example, the peak in Fig. 9(b) would not be taken into consideration because it is below the threshold.

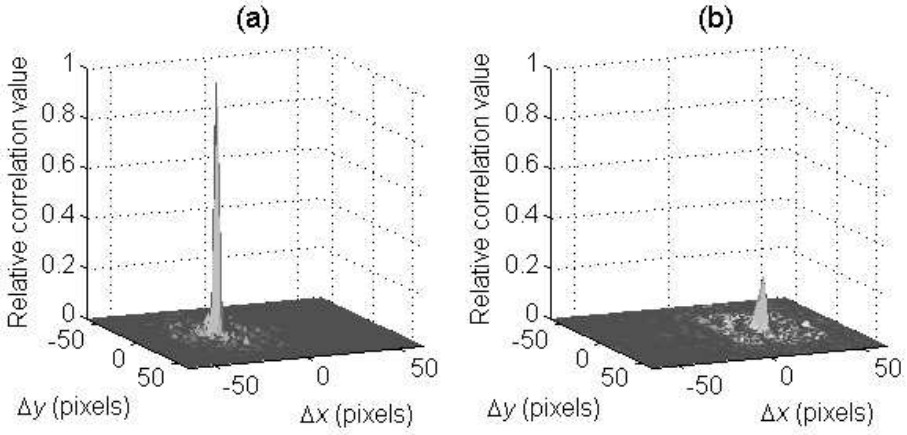


Fig. 9: Two correlation planes extracted from the 3D correlation between Scene 1 and the 3D square reference object – (a) Correlation plane corresponding to Δz = -20 mm – (b) Correlation plane corresponding to Δz = +10 mm.

We present an additional illustration of the 3D localization of the objects. Fig. 10 presents the volume representation of the 3D correlation between Scene 1 and both the square reference object [Fig. 10(a)] and the circle reference object [Fig. 10(b)]. The detection peaks are plotted in 3D. They provide the relative 3D coordinates of the reference objects in the input scene.

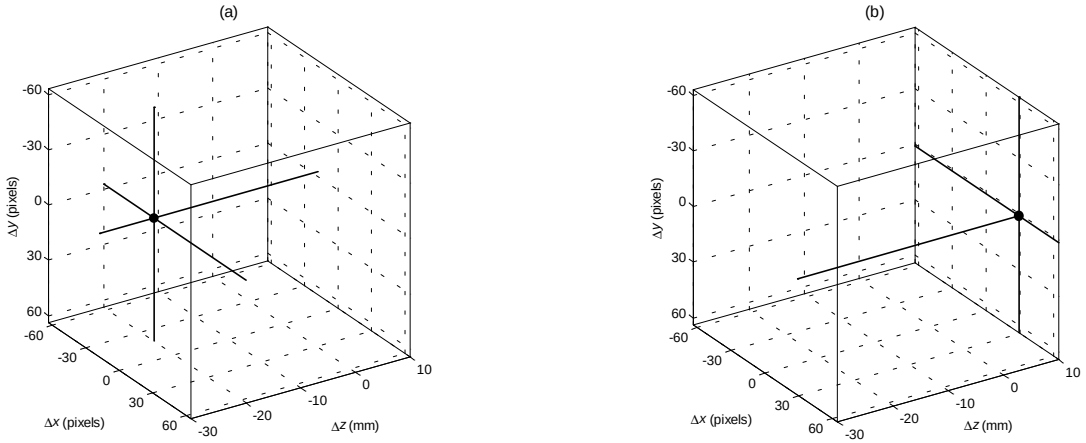


Fig. 10: Three-dimensional localization of the correlation peak – (a) Scene 1, square reference – (b) Scene 1, circle reference.

5.4 Comparison between 2D and 3D correlation

In the previous subsections, we achieved 3D detection of elementary objects (a square and a circle) in complex 3D scenes (Scene 1 and Scene 2). In this subsection, we compare two complex scenes one to another. Scene 1 and Scene 2 have different 3D structures. They therefore constitute two different 3D objects. The correlation of these two 3D objects will be compared by using conventional correlation of 2D images and by proposed 3D correlation. The 2D correlation is obtained between the images shown in Fig. 4(d) and Fig. 4(e) which are the views obtained through the central microlens of the array. We apply a thresholding on the images to remove the background noise. The second comparison method consists in digitally reconstructing the 3D scenes and computing their 3D correlation as described previously.

The values $C_{1\text{-}2}$ of the cross-correlation peaks for both of these methods is given in Table 1. In both cases (2D or 3D), we use a nonlinear correlation with $k = 0.2$. Of course the correlation values by themselves cannot be compared because they are not normalized. In order to give a comparison scale, we also provide the values of the auto-correlations for Scene 1 ($C_{1\text{-}1}$) and Scene 2 ($C_{2\text{-}2}$) for both the 2D and the 3D methods. Lastly, we define the discrimination ratio as the ratio between the value of the auto-correlation and the value of the cross-correlation. Table 1 presents the values of this ratio with respect to both scenes. 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.

Correlation peak	$\mathbf{C_{1\text{-}2}}$	$\mathbf{C_{1\text{-}1}}$	$\mathbf{C_{2\text{-}2}}$	$\mathbf{C_{1\text{-}1} / C_{1\text{-}2}}$	$\mathbf{C_{2\text{-}2} / C_{1\text{-}2}}$
2D	0.157	1.29	1.33	8.2	8.5
3D	1.74×10^9	36.0×10^9	47.5×10^9	21	27

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

6. CONCLUSION

In this paper we have demonstrated the digital reconstruction and recognition of 3D objects by using integral imaging. First, we recalled how it is possible to estimate the depth of an object from an integral image. Then we used this depth information to digitally reconstruct the objects in three dimensions. With these reconstructed 3D objects, we performed numerical 3D correlations in order to recognize the 3D objects. We presented nonlinear correlation results using various nonlinearities to investigate the discrimination capability. It was demonstrated that the proposed technique may be used to recognize and locate 3D objects in a 3D scene. Finally, it was shown that the 3D correlation provides a better discrimination than 2D correlation since it uses the object depth information. The proposed 3D recognition technique is shift-invariant, that is, it is not affected by a lateral or longitudinal shift of the 3D object. However, it has to be mentioned that the estimation of the depth is all the less accurate as the object is far from the microlens array. This problem is inherent to any technique based on triangulation.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. M. R. Taghizadeh, at Heriot-Watt University in Edinburgh, for providing the microlens array.

REFERENCES

1. A. B. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory* **10**, 139-145, 1964.
2. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New-York, 1968.
3. *Image Recognition and Classification*, B. Javidi, ed., Marcel Dekker, New York, NY, 2002.
4. E. N. Leith and J. Upatniecks, "Reconstructed wavefronts and communication theory," *JOSA* **52**, 1123-1130, 1962.
5. J. Caulfield, *Handbook of Optical Holography*, Academic Press, London, 1979.
6. G. Lippmann, "La photographie intégrale," *C.R. Académie des Sciences* **146**, 446-451, 1908.
7. T. Okoshi, *Three Dimensional Imaging Techniques*, Academic Press, New York, 1971.
8. 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.
9. M. Suk and S.M. Bhandarkar, *Three-dimensional object recognition from range images*, Springer, New York, 1992.
10. G. T. Herman, *Image reconstruction from projections: The fundamentals of computerized tomography*, Academic Press, New York, 1980.
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. B. Javidi, "Nonlinear joint power spectrum based optical correlation," *Appl. Opt.* **28**, 2358-2367, 1989.