

Three-dimensional pattern recognition with a single two-dimensional synthetic reference function

Youzhi Li and Joseph Rosen

A novel, to our knowledge, method of distortion-invariant three-dimensional (3-D) pattern recognition is proposed. A single two-dimensional synthetic discriminant function is employed as a reference function in the 3-D correlator. Thus the proposed system is able to identify and locate any true-class object in the 3-D scene. Preliminary simulation and experimental results are presented. © 2000 Optical Society of America

OCIS codes: 100.6890, 070.2580, 100.6740, 070.2590, 070.4550, 070.5010, 100.5760, 070.6020, 070.6110.

1. Introduction

Recently, Rosen extended the optical correlator from operation in two dimensions to three.^{1,2} His method involves fusing images of objects from a few different points of view and allows objects to be identified and located in three-dimensional (3-D) space. This 3-D correlation has been demonstrated on a 3-D joint transform correlator (JTC), in which a reference and tested objects are observed together from a distance. The reference object and the tested objects are projected a few times from different points of view on a spatial light modulator (SLM), and the projected images are electro-optically processed yielding the desired 3-D correlation. However, this scheme suffers from limitations similar to those of the conventional two-dimensional (2-D) optical correlators,³ namely, sensitivity to geometrical distortions. Objects from the same class as the reference (i.e., the true class) must appear in the same in-plane and out-plane orientation, and be the same size as the reference, to be correctly identified. Otherwise, these objects might be mistakenly classified as belonging to the false class.

In this study we present a preliminary solution to the problem of sensitivity to distortions. Instead of

placing the reference object in the observed scene with the tested objects,^{1,2} we propose a different method of computing and employing the reference function. Our proposed synthetic reference is obtained as a function of the training set, and it is invariant to some distortions determined by this training set. The same 2-D synthetic reference function is displayed side by side with each projected image for all the different points of view. The reference function is designed to recognize every object from the true class, with any distortion defined by the training set, and is capable of locating this object in 3-D space. The fact that the reference function is uniform for all the projections simplifies the system and yet enables us to achieve some degree of distortion invariance in the 3-D pattern-recognition system.

In this preliminary project we concentrate on a limited kind of distortion. Let us assume that the observed scene, with the coordinates (x, y, z) , is observed from various points of view, all distributed on the x - z plane. In Ref. 2 and here the points of view are located along an arc lying on the x - z plane, where the origin of the (x, y, z) space is used as the arc's center point (see Fig. 1). The type of distortion we consider in the present study is object rotations on the x - z plane. In other words, the demonstrated system is invariant to any object's rotation, within a limited angular interval, on the x - z plane. Our aim is to guarantee that each object from the true class will be identified and located in 3-D space no matter what its orientation within the limited angular interval on the x - z plane. However, the same concept can be extended toward other kinds of distortions by

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel. J. Rosen's e-mail address is rosen@ee.bgu.ac.il.

Received 11 May 1999; revised manuscript received 4 October 1999.

0003-6935/00/081251-09\$15.00/0

© 2000 Optical Society of America

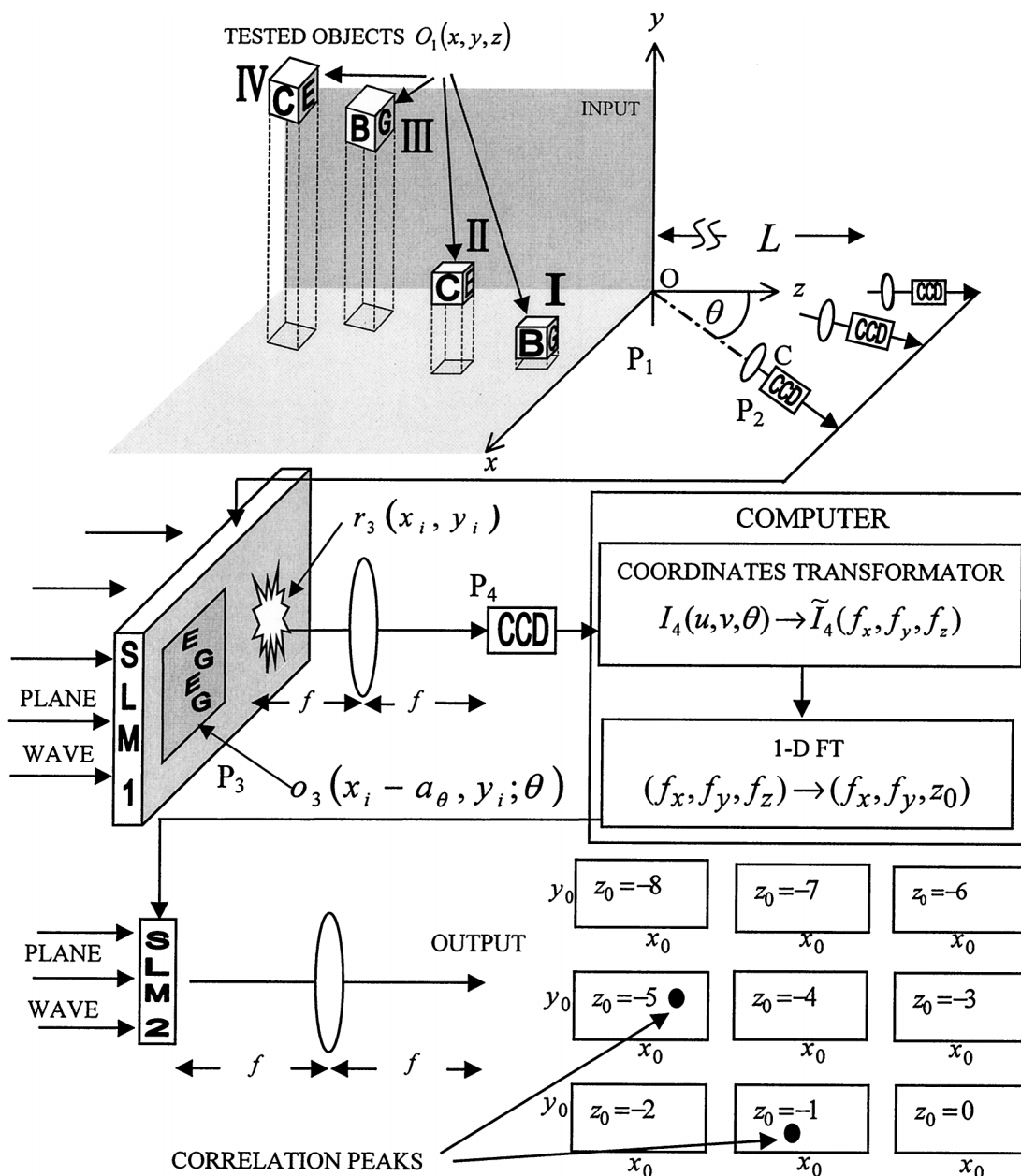


Fig. 1. Schematic of the 3-D joint transform correlator equipped with the SDF.

inclusion of appropriate representatives of these distortions in the training set.

The reference function is actually a synthetic discriminant function (SDF),⁴⁻⁶ appearing side by side with all object projections. The SDF can be computed by any known optimization algorithm.⁵ The fact that the correlation's dimensions have been extended from two to three does not increase the dimensions of the computational problem in our case. Thus one can choose any off-the-shelf 2-D SDF algorithm and implement it in our distortion-invariant 3-D correlator. In recent years many SDF algorithms for distortion-invariant problems have been widely investigated.⁵ Among them, the minimum average correlation energy SDF (MACE-SDF)⁶ has been successfully demonstrated in a few independent

experiments. With the MACE-SDF one can control the whole correlation plane as well as keep a sharp easily detectable correlation peak. For demonstration purposes of our general concept we choose the MACE-SDF as the reference function in our first distortion-invariant 3-D correlator.

2. Three-Dimensional Correlator with a Virtual Reference Function

The distortion-invariant 3-D JTC is shown in Fig. 1. The cameras (or a single camera that moves from point to point) record the input scene from different points of view. Each image is displayed on SLM1 located on plane P_3 . Unlike previous versions of the 3-D JTC,^{1,2} here the reference function is not a real object located in the input scene. Instead, plane P_3 is divided into

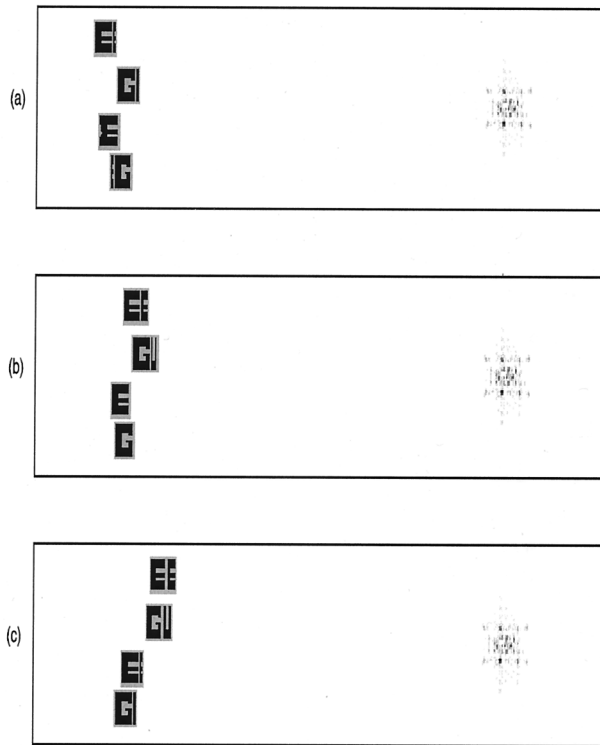


Fig. 2. Three images of plane P_3 out of 19 as observed from different points of view.

two parts. The various projections of the observed scene, from the various points of view, are displayed, one by one, on one side of plane P_3 . On the other side of plane P_3 the same single 2-D reference function constantly appears. In general the reference function is complex valued, and therefore one should consider the technique for implementing complex-valued functions in a JTC. The direct and simple method is to use a transparency with the desired amplitude and phase modulations. Alternatively, if such transparency is not available, one can employ indirect methods such as holographic coding,⁷ or interferometric techniques,⁸ to get an effective complex reference function. In this study we demonstrate the new procedure by both computer simulation and electro-optical experiment. Obviously, in the computer simulation the reference function can in general be complex without any limitation. In the experiment we used a similar technique employed before in Ref. 9. The Fourier holograms of the input objects are multiplied by a filter function in the Fourier plane, and another Fourier transform (FT) of this product yields the desired correlation results. Such a correlator is actually a combination of the JTC (with a delta function as the reference) and the VanderLugt correlator. As a hybrid configuration it combines the superb features of both types of correlators. Explicitly, it enables us to implement an effective complex reference function as is usually possible with VanderLugt correlator. It also lets us perform complicated mathematical manipulations (in our case it is coordinate transformations,

as explained below) on the spatial spectrum of the input function, as is inherently possible with the JTC.

Other than the new concept of the 2-D reference function, the system is similar to the previous one presented in Ref. 2. We briefly summarize the correlation process in the following. A 3-D input function $o_1(x, y, z)$, describing all tested objects in the observed scene, is located in the coordinate system (x, y, z) , where P_1 is the transverse plane $z = 0$ (see Fig. 1). From each point of view the camera observes on plane P_1 through an imaging lens located a distance L from plane P_1 . In each point of view the line OC between the center of the camera's plane and the origin point $(x, y, z) = (0, 0, 0)$ is orthogonal to the camera's plane. The angle between the z axis and the line OC is denoted by θ . For each θ , the projected function $o_3(x_i, y_i; \theta)$ is displayed on SLM1, where (x_i, y_i) are the coordinates of plane P_3 . The relation between (x_i, y_i, θ) and (x, y, z) is given by²

$$(x_i, y_i) = M_0(x \cos \theta + z \sin \theta, y), \quad (1)$$

where M_0 is the magnification factor of the imaging lens. It is assumed that the distance L is much longer than the depth of the object function $o_1(x, y, z)$, and therefore the magnification factor is approximately the same constant M_0 for all the object points.

In addition to the projected function $o_3(x_i, y_i; \theta)$, plane P_3 contains the reference function $r_3(x_i, y_i)$. We choose to locate $r_3(x_i, y_i)$ in the origin of plane P_3 , and $o_3(x_i, y_i; \theta)$ is centered around the point $(x_i, y_i) = (a_\theta, 0)$. Returning to the (x, y, z) space, we can describe the whole distribution on plane P_3 , for all various values of θ , as a collection of projections of the virtual 3-D function $g_1(x, y, z)$ located in plane P_1 and given by

$$g_1(x, y, z) = o_1(x - \tilde{x}, y, z - \tilde{z}) + r_1(x, y, z). \quad (2)$$

From Eq. (1) it is clear that \tilde{x} and \tilde{z} satisfy the relation $M_0(\tilde{x} \cos \theta + \tilde{z} \sin \theta) = a_\theta$. Thus we see indeed that a_θ depends on θ . $r_1(x, y, z)$ is a virtual 3-D function, satisfying the condition that, for any angle θ , its projection function is always $r_3(x_i, y_i)$. The functions $g_1(x, y, z)$ and $r_1(x, y, z)$ are termed virtual, because they do not exist in physical reality. However, the 3-D JTC effectively yields a real 3-D autocorrelation result of the virtual function $g_1(x, y, z)$, and therefore $g_1(x, y, z)$ and $r_1(x, y, z)$ should be considered in our analysis. The collection of 2-D functions that does exist on plane P_3 , each of which for different values of θ , is given by $g_3(x_i, y_i; \theta) = o_3(x_i - a_\theta, y_i; \theta) + r_3(x_i, y_i)$. Next we consider the intensity distribution on plane P_4 , for any angle θ . With the 2-D FT relation between planes P_3 and P_4 , the intensity on plane P_4 is

$$I_4(u, v, \theta) \propto \left| \iint g_3(x_i, y_i; \theta) \times \exp \left[i \frac{2\pi}{\lambda f} (ux_i + vy_i) \right] dx_i dy_i \right|^2, \quad (3)$$

where λ is the optical wavelength and f is the focal distance of the Fourier lens. Let us look now at the

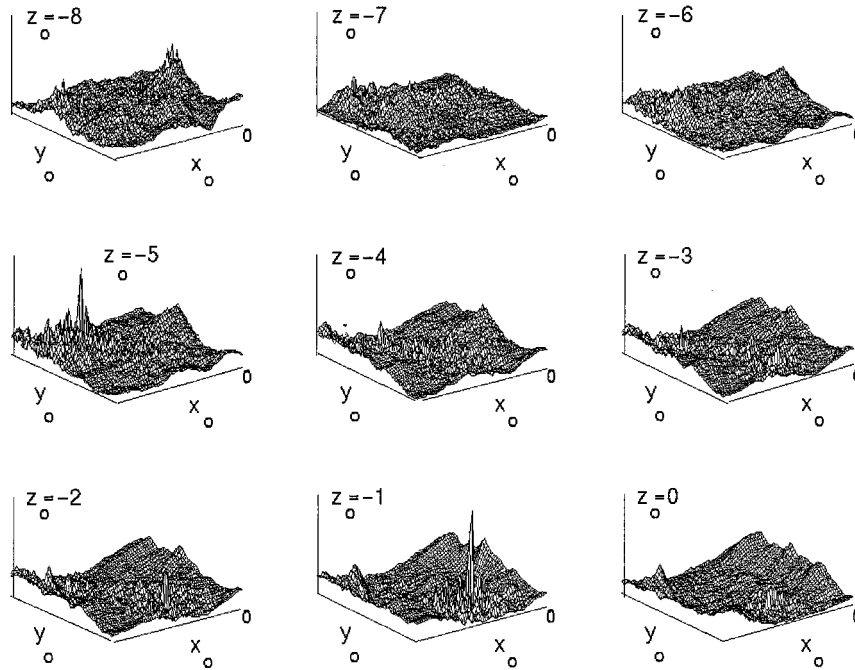


Fig. 3. Intensity of the correlation results of the 3-D joint transform correlator.

intensity distribution $I_4(u, v, \theta)$ as a function of the virtual image $g_1(x, y, z)$. For a single element of the size $(\Delta x, \Delta y, \Delta z)$, and brightness $g_1(x, y, z)$, from the entire 3-D object function the intensity on plane P_4 is

$$I_4(u, v, \theta) \propto \left| g_1(x, y, z) \exp \left[i \frac{2\pi}{\lambda f} (ux_i + vy_i) \right] \Delta x \Delta y \Delta z \right|^2$$

$$= \left| g_1(x, y, z) \exp \left[i \frac{2\pi M_0}{\lambda f} (ux \cos \theta + vy + uz \sin \theta) \right] \Delta x \Delta y \Delta z \right|^2, \quad (4)$$

where the final part of Eq. (4) is obtained after substitution of Eq. (1) into the first part of Eq. (4). Next we examine the influence of all points of the virtual object $g_1(x, y, z)$. Since the SLM is illuminated by a coherent plane wave, the electromagnetic field contributed from each virtual image point is summed. The virtual object is three dimensional, and therefore the overall intensity at plane P_4 is obtained by a 3-D integral as follows:

$$I_4(u, v, \theta) \propto \left| \iiint g_1(x, y, z) \exp \left[i \frac{2\pi M_0}{\lambda f} (ux \cos \theta + vy + uz \sin \theta) \right] dx dy dz \right|^2. \quad (5)$$

Our goal is to get the magnitude square of the 3-D FT of $g_1(x, y, z)$ on plane P_4 . That is because the effect of the 3-D JTC can be obtained when an additional 3-D FT is performed on this magnitude square. Looking over relation (5), we see that it has the form

of a 3-D FT with Fourier coordinates $(u \cos \theta, v, u \sin \theta)$. However, the real coordinates of the physical space in P_4 are (u, v, θ) . Therefore, to get the desired 3-D FT, we must transform the recorded data from coordinate systems (u, v, θ) to $(u \cos \theta, v, u \sin \theta)$. The intensity $I_4(u, v, \theta)$ is recorded into a computer in which this coordinate transform is performed. In this stage the transformed function $\tilde{I}_4(u \cos \theta, v, u \sin \theta)$ is actually (part of) the square absolute magnitude of the 3-D FT of the virtual function $g_1(x, y, z)$. Therefore, following the convolution theorem, another 3-D FT of $\tilde{I}_4(u \cos \theta, v, u \sin \theta)$ yields the auto-correlation of $g_1(x, y, z)$ as follows:

$$c(x_0, y_0, z_0) \propto \iiint \tilde{I}_4(u \cos \theta, v, u \sin \theta)$$

$$\times \exp \left[-i \frac{2\pi M_0}{\lambda f} (x_0 u \cos \theta + y_0 v + z_0 u \sin \theta) \right] d(u \cos \theta) dv d(u \sin \theta)$$

$$= \iiint g_1(x, y, z) g_1^*(x - x_0, y - y_0, z - z_0) dx dy dz. \quad (6)$$

This final 3-D FT is performed by two steps: first, a one-dimensional digital FT from $u \sin \theta$ to z_0 and then, multiple 2-D optical FT's from $(u \cos \theta, v)$ to (x_0, y_0) , each one for a different value of z_0 . In case one has only a single SLM, the series of 2-D optical FT's is done sequentially one at a time. The number of a 2-D transform in the sequence is actually the measure of z_0 .

Similar to a conventional JTC, one of the four terms^{1,2} of this autocorrelation is the requested 3-D cross correlation between the object function $o_1(x, y, z)$ and the virtual reference function $r_1(x, y, z)$. In Section 3 we describe the process of synthesizing $r_3(x_i, y_i)$ to get a distortion-invariant pattern recognition from the 3-D cross correlation between $o_1(x, y, z)$ and $r_1(x, y, z)$.

3. Synthetic Discriminant Function Synthesis

Searching for an efficient SDF algorithm, we wish to maintain the dimensionality of the computation problem, even though the correlation dimensions increase from two to three. For the purpose of the SDF computation only, we use, as the training set, 2-D images of the object instead of the real 3-D objects. The training set contains the object's projections from various angles according to the desired limits of the rotation invariance. The use of 2-D images is a kind of approximation. It avoids the necessity to increase the dimensionality of the SDF algorithm. This approximation is justified under the assumption that the maximal angle θ is small (the maximal angle is 11° in our experiment) and that the object's dimension along the z direction is no more than its dimension along the x axis. These conditions let us further assume that most of the information about the object's pattern is given by the 2-D image obtained from its central projection (i.e., from $\theta = 0$). Looking from other points of view up to the most extreme points does not add significant information that can substantially change the correlation results. However, we realize that this approach is not suitable for all cases of 3-D correlation, and it is demonstrated here only as a preliminary simplified example of the use of a synthetic reference function in a 3-D correlator. It should be emphasized, however, that, although we use a 2-D correlation for the SDF synthesis, after the computation is done, the resulting reference is employed in the 3-D correlator. In other words, although the SDF is computed with 2-D correlations, in the recognition tests we perform 3-D correlation between real-input 3-D objects and the effectively 3-D reference function. Reducing the computation problem to two dimensions enables us to choose any known SDF algorithm.⁵ As mentioned above, for this preliminary demonstration we chose the MACE-SDF as a well-proved algorithm, but we realize that other SDF's might be suitable as well.

The MACE-SDF achieves full correlation-plane control by minimizing the average correlation energy. Under the above-mentioned approximation, it is a minimization of the energy of 2-D correlation functions. It thus produces a sharp and easily detected correlation peak with minimum sidelobes and false peaks. Our objective is to design a reference function that ensures sharp correlation peaks while keeping the object's location accurately and retaining shift invariance. The algorithm of the MACE-SDF⁶ operates in the spatial-frequency domain. Therefore, in the case of the JTC, the final result of the algorithm, the filter function, is converted by a 2-D inverse FT to

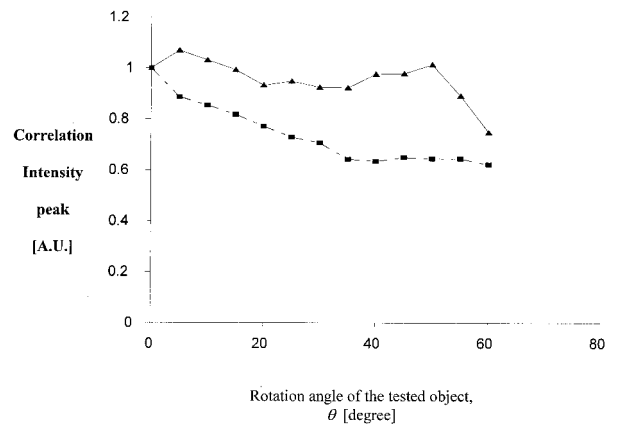


Fig. 4. Correlation peak of one true-class object versus the object's rotation angle for the conventional reference function (rectangles) and for the SDF reference function (triangles).

the image domain to be employed as a reference function. We briefly summarize here the main points of this algorithm, which is extensively described in Ref. 6.

Without losing generality, we consider two classes of the problem. Let $f_1(x, y; \theta_1), \dots, f_1(x, y; \theta_M), f_2(x, y; \theta_1), \dots, f_2(x, y; \theta_M)$ denote the $2M$ training images obtained from the 2-D projections of the 3-D objects from different view angles θ_j 's. f_1 and f_2 designate the true and the false classes, respectively. $F_1(u, v; \theta_1), \dots, F_1(u, v; \theta_M), \dots, F_2(u, v; \theta_M)$ denote the 2-D FT's of the training images, and $R^*(u, v)$ denotes the filter function. We constrain the filter to satisfy the following correlation-peak values:

$$\iint F_i(u, v; \theta_j) R^*(u, v) du dv = c_{i,j}, \quad i = 1, 2, \quad j = 1, \dots, M, \quad (7)$$

where $c_{i,j}$ is the expected correlation-peak value between the object f_i rotated by θ_j and the reference function. The energy of the (i, j) th correlation plane is defined by

$$E_{i,j} = \iint |F_i(u, v; \theta_j)|^2 |R^*(u, v)|^2 du dv. \quad (8)$$

The optimization is carried out with a vector notation.^{5,6} Assuming that each 2-D image has d pixels overall, $F_{i,j}$ and R denote the d -dimensional vectors obtained by sampling in a lexicographic order the functions $F_i(u, v; \theta_j)$ and $R^*(u, v)$, respectively. The column vector c contains the expected correlation value $(c_{1,1}, \dots, c_{1,M}, c_{2,1}, \dots, c_{2,M})^T$, and \bar{F} is a $d \times 2M$ matrix with $F_{i,j}$ as its $[j + (1 - 1)M]$ th column. To minimize the average correlation-plane energy, we average all the energy values given by Eq. (8). In a vector-matrix notation the function to be minimized is

$$E_{av} = (1/M) R^+ D R, \quad (9)$$



(a)



(b)



(c)

Fig. 6. Three images of the input scene out of 23, as observed from different points of view by the CCD camera on plane P_2 .

above-mentioned algorithm. In our simulation the value of c for the true class is 1 and for the false class is 0.2.

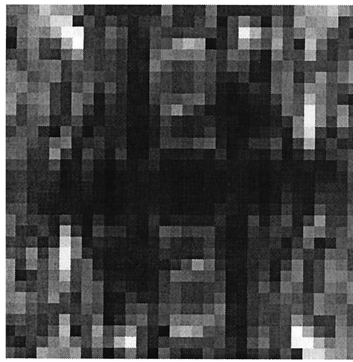
When the reference function is displayed on plane P_3 , the system should be able to recognize and locate

true class members, which appear in any orientation within $\pm 50^\circ$ in the x - z plane. In the test shown in Fig. 1 the input space contains four objects. Two of them, I and III, belong to the true class. Object III is rotated 20° from the z axis in the x - z plane. The other two objects, II and IV, belong to the false class, and they should be ignored by the system. In the test stage of our simulation the point of view was shifted along an arc of 18° , 9° to each side of the z axis. The arc's center is in the origin of the xyz space, and all the simulated cameras are directed to this point. The total number of projections in the test stage is 19 in 1° increments. Each projection was mapped on plane P_3 side by side with the reference function. Three examples of plane P_3 , out of 19, are shown in Fig. 2. Figures 2(a), 2(b), and 2(c) show the most extreme projections to the left-hand side of the z axis, on the z axis, and to the right-hand side of the z axis, respectively. The magnitude of the same reference is shown in the right-hand side of these figures. All 19 images of plane P_3 were processed by the 3-D JTC as described in Section 2. The output result of this system appears as a collection of correlation planes, each for a different value of the output longitudinal axis z_0 . As with every JTC, the expected cross-correlation results between the tested objects and the reference are obtained near the first diffraction order. This region is shown in Fig. 3 for a few values of z_0 . The high correlation peaks on planes $z_0 = -5$ and $z_0 = -1$ indicate the successful identification of the two cubes, I and III, of the true class.

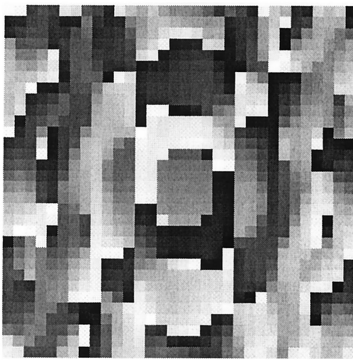
A comparison between two methods of 3-D correlation is presented in Fig. 4. In the first method, indicated by triangles, the true class object is correlated with the MACE-SDF for various values of rotation angle in the x - z plane. In the second method, indicated by rectangles, the same object is correlated with the conventional reference function described in Refs. 1 and 2. In the range between 0° and 50° , which is the angular range of the training set, the MACE-SDF keeps a stable peak value, whereas the correlation peak of the conventional reference function gradually decreases.

5. Experimental Results

The optical experiment was carried out by the system shown in Fig. 5. Instead of displaying the reference function on the SLM, we display only the projections of the 3-D input scene. Three examples, out of 23, of these projections are shown in Fig. 6. Figures 6(a), 6(b), and 6(c) show the scene seen by the CCD positioned in angles of -11° , 0° , and 11° from the optical axis, respectively. The angular increment between every two successive projections was 1° , and therefore $\pm 11^\circ$ were the extreme angles of this experiment. The distance between the center of the scene and the CCD on plane P_2 was 64 cm. The observed scene contained three cubes of size $5 \text{ cm} \times 5 \text{ cm} \times 5 \text{ cm}$. The lower left has the letters BGU on three of its faces, and its lowest forward left-hand corner is used as the origin of the input space. From this origin point, at the point $(x, y, z) = (6, 5.6, -7.5) \text{ cm}$



(a)



(b)

Fig. 7. (a) Magnitude and (b) phase angle of the MACE-SDF filter used in the experiment.

was located the lowest forward left-hand corner of the other cube with the letters BGU. This cube, however, was rotated 22° to the left compared with the lower cubes. The third cube was used as the false object, and the letters LRL appeared on three of its faces, whereas R was on the front face. The location of its lowest forward left-hand corner is at the point $(10.2, 0, -2.5)$ cm.

From the SLM each projection was Fourier transformed and interfered with a reference plane wave on the CCD plane P_4 . As a result, 23 Fourier holograms of the various projections went into the computer. Inside the computer each hologram was first multiplied by the SDF filter, computed by the algorithm described in Section 3. Of course, in this case we use the direct result from the SDF algorithm in the Fourier plane, without transforming it to the image plane. We obtained the training sets for the algorithm by recording the BGU, and later the LRL, cubes at the origin from an angle -20° to 35° with 5° increments. The magnitude and the phase angle of the computed 32×32 pixel SDF filter are shown in Figs. 7(a) and 7(b), respectively. As expected, the MACE-SDF is more transparent in the high-

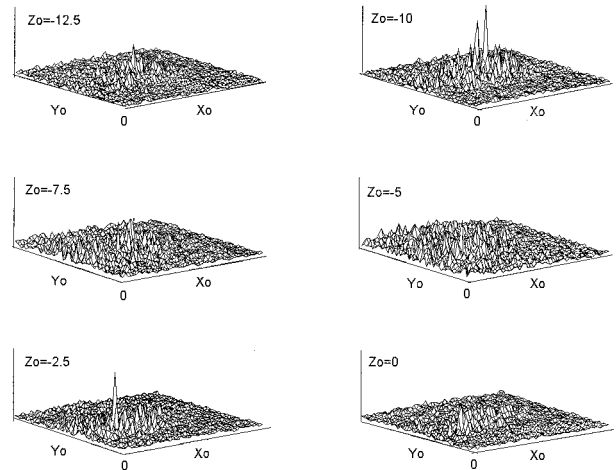


Fig. 8. Intensity of the correlation plane resulting from the experiment of the 3-D correlation. The values of z_0 are given in centimeters.

frequency region, and this feature is responsible for the sharp correlation peaks for the true-class objects.

After multiplying all 23 holograms by the complex SDF filter, we performed the coordinate transformation mentioned in Section 2. Finally, the computer calculated the second 3-D FT, whereas the proposal to cooperate optics in this final transform is postponed for a future study. The correlation results are depicted in Fig. 8. Each 3-D plot shows the intensity distribution on a transverse plane along the z_0 axis. Two recognizable peaks appear in the locations of the two true-class objects, whereas the false cube did not grow any peak at all.

6. Conclusions

In conclusion, we have demonstrated a process of 3-D correlation between a real-world 3-D function and a 3-D virtual complex reference function. This final function is virtual, because it does not exist in its 3-D form. The only thing we see from this function is its 2-D projections on SLM1. All reference's 2-D projections are equal to the same function computed by the off-the-shelf algorithm MACE-SDF. On one hand, it is convenient to maintain the low dimensionality, at least for SDF calculation. However, on the other hand, such is also the limitation of this system. In case objects from the training set have crucial information along their depth dimension, the 2-D algorithm does not achieve the goal of distortion-invariant recognition. In such cases a 3-D function should be synthesized, and currently we seek an efficient 3-D algorithm for this purpose.

In this preliminary study we wanted only to examine and demonstrate the evolution from a real-object reference 3-D function^{1,2} to a synthetic complex reference function. The goal of this new correlation method has been achieved, namely, a limited distortion-invariant optical pattern recognition together with locating the true-class object in 3-D space.

This research was supported by the Israel Science Foundation.

References

1. J. Rosen, "Three-dimensional electro-optical correlation," J. Opt. Soc. Am. A **15**, 430–436 (1998).
2. J. Rosen, "Three-dimensional joint transform correlator," Appl. Opt. **37**, 7538–7544 (1998).
3. J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996), Chap. 8, p. 251.
4. D. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition," Appl. Opt. **19**, 1758–1761 (1980).
5. B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Appl. Opt. **31**, 4773–4801 (1992).
6. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filters," Appl. Opt. **26**, 3633–3640 (1987).
7. D. Mendlovic, E. Marom, and N. Konforti, "Complex reference-invariant joint-transform correlator," Opt. Lett. **15**, 1224–1226 (1990); U. Mahlab, J. Rosen, and J. Shamir, "Iterative generation of complex reference functions in a joint-transform correlator," Opt. Lett. **16**, 330–332 (1991).
8. R. Piestun, J. Rosen, and J. Shamir, "Generation of continuous complex-valued functions for a joint transform correlator," Appl. Opt. **20**, 4398–4405 (1994).
9. J. Rosen, T. Kotzer, and J. Shamir, "Optical implementation of phase extraction pattern recognition," Opt. Commun. **83**, 10–14 (1991).