Three-dimensional electro-optical correlation

Joseph Rosen

Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel

Received May 20, 1997; accepted September 18, 1997; revised manuscript received October 1, 1997

Electro-optical implementation of a three-dimensional (3-D) spatial correlation is proposed. A 3-D scene of objects, seen from the paraxial zone, is correlated with a reference object. As an example of application, we describe a 3-D joint transform correlator that is capable of recognizing targets in the 3-D space. © 1998 Optical Society of America [S0740-3232(98)02102-4]

OCIS codes: 070.2590, 070.5010.

1. INTRODUCTION

A correlation is essential in signal processing in general, and in optical image processing in particular. A spatial correlation is employed extensively in various schemes of edge enhancement,[1] pattern recognition,[2] target tracking,[3] and more. In most of these schemes the functions involved are at most two dimensional (2-D). However, our real spatial world is three dimensional (3-D), and in some applications one needs to process 3-D objects in their natural 3-D environment. Pattern recognition and target tracking in the 3-D space are examples of applications that can benefit by use of the 3-D correlation. In these applications one employs the information obtained from the 3-D shape of the target and learns its location in the 3-D space. In this work we describe a process of the 3-D correlation between two 3-D real-world functions. This correlation is demonstrated on a 3-D joint transform correlator (JTC) for the application of pattern recognition.

A correlation of any dimension can be expressed by two successive Fourier transforms (FT's). Therefore, we first propose a method to perform 3-D FT of a real-world 3-D function. The idea behind this 3-D FT is inspired by the work of Rosen and Yariv.[4] They have proposed a method of receiving a 3-D FT of the intensity distribution of an incoherent radiation source by measuring the far-field 3-D degree of coherence. In the present method the 3-D FT is composed from a series of 2-D FT's, each of which is performed on a 2-D projection of the 3-D input function, observed by a camera from a different point of view. The 3-D object is observed from various points of view distributed on a finite transverse plane located far from the object. It is assumed that the field of view is wider than the transverse dimension of the input function and that the depth of focus of the camera is longer than the longitudinal dimension of the input function.

2. THREE-DIMENSIONAL FOURIER TRANSFORM

A 3-D input function $o(x_s, y_s, z_s)$, shown in Fig. 1, is located in the coordinate system (x_s, y_s, z_s), where P_1 is the transverse plane $z_s = 0$. A camera observes plane P_1 through an imaging lens located a distance L from plane P_1. The lens and the camera are transversely displaced a distance (D_x, D_y) from the z_s axis. To express the relation between the 3-D input function and the distribution on the Fourier plane, let us first look at a single point (x_s', y_s', z_s') from the entire input object. The observed point is imaged on plane P_2 at point (x_i, y_i). Assuming that plane P_1 is imaged with a magnification factor M, we can calculate the location of the observed point (x_i, y_i) as a function of its location in the object space and the amount of the camera's displacement (D_x, D_y). According to Fig. 1, the location of the imaged point on plane P_2 is

$$x_i = \frac{M(D_x + x_s')}{1 - z_s'/L}$$

$$y_i = \frac{M(D_y + y_s')}{1 - z_s'/L}. \tag{1}$$

Assuming that $L \gg z_{s,\max}'$, we approximate (x_i, y_i), by taking only the first two terms of the binomial expansion of $(1 - z_s'/L)^{-1}$, as follows

$$x_i \cong M(D_x + x_s' + z_s'D_x/L + z_s'x_s'/L),$$

$$y_i \cong M(D_y + y_s' + z_s'D_y/L + z_s'y_s'/L). \tag{2}$$

Each imaged point (x_i, y_i) is Fourier transformed to a linear phase function $\exp[i2\pi\kappa(x_iu + y_iv)]$, where (u, v) are the coordinates of the 2-D Fourier plane and κ is a constant. This FT can be done optically by displaying the imaged point on an electrical-addressed spatial light modulator. The spatial light modulator is located in the front focal plane of a spherical lens and illuminated by a plane wave. Thus a 2-D FT of the image on the spatial light modulator is obtained on the lens's back focal plane. In this case $\kappa = 1/\lambda f$, where f is the lens's focal length and λ is the wavelength of the plane wave. Alternatively, the FT can be done by an electronic computer that

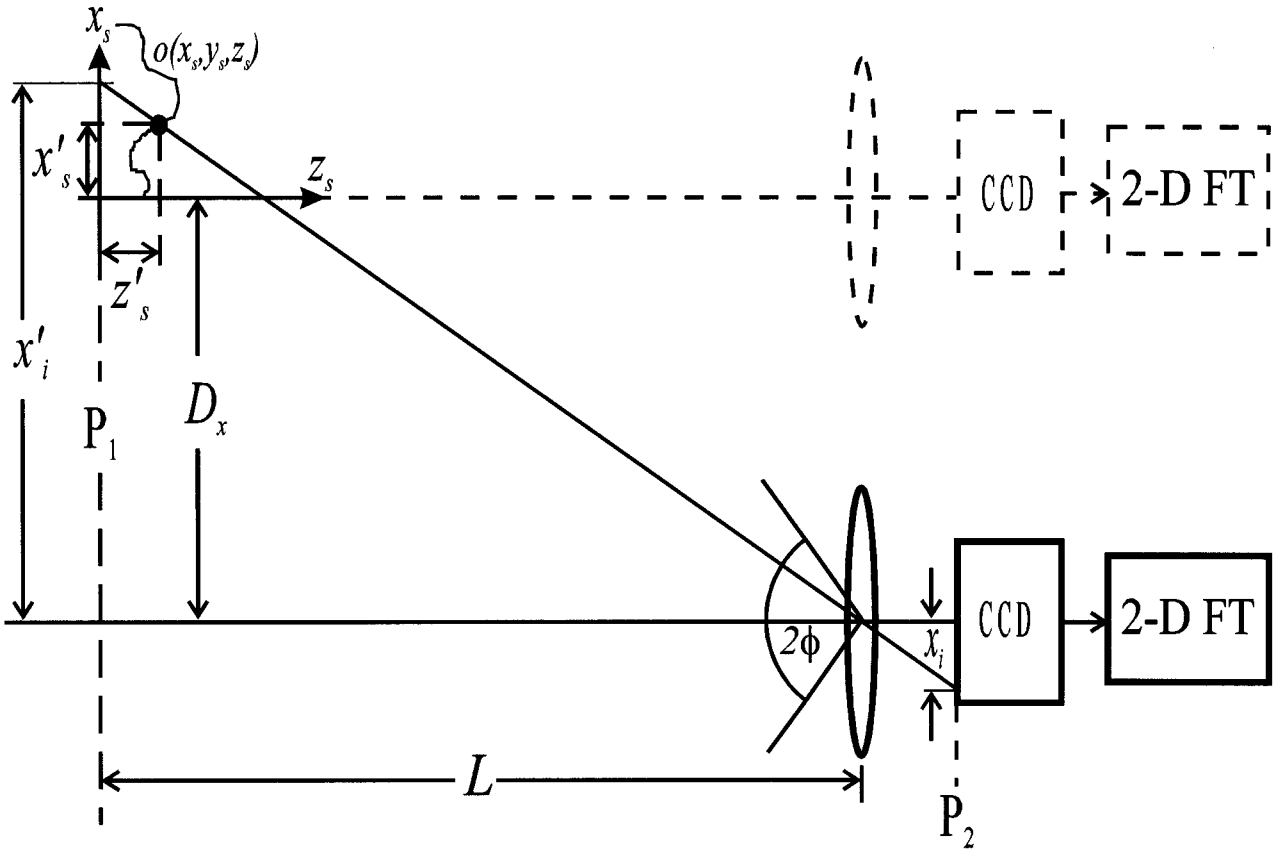

Fig. 1. Illustration of the imaging system for calculating Eq. (1).

receives the image directly from the camera. We obtain the overall distribution on the Fourier plane for a given displacement (D_x, D_y) by integrating over the linear phases contributed by all the 3-D object points, as follows

$$O(u, v, D_x, D_y) = \iiint o(x_s, y_s, z_s) \times \exp[i2\pi\kappa(x_i u + y_i v)]\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}z_s . \tag{3}$$

At this point we assume that $L \gg 2\pi M\kappa u_{\max}\Delta z_s\Delta x_s$ and $L \gg 2\pi M\kappa v_{\max}\Delta z_s\Delta y_s$, where $(\Delta x_s, \Delta y_s, \Delta z_s)$ denote the size of the input function, and $(u_{\max}, v_{\max})$ are the maximum values of the Fourier plane. In these assumptions we require that the maximum phase change contributed by the fourth terms of Eqs. (2) be much less than 1 rad. For a given input object, the validity of these assumptions depends on the system parameters, and therefore we can always design the system to satisfy these assumptions. Following the assumptions, the fourth terms in the right-hand side of Eqs. (2) can be neglected, and we approximate the location of each image point as

$$x_i \cong M[D_x + x'_s + z'_s D_x/L],$$
$$y_i \cong M[D_y + y'_s + z'_s D_y/L]. \tag{4}$$

Substituting Eqs. (4) into Eq. (3) yields

$$O(u, v, D_x, D_y) = \exp[i2\pi M\kappa(D_x u + D_y v)] \times \iiint o(x_s, y_s, z_s) \times \exp\{i2\pi M\kappa[x_s u + y_s v + (z_s/L)(D_x u + D_y v)]\}\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}z_s . \tag{5}$$

Eq. (5) is a 3-D FT multiplied by a linear phase function, which transforms an object function $o(x_s, y_s, z_s)$ into a 3-D spatial-frequency function $O(\omega_x, \omega_y, \omega_z)$, where $\omega_x = M\kappa u$, $\omega_y = M\kappa v$, and $\omega_z = M\kappa(D_x u + D_y v)/L$. We note that ω_z is unusually dependent on the transverse spatial-frequency variables u and v. The three independent variables in the transformation space are u, v, and the vector (D_x, D_y), but the longitudinal frequency variable ω_z is a linear combination of (D_x, D_y) scaled by u and v. Nevertheless, it turns out that this transform [defined by Eq. (5) without the linear phase function] has features similar to those of the conventional 3-D FT. In particular, it satisfies the convolution theorem,[5] and therefore this peculiar transform can be used as a building block in the spatial correlation process.

To recognize the limitations of the system, we first consider the required maximum displacement of the camera $(D_{x,\max}, D_{y,\max})$. Following a conventional Fourier analysis, we know that the maximum camera's displacement depends on the longitudinal size δz_s of the smallest input element. Assuming $D_{x,\max} = D_{y,\max} = D$ and

$u_{\max} = v_{\max} = B$, the condition $DB \geqslant L/M\kappa\delta z_s$ should be satisfied so that the information on the smallest longitudinal element is not lost. If this condition is not satisfied, the spectral width of the smallest longitudinal element cannot be detected by the system. The bandwidth of the system in the third dimension ω_z depends directly on the product of the transverse bandwidth and the maximal camera's displacement. On the other hand, the condition on the transverse bandwidth is $B \geqslant 1/M\kappa\delta x_s$, where δx_s is the transverse size of the smallest input element. Therefore the ratio between the longitudinal and the transverse resolutions is $\delta z_s/\delta x_s = L/D$. If a system with a better longitudinal than transverse resolution is desired, a maximum displacement D longer than L should be chosen. That means that the imaging system has a field angle greater than 90°, which may lead to the undesirable phenomenon of vignetting.[6]

The observed object should remain in the field of view of the camera, and therefore another limitation on the camera's displacement is given by the condition $D \leqslant L \tan\phi - \Delta x_s/2$, where 2ϕ is the field angle of the imaging system (assume that $\Delta y_s = \Delta x_s$). Shifting the camera beyond this limitation causes the object to disappear from the field of view. On the other hand, the condition $D \gg \Delta x_s$, should be satisfied; otherwise, there is no justification for keeping the third terms on the right-hand side of Eqs. (2) while neglecting the fourth terms.

Although it is convenient to analyze the system in terms of continuous signals, our detected 3-D signal is discrete in all its dimensions. This is so because each 2-D image is recorded separately as a collection of discrete pixels inside the computer. Therefore the limitations on the sampling interval along the camera's translation should be considered. Let us assume that the maximal sampling intervals $(\delta D_{x,\max}, \delta D_{y,\max})$ satisfy the equation $\delta D_{x,\max} = \delta D_{y,\max} = d$ and that the maximal transverse sampling intervals $(\delta u_{\max}, \delta v_{\max})$ satisfy the equation $\delta u_{\max} = \delta v_{\max} = b$. The criterion $db \leqslant L/2M\kappa\Delta z_s$ should be satisfied for a signal to be reconstructed completely, along the z_s direction, from its samples in the spectral domain. When the well-known criterion on the maximal transverse sampling interval $b \leqslant 1/2M\kappa\Delta x_s$ is substituted into the longitudinal criterion, we obtain that the maximal sampling interval along the camera's translation should follow the condition $d \leqslant L\Delta x_s/\Delta z_s$.

3. THREE-DIMENSIONAL SPATIAL CORRELATION

To perform a 3-D correlation with the 3-D FT, one needs first to transform the coordinates of $O(u, v, D_x, D_y)$ from the actual space (u, v, D_x, D_y) to the spectral space $(\omega_x, \omega_y, \omega_z)$. Then the transformed function

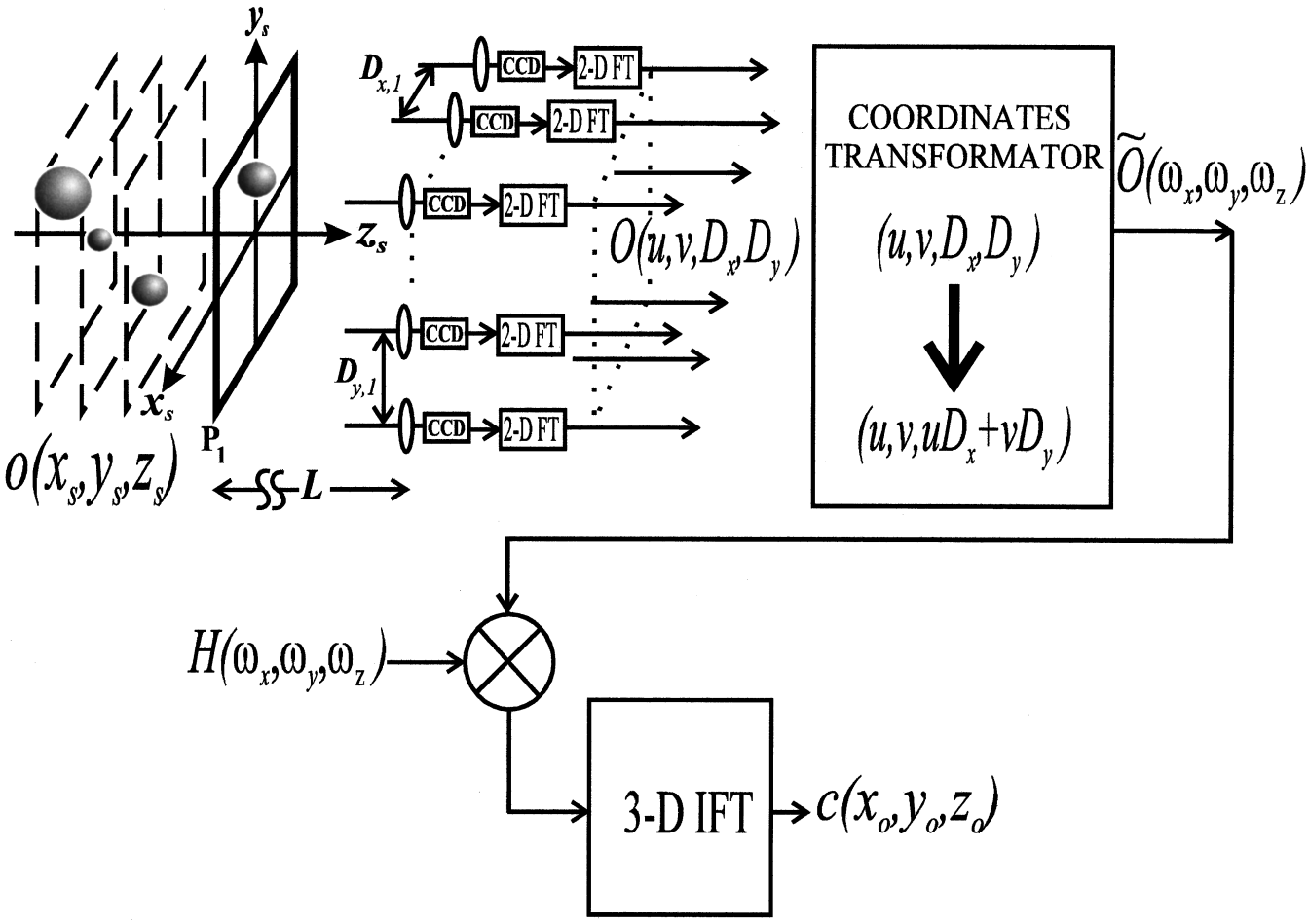

Fig. 2. Schematic of the 3-D spatial correlator. The 3-D scene is observed by an array of cameras distributed on a transverse plane located a distance L from the scene.

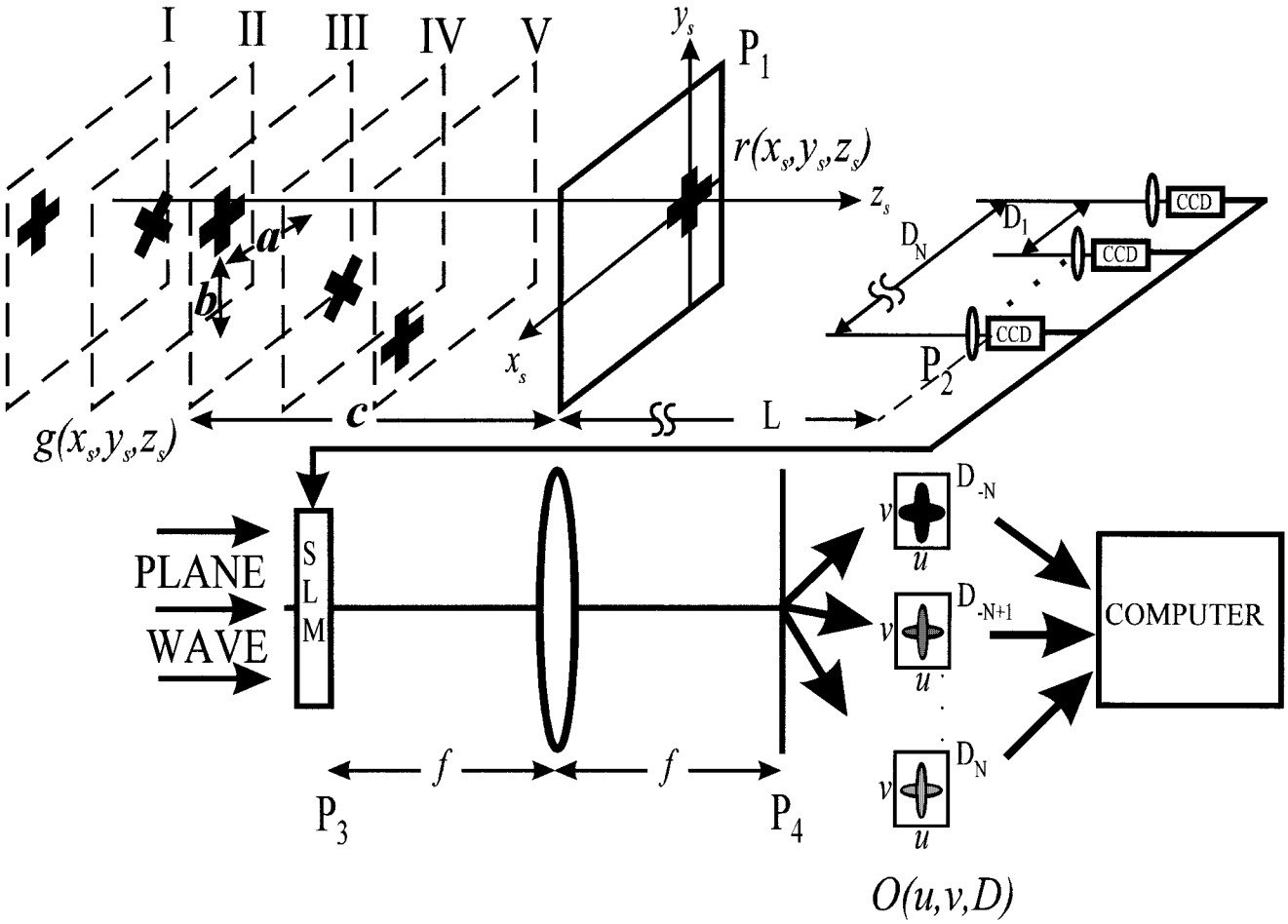

Fig. 3. Schematic of the 3-D JTC. The reference is located on plane P_1 at the origin of the axes. The tested objects are distributed along planes I–V, around the point (a, b, c).

$\tilde{O}(\omega_x, \omega_y, \omega_z)$ is multiplied by a filter function $H(\omega_x, \omega_y, \omega_z)$. Finally, the output correlation result is obtained from the product $\tilde{O}H$ by an inverse 3-D FT. The overall scheme of the 3-D correlation is shown in Fig. 2. Note that in the case $H(\omega_x, \omega_y, \omega_z) \equiv 1$, i.e., the input spectrum is not modified by any filter, the output correlation distribution is a 3-D reconstruction of the observed input pattern. In other words, in the special case of the unit filter, the present system can be used as a reconstruction system for 3-D patterns. This special case will be considered in a future publication.

At least in principle, the scheme described in Fig. 2 might be implemented optically, and thus the 3-D correlation could be executed rapidly and in parallel. However, such a 3-D correlator would become complicated and sensitive to noise. On the other hand, the stage of the multiple 2-D FT's can be done optically without difficulties, whereas it is preferable for the rest of the process to be done electronically, as shown in Fig. 3. The hybrid system offers the best of the two worlds, i.e., the parallelism and the speed of an optical 2-D FT together with the flexibility of the electronic processing. This concept of a hybrid system is considered in the rest of this article.

The problem with the hybrid system arises in the interface between the optics and the electronics. The phase distribution on the output plane of the optical stage is lost when one records the optical intensity by a camera. To overcome this difficulty, without losing the phase information, we adopt the concept of the JTC[7] into our new 3-D correlator. Although the intensity of the spatial spectrum is recorded by the camera, the JTC yields a real correlation between two arbitrary functions without losing the phase information.

The 3-D input space of the JTC contains a reference object $r(x_s, y_s, z_s)$ at some point—say, the origin—and a few tested objects, denoted together by the function $g(x_s, y_s, z_s)$, and located around some other point—say, the point (a, b, c) (see Fig. 3). Therefore the JTC input function is given by

$$o(x_s, y_s, z_s) = r(x_s, y_s, z_s) + g(x_s + a, y_s + b, z_s + c). \quad (6)$$

Substituting Eq. (6) into Eq. (5) and computing the squared magnitude of the complex amplitude $O(u, v, D_x, D_y)$ yield the intensity distribution on the plane P_4:

$$
\begin{aligned}
&I_4(u, v, D_x, D_y) \\
&= \left| R(u, v, D_x, D_y) + G(u, v, D_x, D_y) \right. \\
&\quad \left. \times \exp\left\{ (i2\pi M/\lambda f)\left[au + bv + \frac{c}{L}(D_x u + D_y v) \right] \right\} \right|^2 \\
&= |R(u, v, D_x, D_y)|^2 + |G(u, v, D_x, D_y)|^2 \\
&\quad + G(u, v, D_x, D_y) R^*(u, v, D_x, D_y) \\
&\quad \times \exp\left\{ (i2\pi M/\lambda f)\left[au + bv + \frac{c}{L}(D_x u + D_y v) \right] \right\} \\
&\quad + G^*(u, v, D_x, D_y) R(u, v, D_x, D_y) \\
&\quad \times \exp\left\{ -(i2\pi M/\lambda f)\left[au + bv + \frac{c}{L}(D_x u + D_y v) \right] \right\},
\end{aligned}
\tag{7}
$$

where R and G are 3-D FTs, defined by Eq. (5), of r and g, respectively. The intensity distribution I_4 is recorded by another camera into the computer, in which the coordinate transform of $I_4(u, v, D_x, D_y)$ to $\tilde{I}_4(u, v, D_x u + D_y v)$ is done relatively easily.

After an inverse 3-D FT of $\tilde{I}_4(u, v, D_x u + D_y v)$, the output result is

$$
\begin{aligned}
c(x_o, y_o, z_o) &= \int \tilde{I}_4(\omega_x, \omega_y, \omega_z) \exp[-i2\pi(x_o\omega_x \\
&\quad + y_o\omega_y + z_o\omega_z)] \mathrm{d}\omega_x \mathrm{d}\omega_y \mathrm{d}\omega_z \\
&= r \otimes r + g \otimes g \\
&\quad + (r \otimes g) * \delta(x_o - a, y_o - b, z_o - c) \\
&\quad + (g \otimes r) * \delta(x_o + a, y_o + b, z_o + c),
\end{aligned}
\tag{8}
$$

where $\otimes$ and $*$ stand for the correlation and the convolution, respectively and $\delta(\cdot)$ is the Dirac delta function. Similarly to an ordinary 2-D JTC, the last two terms of Eq. (8) are the cross correlations between the reference and the tested objects. The third and the fourth terms are centered around the points (a, b, c) and $(-a, -b, -c)$, respectively. The first two terms of the autocorrelation in Eq. (8) are centered around the origin. Therefore, if one of the distances (a, b, c) is longer than the respective size of the tested function g, the cross correlation is spatially separated from the autocorrelation terms and becomes detectable. Note that even if two of the three distances (a, b, c) are zero, the desired cross correlation is still separated from the autocorrelations of g and r as long as the third nonzero distance is longer than the respective size of the tested function.

4. SIMULATION RESULTS

We simulated the optical system shown in Fig. 3 by a computer. In our example the input plane contains six objects. The reference in the shape of a cross is located on plane P_1. The other five objects are used as the test patterns, and they are distributed from planes I to V. Three of them on planes I, III, and V are identical to the reference and therefore should be recognized by the system. The other two images on planes II and IV are in the shape of an X, which is different from the reference and should be ignored by the system. We can also see in Fig. 3 the location of the camera in three states $D = 0$, D_1, D_N, while D_N is the maximum displacement of the camera to the left side. To simplify the process in this simulation, the camera is shifted only along the x axis $2N$ displacements, N for each side, where $N = 24$. Therefore in this example the longitudinal frequency variable is $\omega_z = MD_n u/\lambda f L$, where $n = 0, \pm 1, \ldots \pm N$. We also use the following relations: $M/\lambda f = 1$ and $D_N/L = 0.1$. The dimensions of the input space are $\Delta x_s = \Delta y_s = 0.025L$ and $\Delta z_s = 0.1L$. Because of memory limitations of our computer, we used objects with a depth of one pixel only. Simulations with other thicker objects are given in Ref. 8.

Figures 4–8 are the summary of the simulation results. Figure 4 presents 5 out of 49 images taken by the camera from a few positions along the baseline. These images, and all other 44 images with other D_n values, are Fourier

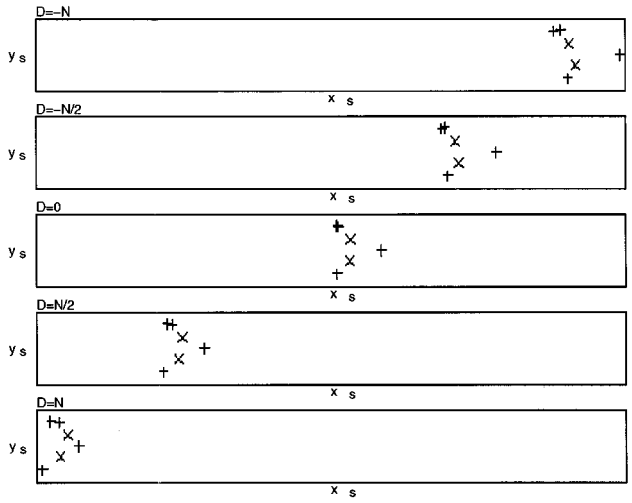

Fig. 4. Five projections, out of 49, of the input space as seen from five different points of view along the camera's baseline.

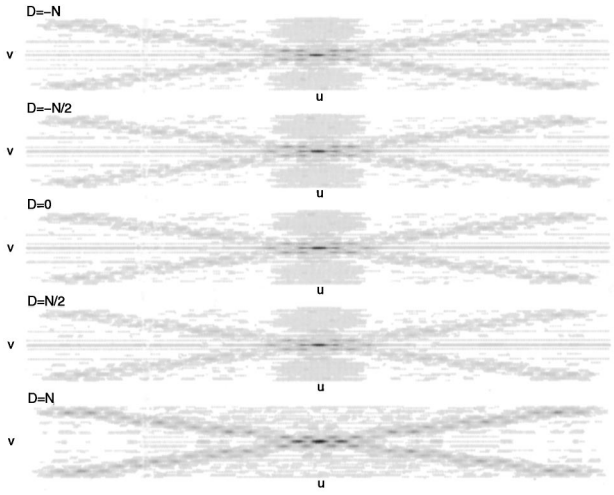

Fig. 5. Intensity distribution on plane P_4 obtained by 2-D FT of each projection of Fig. 4.

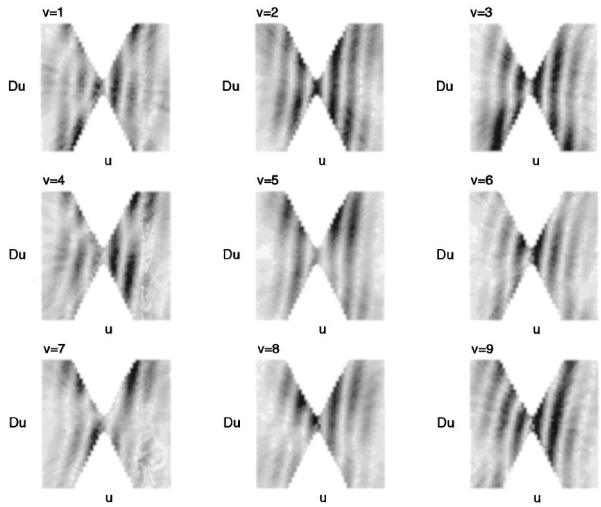

Fig. 6. 3-D spectral intensity distribution after the coordinate transformation from (u, v, D_n) to (u, v, uD_n).

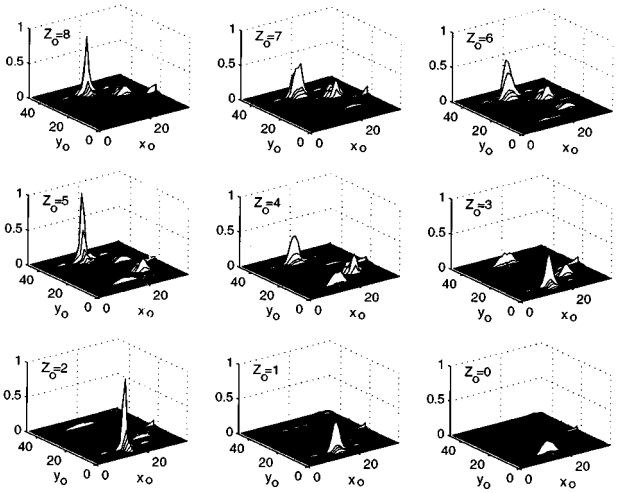

Fig. 7. Intensity of the cross correlation between the reference with the shape of the cross and the test objects obtained by a 3-D inverse FT of $\tilde{I}_4(u, v, uD_n)$ that is partly shown in Fig. 6. The three high peaks indicate the existence of the three crosses among the group of the tested objects.

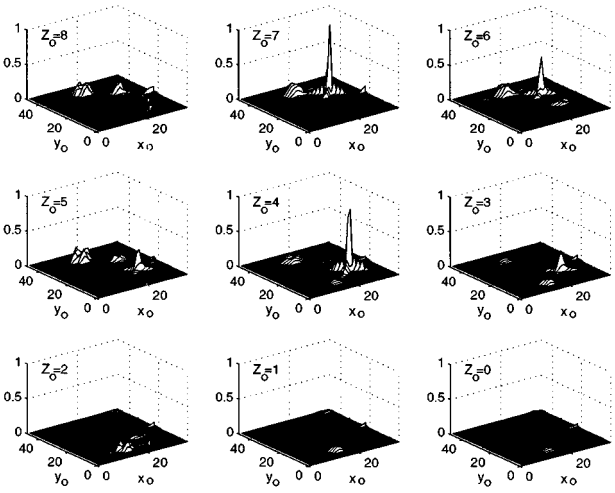

Fig. 8. Same as Fig. 7, except that this time the reference is in the shape of an X.

transformed. The intensity of the respective FT's of the images in Fig. 4 are shown in Fig. 5. From the set of 49 2-D FT's we compose the 3-D spectrum using the coordinates transform from (u, v, D_n) to (u, v, uD_n). Nine samples of the obtained 3-D spectrum $\tilde{I}_4(u, v, uD_n)$, for nine values of v, are shown in Fig. 6. The horizontal lines D_n = constant, v = constant in Fig. 5 are transformed into diagonal lines, with a slope D_n, that cross through the origin in Fig. 6.

A 3-D FT of $\tilde{I}_4(u, v, uD_n)$ (shown in Fig. 6) yields the required cross correlation between the reference and the tested objects. A collection of 3-D plots of the output space around the region of the first diffraction order [third term in Eq. (8)] are shown in Fig. 7. The output correlation space is given in the coordinates (x_o, y_o, z_o), which are equivalent to the input coordinates (x_s, y_s, z_s). Each 3-D plot in Fig. 7 presents the transverse intensity distribution at some z_o. The three strong correlation peaks on planes $z_o = 2$, 5, and 8 indicate the locations of the three recognized crosses, which are identical to the reference. The other less-strong peaks on the other planes are the tails of the above-mentioned identification peaks. In the same manner that a correlation peak has a finite width along the x_o or y_o axes, it also has a width along the z_o axis. For instance, the strongest peak on plane $z_o = 3$ is part of the tail of the autocorrelation peak whose maximum is seen on plane $z_o = 2$. The same relation exists between the strong peak on plane $z_o = 5$ and its tails on planes $z_o = 4$ and $z_o = 6$. The cross correlation between the reference and the two X-shaped objects can hardly be seen on planes $z_o = 4$ and $z_o = 7$ along the line $x_o = 20$. These peaks are less than 30% of the autocorrelation peaks.

The same demonstration was repeated with a different X-shaped reference object while the tested objects were introduced to the system exactly as in Fig. 3. The results of the correlation are shown in Fig. 8. This time, as expected, the two strongest peaks are on planes $z_o = 4$ and $z_o = 7$, indicating the presence of the X shapes in these planes. We can conclude this demonstration by saying that from the five input objects the three crosses were recognized, and their exact locations in the 3-D space were identified when the reference object was the cross. When the reference was changed to the X shape, the two corresponding objects were recognized in their locations.

5. CONCLUSION

We have developed a method for electro-optical 3-D spatial correlation. The correlation process contains a series of 2-D FT's, a coordinate transform, multiplication by a filter, and finally a numerical inverse 3-D FT. Alternatively, the 3-D correlation can be done in a JTC configuration in which we apply the series of 2-D FT's jointly on the test and reference objects. Then the coordinate transform is operated on the spectral intensity, and finally we obtain the correlation output using a numerical 3-D FT. The 3-D JTC's were demonstrated by computer simulation, in which objects seen from the paraxial zone were recognized in their 3-D natural space. Other applications of 3-D spatial filtering are expected in the near future.

REFERENCES

1. B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," Appl. Opt. **5**, 967 (1966).
2. J. W. Goodman, *Introduction to Fourier Optics*, 1st ed. (McGraw-Hill, New York, 1968), Chap. 7, p. 171.
3. E. C. Tam, F. T. S. Yu, D. A. Gregory, and R. D. Juday, "Autonomous real-time object tracking with an adaptive joint transform correlator," Opt. Eng. **29**, 314–320 (1990).
4. J. Rosen and A. Yariv, "Three-dimensional imaging of random radiation sources," Opt. Lett. **21**, 1011–1013 (1996); J. Rosen and A. Yariv, "Reconstruction of longitudinal distributed incoherent sources," Opt. Lett. **21**, 1803–1805 (1996).
5. A. Papoulis, *System and Transforms with Applications in Optics*, 1st ed. (McGraw-Hill, New York, 1968), Chap. 3, p. 61.
6. M. Born and E. Wolf, *Principles of Optics*, 5th ed. (Pergamon, Oxford, 1975), Chap. 4, p. 133.
7. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," Appl. Opt. **5**, 1248–1249 (1966).
8. J. Rosen, "Three-dimensional optical Fourier transform and correlation," Opt. Lett. **22**, 964–966 (1997).