Three-dimensional joint transform correlator

Joseph Rosen

The three-dimensional (3-D) joint transform correlator is demonstrated with realistic targets. Three-dimensional objects observed by multiple cameras are correlated with a 3-D reference object. The number of cameras and their directions of observation are particularly considered. © 1998 Optical Society of America

OCIS codes: 070.2580, 070.2590, 070.4550, 070.5010, 070.6110, 100.0100, 100.6890.

1. Introduction

Three-dimensional (3-D) optical correlators open opportunities for processing 3-D images directly and rapidly. Targets distributed in 3-D space can be recognized or tracked by optical correlators in the same fast and parallel manner that the well-known two-dimensional (2-D) correlators[1] have demonstrated for the past three decades. A method for performing 3-D optical correlation was proposed by Bamler and Hofer-Alfeis[2] and revisited in Ref. 3. In these studies the 3-D observed scene is first mapped, slice by slice, onto a 2-D plane. Then a conventional 2-D optical correlation, with an increased space–bandwidth product, is performed. However, the 3-D distribution of the observed scene must be known *a priori* to the digital computing system, prior to the mapping stage on the 2-D plane. In other words, the scene must be processed with intensive digital algorithms to reconstruct the 3-D image inside the computer memory before any correlation can be employed. The advantages of the optical processing, namely, the directness and the high processing speed, vanish in such schemes.

Recently[4,5] a process of direct 3-D electro-optical correlations between two 3-D real-world functions was developed, without the need for any preprocessing stage. This 3-D correlation was demonstrated by simulation of a 3-D extension of the well-known 2-D joint transform correlator (JTC).[6] In the proposed scheme, a reference object and tested objects are observed from a single distant transverse plane. Several cameras distributed on this observation plane record 2-D projections of the 3-D input scene from various points of view. The correlation process in the JTC configuration contains a series of optical 2-D Fourier transforms (FT's) applied jointly on the reference object and the tested objects. Then the accumulated intensity distributions of the 2-D FT's are mapped on a 3-D spectral space, and, finally, the correlation output is obtained by use of an additional 3-D FT. The output result of this process is a 3-D correlation between the reference object and the tested objects. This algorithm is composed of a series of automatic image-processing operations. Some of these operations can be implemented optically, and together they yield the 3-D correlation between any two arbitrary functions. The key concept of the 3-D correlation is first to implement electro-optically the 3-D FT of the observed scene. Then, with the convolution theorem, the desired 3-D correlation result is obtained. The 3-D correlator enables one to identify a target from its depth pattern as well as from its regular 2-D information. Of no less importance is that, instead of knowing the target's location in only the 2-D scene, we learn its position in 3-D space.

One main drawback of the system proposed in Ref. 5 is the need for cameras with wide fields of view in the extreme points of view. All the directions of observation for all the points of view are parallel to the same longitudinal axis. Therefore the field of view must be much wider than the total width of the observed objects [see Fig. 1(a)]. Otherwise, one cannot shift the point of view far from the origin without missing parts of the objects in the extreme frames. In this paper the previously proposed system is modified to overcome this drawback. The first experiment of the 3-D JTC with real-world objects is also

The author (rosen@eesrv.ee.bgu.ac.il) is with Ben-Gurion University of the Negev, Department of Electrical and Computer Engineering, P.O. Box 653, Beer-Sheva 84105, Israel.

Received 11 February 1998; revised manuscript received 6 May 1998.

0003-6935/98/327538-07$15.00/0

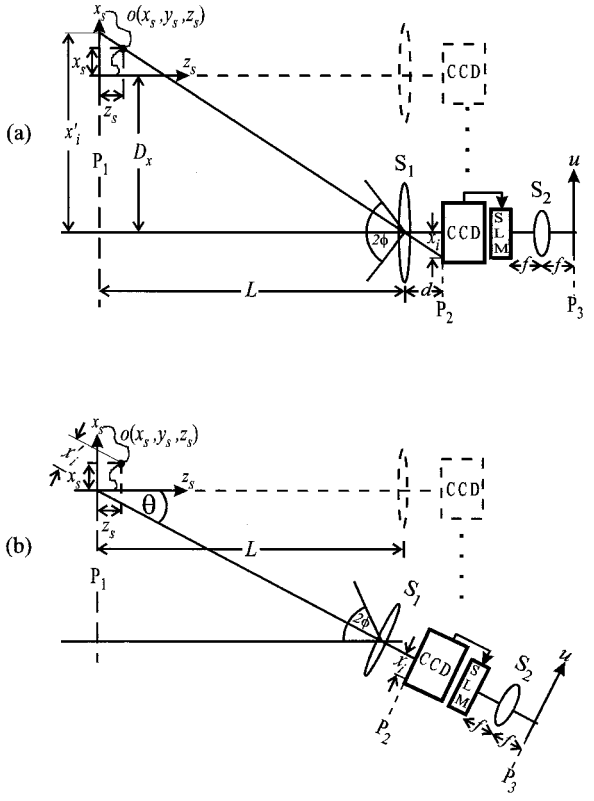

Fig. 1. Illustration of the imaging system in the case of (a) parallel and (b) converging observations.

described. Finally, the relation between the number of observation points and the quality of the correlation results is examined experimentally. In other words, the question of how much one can reduce the number of 2-D projections and still receive reasonably acceptable results is asked.

2. Analysis

The two options for observing the input scene by multiple cameras from some distance are shown in Fig. 1. In both configurations the plane $z_s = 0$, located at a distance L from the imaging lens S_1, is imaged on the camera with a magnification factor M_0. For a distance d between the lens S_1 and the camera, M_0 satisfies the relation $M_0 = d/L$. Therefore any arbitrary point (x_s, y_s, z_s) of the observed scene is imaged to the point (x_i, y_i) on plane P_2 with a magnification factor of

$$M(z_s) = \frac{d}{L - z_s} = \frac{M_0}{1 - z_s/L}. \tag{1}$$

Plane P_2 is two-dimensionally Fourier transformed onto plane P_3. Thus the point (x_s, y_s, z_s) with brightness o is transformed into a linear phase function of the form $[M(z_s)]^2 o \exp[i(k/f)(x_i u + y_i v)]$, where (u, v) are the transverse coordinates of plane P_3, $k = 2\pi/\lambda$, λ is the wavelength, and f is the focal length of the Fourier lens S_2. Under the far-field approximation, we assume that $L \gg \Delta z_s$, where Δz_s is the existence interval of the input function along z_s. Thus we can assume that all the object points are imaged with the same magnification factor [i.e., Eq. (1) is approximated such that $M(z_s) \approx M_0, \forall z_s \in \Delta z_s$]. The overall distribution on plane P_3 is given by integration over the linear phases contributed by all the object points, as follows:

$$O_3(u, v) = M_0^2 \iiint o(x_s, y_s, z_s) \exp[i(k/f)(x_i u + y_i v)] \mathrm{d}x_s \mathrm{d}y_s \mathrm{d}z_s, \tag{2}$$

where $o(x_s, y_s, z_s)$ is the 3-D input function in the coordinate system (x_s, y_s, z_s).

In the parallel observation shown in Fig. 1(a), the dependence of (x_i, y_i) on M_0, L, (x_s, y_s, z_s), and the lateral shift of the camera from its central position D_x are given in Ref. 5. On the basis of this relation and if we assume the far-field approximation, the 3-D function accumulated on plane P_3 from all the 2-D projections is[5]

$$O_3(u, v, D_x) = A \exp(iM_0 k D_x u/f) \times \iiint o(x_s, y_s, z_s) \exp[i(M_0 k/f)(x_s u + y_s v + z_s D_x u/L)] \mathrm{d}x_s \mathrm{d}y_s \mathrm{d}z_s, \tag{3}$$

where A is a complex constant. The distribution of Eq. (3) is used, after a proper coordinate transform, as the 3-D spatial spectrum of the input scene in the 3-D JTC.[5]

According to Fig. 1(a), for a maximum lateral shift $D_{x,\max}$ and a maximum total object width Δx_s, the camera field angle of 2ϕ must satisfy the condition $\tan\phi \geq (D_{x,\max} + \Delta x_s/2)/L$ to prevent loss of object information. For a given camera field angle, limitations were set by the sizes of the objects that can be processed or by the amount of the lateral shift that can be employed. The lateral shift $D_{x,\max}$ determines the longitudinal resolution of the system,[5] and this resolution is a merit that we always wish to increase.

To overcome this limitation, we examine the converging observation as an alternative configuration. In this architecture, shown in Fig. 1(b), the detection plane P_2 of each camera is orthogonal to the line that connects the origin points of the object space and the detection plane. In this case the field angle is independent of the lateral shift D_x, and therefore the field of view can be as narrow as the total object width. Formally, the condition that must be satisfied in this case is only $\tan\phi \geq \Delta x_s/2L$. However, Eq. (3) is not valid for the converging observation, and we should recalculate the 3-D accumulated distribution on plane P_3, as described below.

To express the relation between the 3-D input function and the distribution of the Fourier space for the new configuration, we first look again 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).

The location of the imaged point (x_i, y_i) depends on the coordinates (x_s, y_s, z_s) and on the angle θ between the longitudinal axis z_s and the direction of observation [see Fig. 1(b)]. According to Fig. 1(b) and if we assume that plane P_1 is imaged with a magnification factor M_0, the location of the imaged point on plane P_2 is

$$(x_i, y_i) = M_0(x_s \cos\theta + z_s \sin\theta, y_s). \tag{4}$$

Each imaged point (x_i, y_i) is Fourier transformed to a 2-D linear phase function. The overall distribution on the Fourier plane P_3 from the entire 3-D object $o(x_s, y_s, z_s)$ (under the assumption that $L \gg \Delta z_s$) is obtained by integration over the linear phases contributed by all object points, as given by Eq. (2). Substituting Eq. (4) into Eq. (2) yields

$$O_3(u, v, \theta) = A \iiint o(x_s, y_s, z_s)\exp[i(kM_0/f)(ux_s \cos\theta + vy_s + uz_s \sin\theta)]\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}z_s. \tag{5}$$

Equation (5) is similar to a 3-D FT, which transforms an object function $o(x_s, y_s, z_s)$ into a 3-D spatial-frequency function $O_3(f_x, f_y, f_z)$, where $f_x = M_0 u \cos\theta/\lambda f$, $f_y = M_0 v/\lambda f$, and $f_z = M_0 u \sin\theta/\lambda f$. Note that f_x and f_z are both dependent on the transverse variable u and on the angle θ. This dependence distinguishes Eq. (5) from a pure 3-D FT, but apparently we can obtain 3-D spectral functions of three independent orthogonal spatial-frequency variables with a proper coordinate transform, as discussed below.

Note that, in distant imaging, whereas $L \gg D_x$, the approximations $\cos\theta \approx 1$ and $\sin\theta \approx \theta \approx D_x/L$ are valid. In this case Eq. (5) is reduced to the form of Eq. (3) without the linear phase function given before the integral. Therefore, in comparison with the parallel observation, the converging observation is less restricted by the distance between the imaging system and the objects. However, in the regime of close imaging we have to map the function $O_3(u, v, \theta)$ from coordinates (u, v, θ) to $(u \cos\theta, v, u \sin\theta)$. This mapping is more complicated than the mapping from $(u, v, D_x/L)$ to $(u, v, uD_x/L)$, which is suitable for both cases, i.e., the parallel and the converging observations in the regime of distant imaging. Mapping the function $O_3(u, v, \theta)$ on the 3-D spatial spectral space enables us, by an additional 3-D FT, to obtain the desired 3-D correlation distribution.

A conventional linear 2-D JTC[6] is basically composed of three sequential mathematical operations: a FT of the input scene, calculation of the square magnitude of the FT, and finally another FT. In our 3-D JTC we add an additional operation of coordinate transformation before the final 3-D FT in which every plane in the space (u, v, D_x) with a constant $D_x = \tilde{D}_x$ is mapped to a tilted plane with the slope $(\tilde{D}_x, 0)$ in the space (u, v, uD_x). It is suggested that such mapping be implemented electronically, either with software, as was done in the present study, or with electronic hardware.

The complete process of the 3-D JTC is shown schematically in Fig. 2. 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). Therefore the JTC input function is given by

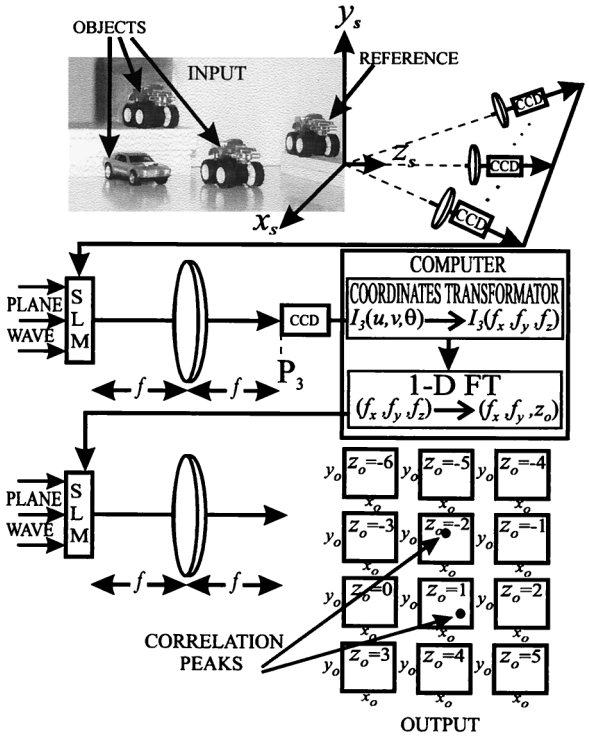

Fig. 2. Schematic of the 3-D JTC.

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

Following the above-mentioned analysis, we can conclude that, for converging observation from the far-field, the obtained 3-D function from the set of 2-D intensity images on plane P_3 for various angles θ is

$$I_3(u, v, \theta) = |O_3(u, v, \theta)|^2 = \left|A \iiint o(x_s, y_s, z_s)\exp[i(M_0 k/f)(x_s u + y_s v + z_s D_x u/L)]\mathrm{d}x_s\mathrm{d}y_s\mathrm{d}z_s\right|^2. \tag{7}$$

This intensity expression [Eq. (7)] is obtained by substitution of the far-field approximation $\sin\theta \approx \theta \approx D_x/L$ into the absolute squared value of the amplitude of Eq. (5). Substituting Eq. (6) into Eq. (7) yields

$$I_3(u, v, \theta) = |R(u, v, \theta) + G(u, v, \theta)\exp[i(M_0 k/f) \times (au + bv + cD_x u/L)]|^2, \tag{8}$$

Fig. 3. Twelve projections out of 25 of the input scene as observed from different points of view along the baseline.

where R and G are the 3-D FT's of r and g, respectively, and are defined by

$$\begin{pmatrix} R \\ G \end{pmatrix}(u, v, \theta) = \iiint \begin{pmatrix} r \\ g \end{pmatrix}(x_s, y_s, z_s)\exp[i(M_0 k/f) \times (x_s u + y_s v + z_s D_x u/L)]\mathrm{d}x_s \mathrm{d}y_s \mathrm{d}z_s. \quad (9)$$

The intensity distribution I_3 is recorded by another camera into the computer, in which the coordinate transform from $(u, v, \theta \approx D_x/L)$ to $(u, v, uD_x/L)$ is performed with proper software. The obtained function in the new coordinate system is

$$\begin{aligned} \tilde{I}_3(u, v, uD_x/L) &= |\tilde{R}(u, v, uD_x/L) \\ &\quad + \tilde{G}(u, v, uD_x/L)\exp[i(M_0 k/f) \\ &\quad \times (au + bv + cD_x u/L)]|^2 \\ &= |\tilde{R}(u, v, uD_x/L)|^2 + |\tilde{G}(u, v, uD_x/L)|^2 \\ &\quad + \tilde{R}(u, v, uD_x/L)\tilde{G}^*(u, v, uD_x/L) \\ &\quad \times \exp[-i(M_0 k/f)(au + bv \\ &\quad + cD_x u/L)] + \tilde{G}(u, v, uD_x/L) \\ &\quad \times \tilde{R}^*(u, v, uD_x/L)\exp[i(M_0 k/f) \\ &\quad \times (au + bv + cD_x u/L)], \end{aligned} \quad (10)$$

where $\tilde{R}$ and $\tilde{G}$ are the transformed functions R and G in the new spatial-frequency coordinates given by $f_x = M_0 u/\lambda f$, $f_y = M_0 v/\lambda f$, and $f_z = M_0 u D_x/L\lambda f$. Without this mapping to the 3-D spatial-frequency space, we cannot use the convolution theorem and get the desired 3-D correlation. After another 3-D FT of $\tilde{I}_3(f_x, f_y, f_z)$, the output result is

$$\begin{aligned} c(x_o, y_o, z_o) &= \iiint \tilde{I}_3(f_x, f_y, f_z) \\ &\quad \times \exp[-i2\pi(x_o f_x + y_o f_y + z_o f_z)]\mathrm{d}f_x \mathrm{d}f_y \mathrm{d}f_z \\ &= r \otimes r + g \otimes g + (r \otimes g) \\ &\quad * \delta(x_o - a, y_o - b, z_o - c) \\ &\quad + (g \otimes r) * \delta(x_o + a, y_o + b, z_o + c), \end{aligned} \quad (11)$$

where the symbol $\otimes$ and the asterisk stand for the 3-D correlation and the 3-D convolution, respectively, and $\delta(\cdots)$ is the Dirac delta function. Similar to an ordinary 2-D JTC,[6] the last two terms of Eq. (11) are the cross correlations between the reference object and the tested objects. The third and fourth correlation terms are centered around the points (a, b, c)

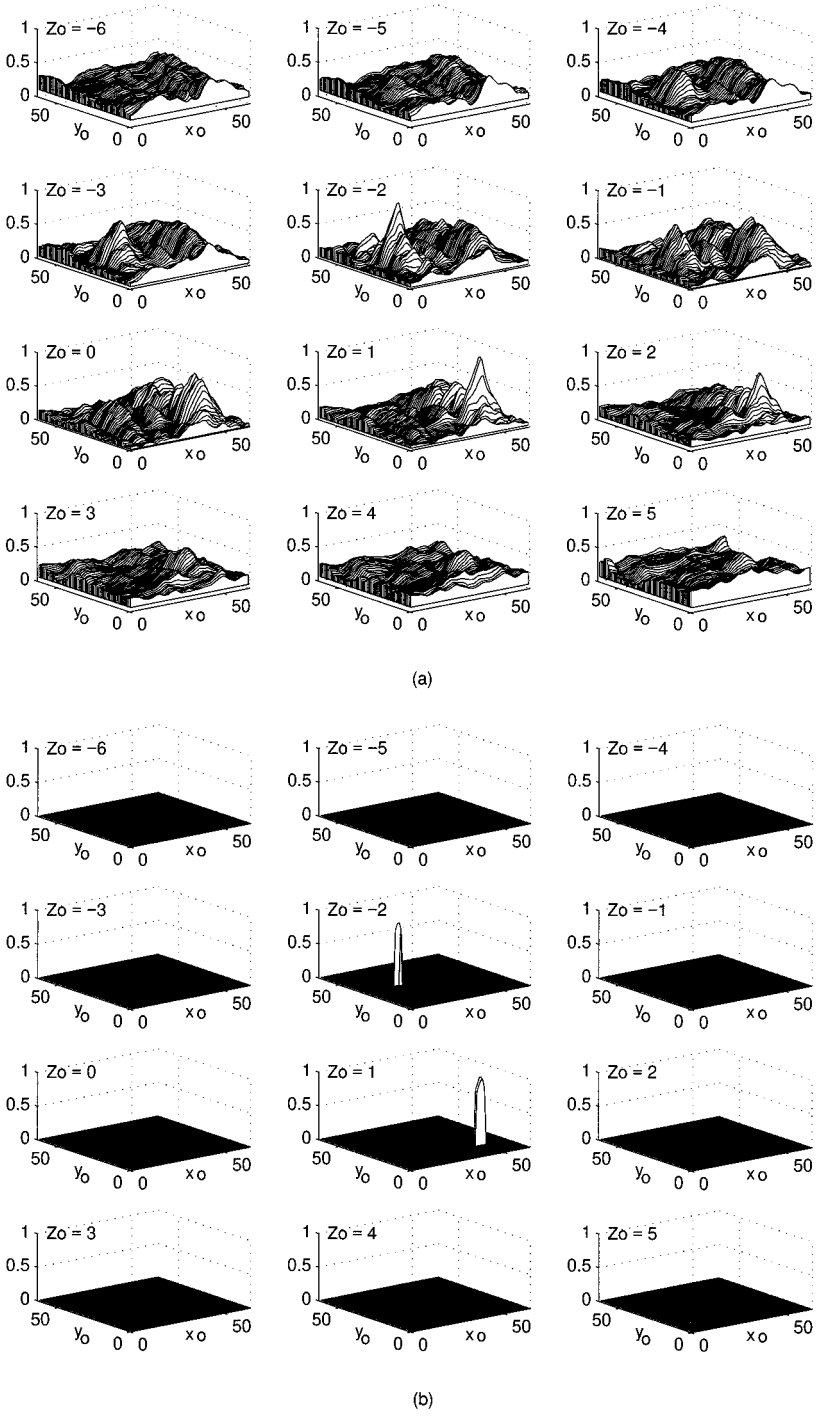

Fig. 4. Intensity of the correlation results of the 3-D joint transform correlator (a) before and (b) after a threshold operation.

and $(-a, -b, -c)$, respectively. Each term can be written explicitly as

$$(g \otimes r)(x_o, y_o, z_o) = \iiint g(x, y, z) \times r(x - x_o, y - y_o, z - z_o)\mathrm{d}x\mathrm{d}y\mathrm{d}z. \tag{12}$$

The first two terms of the autocorrelation in Eq. (11) 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 thus becomes detectable.

There are two new elements in the present process, shown schematically in Fig. 2, in comparison with the previous schemes.[4,5] First, the camera observation is modified to be converging instead of parallel as discussed above. Second, to employ the full processing power of optics, it is suggested that the final 3-D

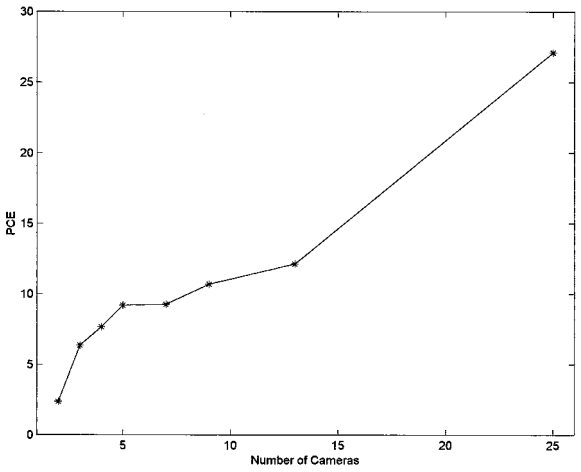

Fig. 5. The peak-to-correlation energy (PCE) versus the number of cameras along the baseline.

FT be split into an electronic one-dimensional FT from f_z to z_o followed by multiple optical 2-D FT's from (f_x, f_y, z_o) to (x_o, y_o, z_o).

3. Experimental Results

In our example the input scene contains four vehicles, as shown in Fig. 2. One of them, the reference, is located on the right-hand side of the scene and is identical to two cars from the left-hand group of three vehicles used here as the observed objects. Note that the two lower vehicles are located in front of the reference, whereas the upper vehicle is in back of it. The system should recognize the two cars that are identical to the reference and ignore the other vehicle. Three locations of the CCD cameras are also shown in Fig. 2. In this experiment a single camera was shifted along the x axis by 24 equal displacements, 12 for each side. Each projection was recorded with the camera and Fourier transformed. A temporary lack of a spatial light modulator (SLM) in the laboratory made it necessary to calculate all the 2-D FT's in this project with a digital computer. The intensity of all the 2-D FT's on plane P_3, designated as $I_3(u, v, \theta)$, was stored in the computer. Note that the reference can first be recorded alone (off line) without the observed objects. Then, at the stage of target recognition, each reference projection is displayed on the SLM beside the corresponding object projection. Alternatively, all the reference projections can be created synthetically by some computer algorithm[7] before they are displayed on the SLM beside the object projections.

Figure 3 presents 12 out of 25 images taken with the camera from every even position along the baseline. In each image the group of three tested objects is located on the left-hand side, while the reference is separated on the right-hand side. From the set of 25 intensity patterns of 2-D FT's, the 3-D spectrum $\tilde{I}_3(u, v, uD_x/L)$ is composed with the coordinate transform from (u, v, θ) to $(u, v, uD_x/L)$, where $\theta \cong D_x/L$. This transformation is justified in this case of converging observation because the value of θ in this experiment is no more than 15°. An additional 3-D FT of $\tilde{I}_3(u, v, uD_x/L)$ yields the required 3-D cross correlation between the reference object and the tested objects in the first diffraction order. In the present experiment the JTC yields a linear 3-D correlation, as defined in Eq. (11). No additional processing (such as binarization, dc removal, etc.) was performed on the joint spectrum $\tilde{I}_3(u, v, uD_x/L)$. However, these operations can be applied in the future with the same considerations as for the known 2-D JTC. Note that this process is 3-D shift invariant but is not a distortion-invariant recognition. Therefore a recognition peak is obtained only when the reference object and the observed object are identical in structure, size, and orientation. In other words, there is no need to train or to synthesize a special reference function. The reference is simply an exact copy of the object to be recognized. The orientation of the left-hand vehicles seems different from that of the right-hand reference only because their locations in relation to the camera are different. Also, it should be emphasized that Fig. 3 does not show an input to 12 separated 2-D JTC's. It does show the input to a single 3-D JTC, in which the high peaks in the output 3-D space (see Fig. 4) indicate the presence of recognized objects in the 3-D input space.

Three-dimensional plots of the output space around the region of the first diffraction order are shown in Fig. 4(a). The output correlation space is given by 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 of Fig. 4 presents the transverse intensity distribution at some z_o. The correlation distribution above the threshold value of 0.8 is depicted in Fig. 4(b). The two strong correlation peaks on planes $z_o = -2, 1$ indicate the locations of the two recognized vehicles, which are identical with the reference.

One of the main considerations of this system is the number of 2-D projections required for a proper operation. To save resources and computation time, as few projections as possible should be used. However, it is clear that some penalty in the sense of performance reduction should come with reducing the number of cameras. This issue is examined by measurement of the peak-to-correlation energy[8] (PCE) versus the camera number, as shown in Fig. 5. Reducing the number of cameras means reducing, by the same amount, the number of tilted planes [with the slopes of $(\tilde{D}_x, O)$] in the 3-D spatial spectral space without changing other dimensions in system. The PCE is defined as the ratio between the correlation-peak intensity and the average intensity over the entire distribution on the correlation plane. For the present experiment we find that correct recognition is achieved within only two projections, i.e., by recording of the scene with only two cameras from two points of view. However, increasing the reliability of the system demands increasing the number of projections. The reliability is expressed here by the PCE, since higher values of the PCE mean a higher

probability for the system to identify the correct correlation peaks from among their noisy vicinities.

4. Conclusions

An electro-optical 3-D spatial correlator has been demonstrated experimentally. The previously proposed system[5] was modified to eliminate the need for cameras with wide fields of view. The penalty that comes with the method of converging observation is the need for careful camera calibration to guarantee that all cameras are focused on the same point in the observed scene.

The effect of reducing the number of cameras has also been considered. Recording finite discrete numbers of projections is equivalent to using partial data from the complete 3-D spectrum. Using partial data from the spectrum in conventional imaging systems usually results in degradation and blurring of the output image. Since the single goal of the correlator is to emit correlation peaks in the proper locations, the demand of transferring maximum information about the spectral distribution can be relaxed. This explains why two projections in this experiment are enough to yield a correct correlation result, although with a low signal-to-noise ratio. It is expected that, for more targets in the input scene and therefore with a more complicated correlation pattern, the desired minimum number of points of view should be increased to avoid a mistaken recognition.

The author thanks Boaz Salik and Yevgeny Karasik for fruitful discussions.

References

1. A. B. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory **IT-10,** 139–145 (1964); B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Appl. Opt. **31,** 4773–4801 (1992).
2. R. Bamler and J. Hofer-Alfeis, "Three- and four-dimensional filter operations by coherent optics," Opt. Acta **29,** 747–757 (1982).
3. Y. Karasik, "Evaluation of three-dimensional convolutions by two-dimensional filtering," Appl. Opt. **36,** 7397–7401 (1997).
4. J. Rosen, "Three-dimensional optical Fourier transform and correlation," Opt. Lett. **22,** 964–966 (1997).
5. J. Rosen, "Three-dimensional electro-optical correlation," J. Opt. Soc. Am. A **15,** 430–436 (1998).
6. C. S. Weaver and J. W. Goodman, "A technique for optically convolving two functions," Appl. Opt. **5,** 1248–1249 (1966); F. T. S. Yu and S. Jutamulia, *Optical Signal Processing, Computing, and Neural Networks* (Wiley, New York, 1992), Chap. 2, p. 34.
7. J. Rosen, U. Mahlab, and J. Shamir, "Complex reference discriminant functions implemented iteratively on joint transform correlator," Appl. Opt. **30,** 5111–5115 (1991).
8. J. L. Horner, "Metrics for assessing pattern-recognition performance," Appl. Opt. **31,** 165–166 (1992).