Signal Detection By Complex Spatial Filtering

A. VANDER LUGT

Summary—**In the past, spatial filtering in coherent optical systems has been limited by the inability to realize practically a general complex filter. This paper describes a technique for realizing such a filter, and gives an application of spatial filtering to the problem of detecting isolated signals in a variety of noise backgrounds. The experimental results obtained to date indicate that this technique provides an excellent two-dimensional filtering capability that will play a key role in such problems as shape recognition and signal detection.**

Introduction

THE FUNDAMENTAL theory of optical spatial filtering has been formulated by several writers.[1,2] The close analogy of spatial filtering in coherent optical systems to optimum linear filtering in communication theory promised to open up new techniques for realizing those frequency domain filters which could not be synthesized in the time domain because of the realizability constraint; but spatial filtering has not advanced as far as one might have expected. Perhaps the main reason is that, though the formulation of the theory tacitly assumed that complex filters could be realized, the practical realization has presented a difficult problem. A major part of this problem is that photographic film, the most prominent candidate for use in the realization of spatial filters, can record only non-negative functions, so that auxiliary phase filters must be used to realize the complex filters.

The modulus of the transfer function for spatial filters is restricted to values between zero and one.[3] To demonstrate the theory, early experiments in spatial filtering used occluding filters with transfer functions having the values (0, 1) as band-pass filters.[2,3] Later, continuous amplitude control (*e.g.*, Gaussian weighting) was used to show how equalization could be accomplished in a coherent optical system. The next step was to add binary phase control to extend the range of filter values to the real axis between plus and minus one.[4,5] The binary phase functions were constructed by using ruled phase plates, evaporative techniques, or film reliefing techniques. These techniques are rather awkward to apply. Clearly, if more general filtering operations are to be performed, a fairly easy method for constructing a general complex filter must be found. This paper describes such a technique and gives some experimental results of major importance. This technique is especially attractive because it realizes the complex filter by a special process for recording on film without the use of an auxiliary phase filter.

Mathematical Nature of the Problem

The mathematical model shown in Fig. 1 describes the problem of sensing, recording and processing imagery. The scene will be denoted by $g(x, y)$, an intensity function of two space coordinates. The primary sensor makes a two-dimensional transformation of $g(x, y)$, and the recording process causes a further transformation. Because the recording medium is usually film, this transformation introduces film grain as a secondary source of noise. It has become common practice to refer to $f(x, y) = T_2\{T_1[g(x, y)]\}$ as the "scene." Though this is not strictly correct, transformation T_2 is usually a nonlinear process (and a poorly controlled one at that). Therefore, it becomes necessary to operate on the available data $f(x, y)$ rather than to attempt to recover $g(x, y)$ before the processing operation.

Processing of the imagery could involve several distinctly different operations, depending on the type of information desired. One such operation might be to differentiate the signal to emphasize large gradients in the image. In other cases, band-pass filtering will improve the imagery. The particular problem treated in this paper is that of maximizing the ratio of peak signal energy to mean square noise energy in the output of the processing system. The signal is any object of interest in the scene, and the noise is the rest of the imagery. This problem is familiar in signal detection theory, and the mathematical solution is well known. Suppose $n(x, y)$ is a homogeneous, isotropic random process with spectral density $N(p, q)$, and $s(x, y)$ is a known Fourier transformable function of space coordinates. If $h(x, y)$ denotes the impulse response of a linear space invariant filter with frequency response $H(p, q)$, we wish to find the optimum linear filter which operates on $f(x, y) = s(x, y) + n(x, y)$ to maximize the ratio of peak signal to rms noise.

Several writers have given the solution to this problem, but usually for the case where $N(p, q)$ is uniform. The general solution[6] is that the required filter response is

Manuscript received May 31, 1963. This work was jointly supported by Project MICHIGAN, under U. S. Army Electronics Command, Contract DA-36-039 SC-78801, and by Aeronautical Systems Division, Air Force Contract AF 33(616)-8433. A more detailed discussion of this subject is given in Report 2900-394-T/4594-22-T.

The author is with the Institute of Science and Technology, The University of Michigan, Ann Arbor, Mich.

[1] T. P. Cheatham, Jr. and A. Kohlenberg, "Optical filters: their equivalence to and differences from electrical networks, 1954 IRE Convention Record, pt. 4, pp. 6–12.

[2] E. O'Neill, "Spatial filtering in optics," IRE Trans. on Information Theory, vol. IT-2, pp. 56–65; June, 1956.

[3] L. J. Cutrona *et al.*, "Optical data processing and filtering systems," IRE Trans. on Information Theory, vol. IT-6, pp. 386–400; June, 1960.

[4] A. Marechal, "Filtering of optical images," in "*Communication and Information Theory Aspects of Modern Optics*," General Electric Co., Electronics Laboratory, Syracuse, N. Y.; August, 1962.

[5] J. Tsujiuchi, "Correction of optical images by compensation of aberrations and by spatial frequency filtering," in "Progress in Optics," E. Wolf, Ed., North-Holland Publishing Co., Amsterdam, The Netherlands, vol. 2, pp. 133–180; 1963.

[6] W. M. Brown, "Analysis of Linear Time-Invariant Systems," McGraw-Hill Book Co., Inc., New York, N. Y.; 1963.

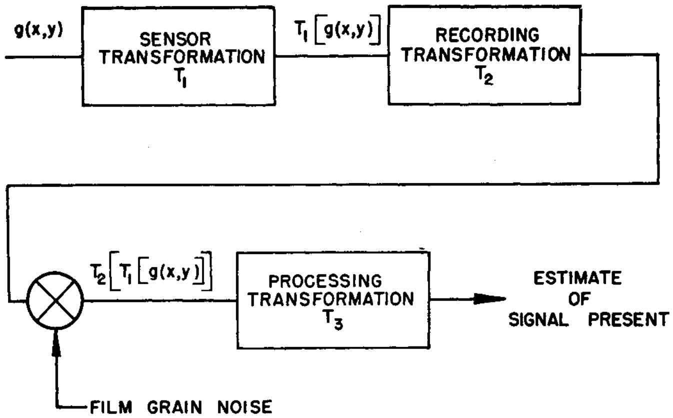


Fig. 1—Mathematical model of image processing.

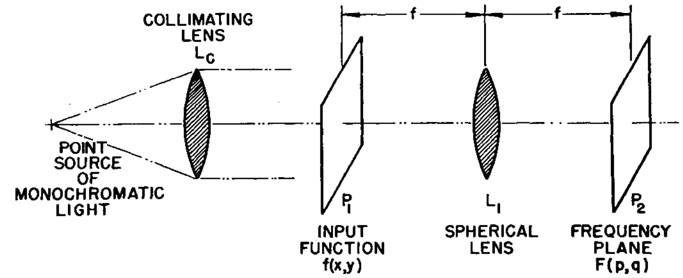


Fig. 2—Coherent Fourier analyzer.

$$H(p, q) = \frac{kS^*(p, q)}{N(p, q)} \tag{1}$$

where

- $S^*(p, q)$ is the complex conjugate of the signal spectrum,
- $N(p, q)$ is the spectral density of the background noise and
- k is a constant chosen to make $|H(p, q)| < 1$ for passive spatial filters.

Eq. (1) reduces to the so-called matched filter case when $N(p, q)$ is uniform for all (p, q), as is usual in treatments of this problem for electronic systems. Almost always, the spectral density of the noise background in photographs is far from uniform. This fact might impose a limitation on the realizability of electronic filters, but, as we shall see later, it poses no problem in the realization of the optimum filter in coherent optical systems.

The output from the linear filter represented by (1), given here for future reference, is

$$r(x, y) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} F(p, q)H(p, q)e^{j(px+qy)}\, dp\, dq. \tag{2}$$

Coherent Optical Processing Systems

Rhodes has shown that the optical system sketched in Fig. 2 functions as a two-dimensional Fourier analyzer.[7] If $f(x, y)$ denotes the specular amplitude transmission of the transparency in plane P_1, and $F(p, q)$ denotes the complex distribution of light in plane P_2, then

$$F(p, q) = \iint_{-\infty}^{\infty} f(x, y) \exp\,[j(px + qy)]\, dx\, dy. \tag{3}$$

In (3), p and q represent spatial frequency variables having the dimensions of radians/unit distance. The variables in plane P_2, which are in units of distance, are related to the frequency variables by

$$\xi = \frac{\lambda f}{2\pi} p \tag{4a}$$

$$\eta = \frac{\lambda f}{2\pi} q \tag{4b}$$

where

- ξ is the direction parallel to x,
- η is the direction parallel to y,
- λ is the wavelength of the illumination and
- f is the focal length of the spherical lens.

When a distribution of light is a function of frequency, the variables (p, q) will be used to emphasize that fact, as well as to simplify the notation associated with Fourier transform theory.

Since a spherical lens can take the Fourier transform of a complex distribution of light, one can construct an optical system by arranging a sequence of lenses which forms a succession of Fourier transform planes. An image of the input plane can be effected by placing a lens behind plane P_2 (see Fig. 3) which takes the transform of $F(p, q)$. A positive spherical lens always introduces a positive kernel in the transform relationship, hence the distribution in plane P_3 is

$$r(x, y) = \frac{1}{4\pi^2} \iint_{-\infty}^{\infty} F(p, q)e^{j(px+qy)}\, dp\, dq = f(-x, -y). \tag{5}$$

We have assumed that the system has unity magnification and sufficient bandwidth to pass the highest spatial frequency of the input function. Note that the output is an inverted image of the input function, which is what one expects from an imaging system operating under any type of illumination.

It is convenient to realize separately the numerator and denominator of the optimum filter. If two transparencies are placed in contact, their complex transmissions

[7] J. E. Rhodes, Jr., "Analysis and synthesis of optical images," *Am. J. of Phys.*, vol. 21, pp. 337–343; May, 1953.

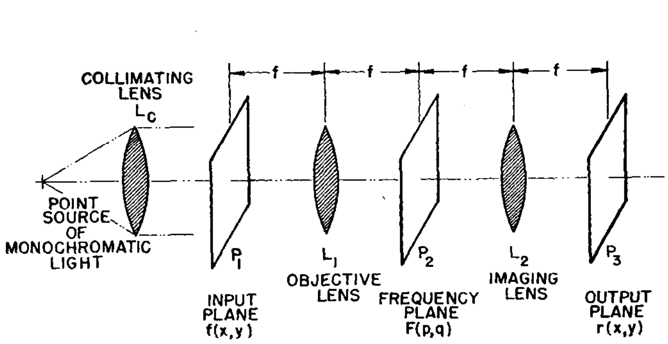


Fig. 3—Coherent optical processing system.

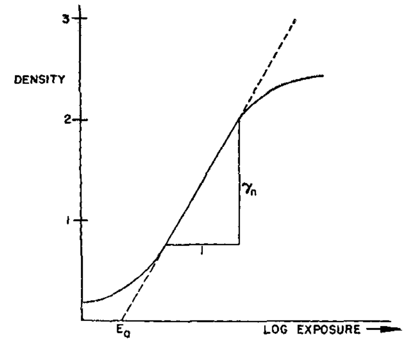


Fig. 4—Typical transfer curve for photographic film.

are multiplicative.[8] Consequently, to realize the filter described by (1), we need to insert in plane P_2 a transparency whose transmission is $1/N(p, q)$, as well as a transparency representing $S^*(p, q)$. When this filter is used, the output of the system essentially represents the probability that a signal has occurred. A bright spot in the output indicates a high probability, and low-light levels indicate a low one. (The output may not resemble the signal to any extent because the phase content of $S(p, q)$ is modified.) This process simultaneously detects signals with similar orientations regardless of their position in the input, as a result of the fact that the spectrum of a translated signal is modified by a linear phase factor. The phase factor contains precisely the information required to image the signal at the proper position in the output. In contrast, the system is sensitive to the orientation of the signals, but a rotation of the filter relative to the input provides a sequential search through all possible orientations.

Realization of the Optimum Filter

Having determined that the optimum filter which maximizes the ratio of peak signal energy to mean square energy is given by (1), we must find some method to realize $H(p, q)$. The denominator of $H(p, q)$ is always a non-negative function which can be realized on photographic film. The first step in this realization is a brief review of the transfer characteristics of films. Fig. 4 shows a typical curve of density vs log exposure, characterized in its linear region by

$$D_n = \gamma_n[\log E_n - \log E_0] \tag{6}$$

where

γ_n is the slope of the straight line,
E_n is the exposure,
E_0 is the intercept of the straight line,
D_n is the density and
n means that a negative transparency is used.

The coherent system operates on the specular amplitude transmission of the film, but using the relationship that $T = \exp[-D/2]$ and substituting (6), we have

$$T = CE_n^{-\gamma_n/2}. \tag{7}$$

If we make E_n proportional to $N(p, q)$ and require that $\gamma_n = 2$, we have realized the denominator of $H(p, q)$.

Since the numerator of (1) may represent a complex quantity, a different technique must be used to realize $S^*(p, q)$. One problem in realizing the numerator of the optimum filter is to determine both the amplitude and the phase of

$$S(p, q) = \iint_{-\infty}^{\infty} s(x, y) \exp[j(px + qy)]\, dx\, dy. \tag{8}$$

One cannot merely use a spherical lens, as previously described, to take the Fourier transform of $s(x, y)$, because any physical detector measures only the intensity of $S(p, q)$, not its phase. A Mach-Zehnder interferometer, however, can be used to determine the phase in a distribution of light by combining that distribution with a reference wave whose amplitude and phase distributions are known. This interferometer is modified for our purposes as shown in Fig. 5.

The signal $s(x, y)$ whose Fourier transform is to be found, and a spherical lens, are inserted in one beam of the interferometer. The lens displays the Fourier transform of $s(x, y)$ at its back focal plane outside the interferometer. If we neglect aberrations in the interferometer, the observed output in the back focal plane of the lens is

$$G(p, q) = |R(p, q) + S(p, q)|^2 = |R(p, q)|^2 + |S(p, q)|^2 + R^*(p, q)S(p, q) + R(p, q)S^*(p, q) \tag{9}$$

where $R(p, q) = |R(p, q)| \exp[j\phi(p, q)]$ represents the light coming from the reference beam, and $S(p, q) = |S(p, q)| \exp[j\theta(p, q)]$ is the signal spectrum. We can rewrite (9) as

$$G(p, q) = |R(p, q)|^2 + |S(p, q)|^2 + 2Re[R(p, q)S^*(p, q)] = |R(p, q)|^2 + |S(p, q)|^2 + 2\,|R(p, q)|\,|S(p, q)| \cos[\phi(p, q) - \theta(p, q)]. \tag{10}$$

[8] Unless otherwise noted, "transmission" will herein signify the specular amplitude transmission of the transparency.

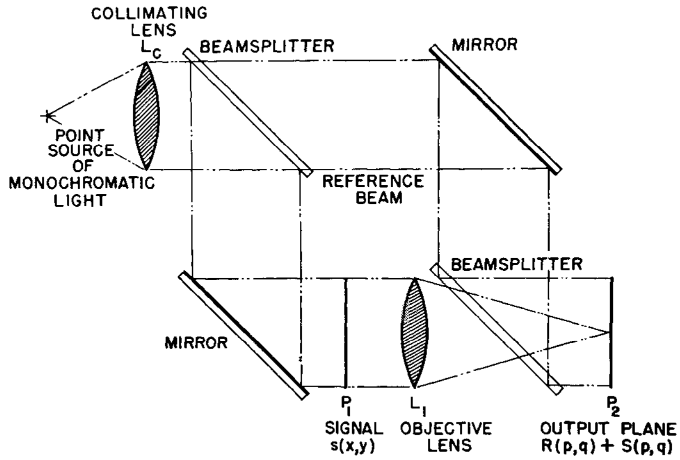


Fig. 5—Modified mach-zehnder interferometer.

Both the amplitude and phase of $R(p, q)$ could be adjusted to determine $\theta(p, q)$, but since the phase information of $S^*(p, q)$ is contained in $G(p, q)$, a non-negative function, $G(p, q)$ can be recorded on film to realize the optimum filter.

The film is exposed so that its transmission is proportional to $G(p, q)$. If we combine this film with the film on which $1/N(p, q)$ is realized, the transmission of the combination is

$$\frac{G(p, q)}{N(p, q)} = \frac{|R(p, q)|^2 + |S(p, q)|^2}{N(p, q)} + \frac{R^*(p, q)S(p, q)}{N(p, q)} + \frac{R(p, q)S^*(p, q)}{N(p, q)}. \tag{11}$$

Suppose we require $|R(p, q)|$ to be a constant, k, and $\phi(p, q)$ to be linear in (p, q), *i.e.*, $\phi(p, q) = bp + cq$. Then

$$\frac{G(p, q)}{N(p, q)} = A(p, q) + H^*(p, q)e^{-j(bp+cq)} + H(p, q)e^{j(bp+cq)} \tag{12}$$

where

$$A(p, q) = \frac{|R(p, q)|^2 + |S(p, q)|^2}{N(p, q)}$$

$$H(p, q) = \frac{kS^*(p, q)}{N(p, q)},$$

b, c are constants.

Thus, the third term of (12) represents the desired filter function, multiplied by a linear phase factor. The problem is to separate this term from the other two terms of (12). This can be accomplished by inserting the filter represented by (12) into the optical system (Fig. 3) at plane P_2. Lens L_2 separates the three terms in the output by taking the inverse transform of the light distribution in P_2; *i.e.*,

$$\text{output} = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} F(p, q)\frac{G(p, q)}{N(p, q)}e^{j(px+qy)}\,dp\,dq. \tag{13}$$

Substituting (12) in (13), we have

$$\text{output} = \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} F(p, q)A(p, q)e^{j(px+qy)}\,dp\,dq + \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} F(p, q)H^*(p, q)e^{j[(x-b)p+(y-c)q]}\,dp\,dq + \frac{1}{4\pi^2}\iint_{-\infty}^{\infty} F(p, q)H(p, q)e^{j[(x+b)p+(y+c)q]}\,dp\,dq. \tag{14}$$

The first term of (14), which appears on the optical axis, is of no particular interest in this discussion. Nor is the second term, which appears at $x = b$, $y = c$. The third term, $r(x + b, y + c)$, where $r(x, y)$ is defined by (2), is exactly the output expected from an optimum filter. This term appears with its center off the optical axis by an amount $x = -b$, $y = -c$. At this point it will be convenient to set $c = 0$ since it is arbitrary. To prevent overlap of the three outputs, the value of b must be such that $b \geq B$, where B is the length of the signal in the x direction. The fact that this output occurs off axis by a distance $x = -b$ is not important, since the output sensor can be located properly to detect it.[9]

In passing, a comment will be made on the significance of the second term of (14). If $N(p, q) = N$ for all (p, q),

[9] This method is the equivalent of recording a complex valued function as a real valued function on a carrier frequency.[6]

we can write the third term of (14) as a cross-correlation integral (for $c = 0$);

$$r(x + b, y) = \iint_{-\infty}^{\infty} f(u, v)s^*(x + b + u, y + v)\, du\, dv. \quad (15)$$

The second term of (14) can then be written as a convolution integral;

$$r(x - b, y) = \iint_{-\infty}^{\infty} f(u, y)s(x - b - u, y - v)\, du\, dv. \quad (16)$$

Thus, when $N(p, q)$ is uniform, the cross correlation and convolution of the signal with the input function are both displayed in the output plane.

Experimental Results

A few experimental results will serve to illustrate the theory and indicate the potential of spatial filtering when complex filters can be realized.

Detection of Simple Geometrical Shapes

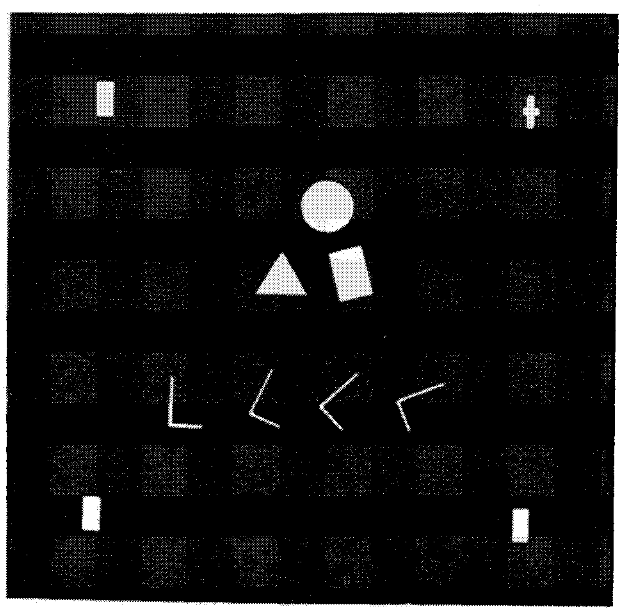
The first example presented is the detection of one of the elementary geometrical shapes shown in Fig. 6(a). Any one of the shapes could be selected; we chose the small rectangle. We realized the filter for this signal (which is real because of symmetry), by the method described. The output of interest is shown in Fig. 6(b). Note the simultaneous detection of all signals with the proper shape and orientation.

The second signal selected was the "L" shape, which has a complex spectrum. The Fourier transform of the complex filter, shown in Fig. 7, illustrates the fidelity with which the filter was realized. The light distribution in the center of the output is the transform of the first two terms of (9). The "L" on the left is the transform of the third term of (9). The other "L" is the transform of the last term. These two images are inverted and reversed relative to each other, which graphically demonstrates that $\mathfrak{F}[S(p, q)] = s(x, y)$ and $\mathfrak{F}[S^*(p, q)] = s^*(-x, -y)$. Of course, since $s(x, y)$ is real, $s^*(-x, -y) = s(-x, -y)$.

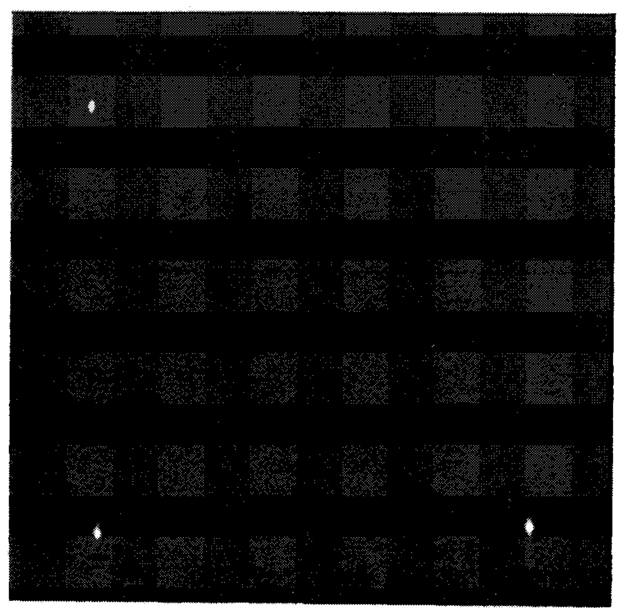
Fig. 8(a) shows the output, which is the cross correlation of the "L" with the input function. Note the symmetry in the output which is a necessary feature of cross correlation. (The other "L" 's in the input, having different orientations, do not give as large an output, but a rotation of the filter relative to the input would sequentially detect them.) Fig. 8(b) shows the output which is the convolution of the signal with the input function. It is asymmetrical, as a consequence of convolution.

Detection of Alphanumerics

The second example is the detection of alphanumerics. An interesting variation of the first example is to record the alphabet shown in Fig. 9(a) via its complex spectrum as the filter function. It is a simple matter to select any one of the alphanumerics as the signal to be detected and use it as the input signal. The output, when the letter "g" is the input, is shown in Fig. 9(b). Since the filter (which is the normal input function in disguise) does not have to be changed while the search is being carried out, this technique suggests a method for scanning a printed page for the presence of any particular alphanumeric.



(a)



(b)

Fig. 6—Detection of geometric shapes. (a) Geometric shapes. (b) Detection of rectangles.

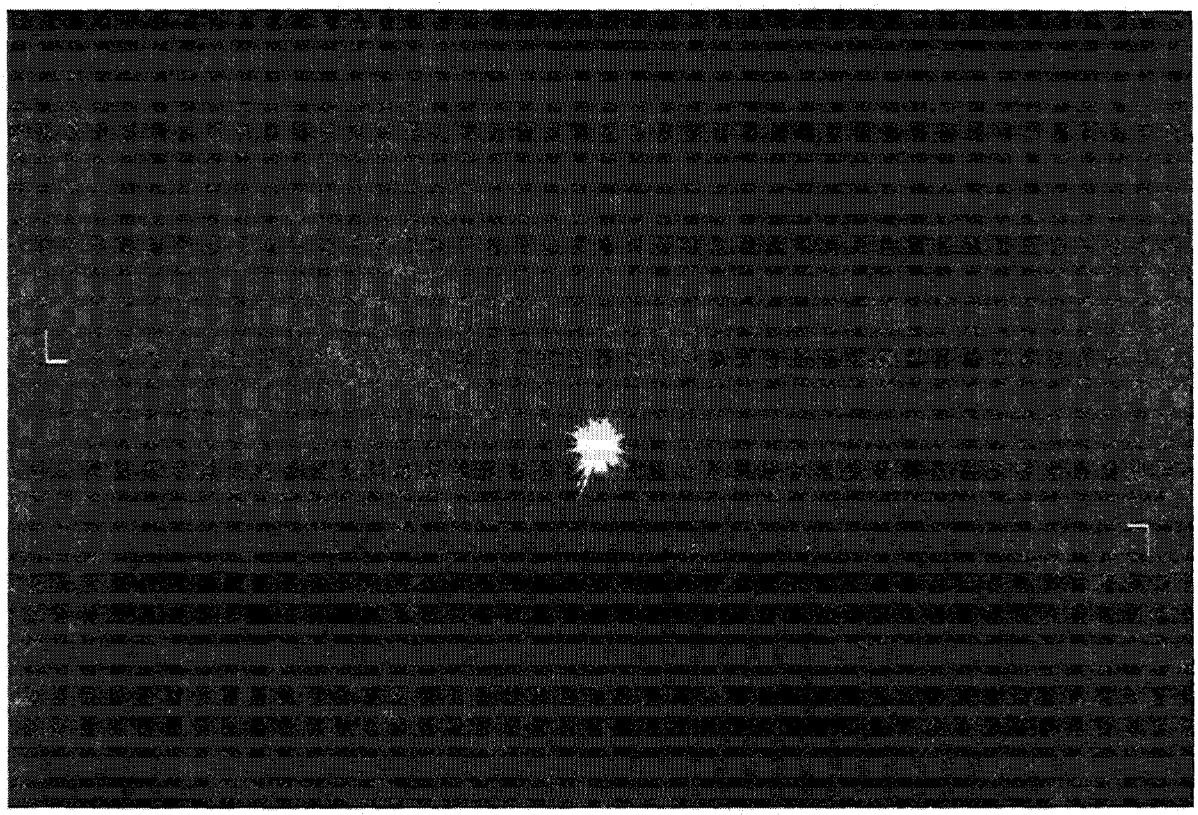
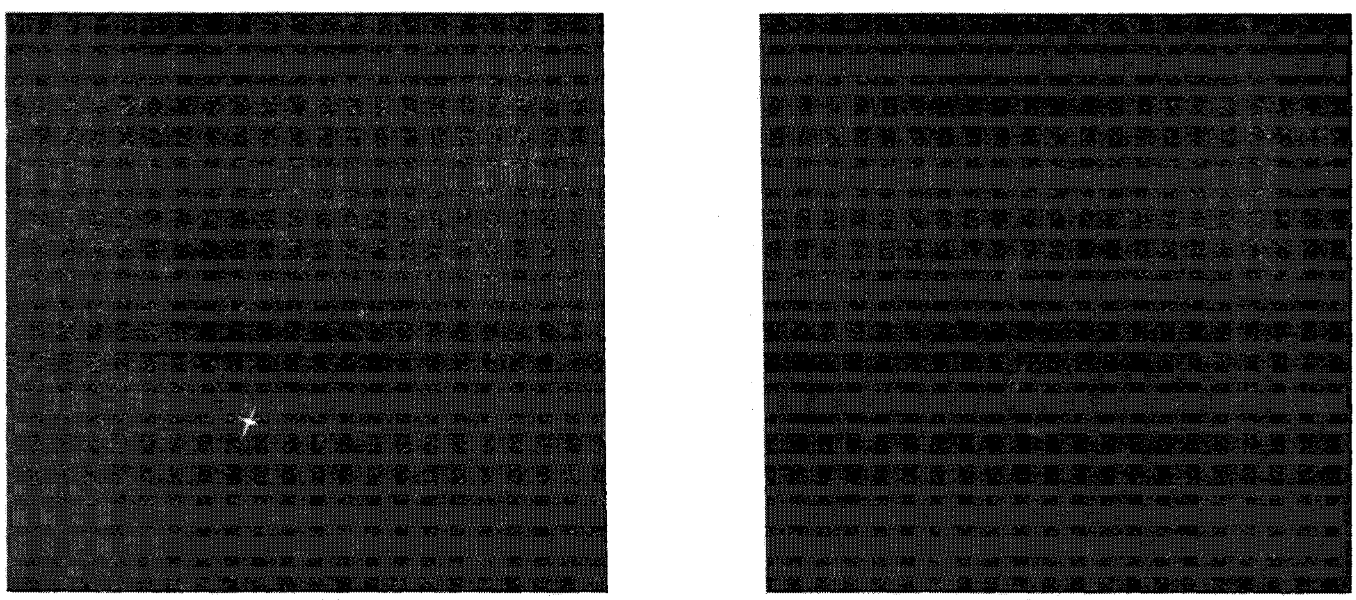
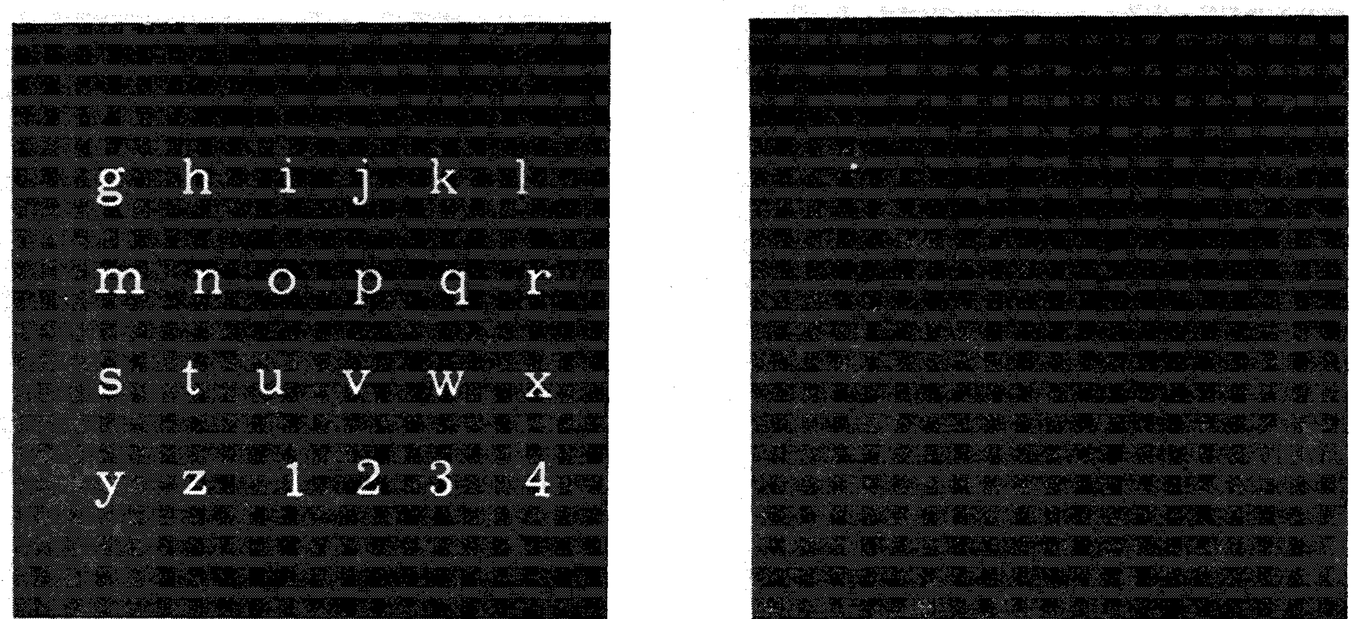


Fig. 7—Fourier transform of complex spatial filter.



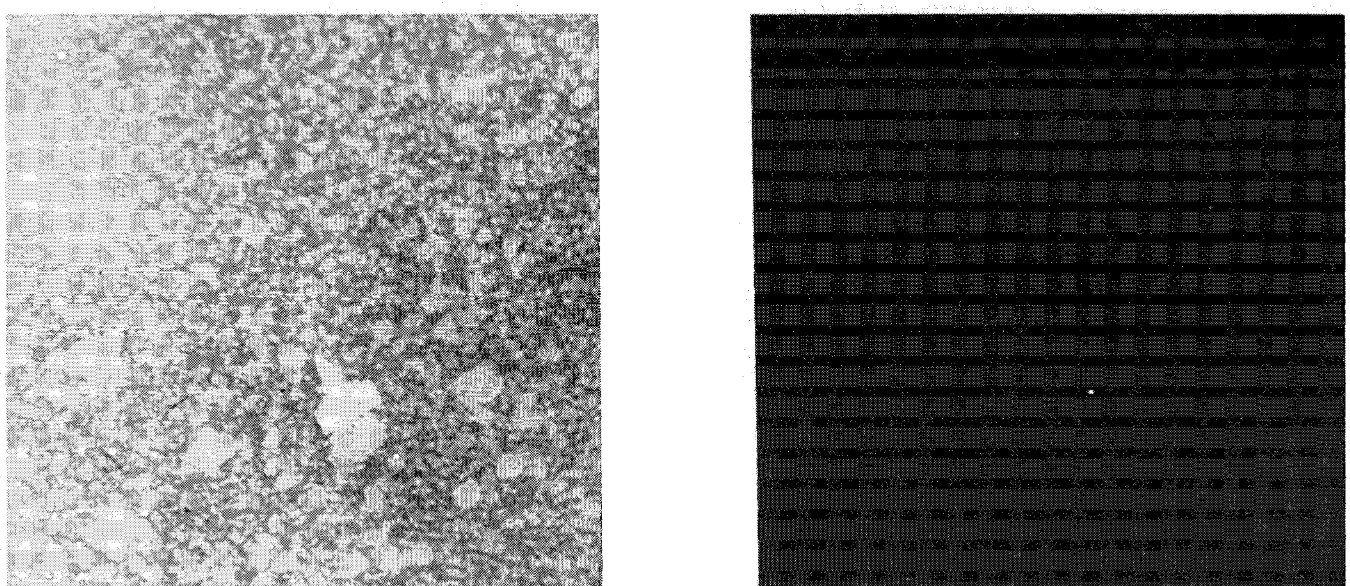
(a) (b)

Fig. 8—Detection of signal having complex spectrum. (a) Cross-correlation of "L" with geometric shapes. (b) Convolution of "L" with geometric shapes.



(a) (b)

Fig. 9—Detection of letter. (a) Alphanumerics. (b) Detection of the letter "*g*".



(a) (b)

Fig. 10—Detection of isolated signal in random noise. (a) Input scene. (b) Detection of signal.

Detection of Isolated Signal in Random Noise

In the preceding two examples the "noise" spectral density was uniform enough to be considered white. This final example shows the detection of a signal which is immersed in a noise background with nonuniform spectral density and demonstrates the power of using a coherent system for signal detection. Fig. 10(a) shows a signal in background noise. Since the noise spectral density is nonuniform, the denominator of the filter must be realized. Fig. 10(b) shows that the background noise has been completely suppressed and the signal has been detected.

Acknowledgment

The author wishes to thank A. Klooster, Jr., who carried out most of the experimental work. Helpful conversations with staff members, particularly E. N. Leith, are also acknowledged.