Convolution-kernel-based optimal trade-off filters for optical pattern recognition

A. Grunnet-Jepsen, S. Tonda, and V. Laude

An architecture for the implementation of optical pattern recognition is proposed that makes use of convolution-kernel-based optimal trade-off filters to allow for an increased speed of operation and filter storage capability. The derivation of these new convolution-kernel-based optimal trade-off filters is presented, and their noise robustness and discrimination capabilities are discussed.

Key words: Correlators, correlation filters, optimal trade-off filters.

In recent years there has been a resurgence of interest in optical pattern recognition implemented by both coherent and incoherent correlators. The impetus has been that optical correlation is particularly well suited to the tasks of locating, identifying, and tracking objects in a large field of view that contains both noise and clutter, tasks at which the correlators excel even in comparison with the latest specialized electronic signal processors or supercomputers (see, e.g., Ref. 1). One area in particular that has seen important developments is the design of correlation filters. From early on it was clear that the classical spatially matched filter performed suboptimally when it came to coping with object discrimination, input-image distortions, and image deterioration caused by noise. As a result, numerous alternative filter designs have been proposed.[2,3] When a filter is designed for optical correlation systems, three separate criteria are often considered[4]: the noise robustness, the sharpness of the correlation peak, and the optical efficiency. By the adoption of a method of multicriteria optimization, it has been possible to obtain filters that represent the optimal trade-off (OT) between these criteria.[5–7]

In this paper we concern ourselves only with optical correlators that make use of direct-space correlation filters. In contrast to Fourier filters, the direct-space filters can be displayed by an amplitude-modulating spatial light modulator (SLM). This display significantly reduces the demands on the dynamic range and coding domain of the SLM, a demand that has plagued the realization of efficient Fourier correlation filters.[8]

The working principle behind most optical correlators is that an image of a scene is acquired, after which the image is correlated with numerous precalculated filters representing separate reference objects or training images that are sufficiently descriptive of object distortions such as changes in size, scale, or viewing angle. It has recently been noted, however, that for the optimum performance to be obtained in terms of adaptive discrimination in pattern recognition and tolerance to variations of the objects, it is necessary for nonlinear filtering techniques to be employed.[7–10] Unfortunately, the cost of this improved performance is that the filters must incorporate the spectral density of the image scene. Although the nonlinear joint-transform correlator inherently accounts for the power spectrum of the input image, it does not allow for adjustment of the trade-off between the discrimination capabilities and the noise robustness. Because this freedom may be desirable in many practical applications, we concentrate here on such correlators as the linear joint-transform correlator[11] and the shadow-casting correlator.[12] Thus, in general, the incorporation of the spectral density of the image scene means that it becomes necessary to recalculate all the filters prior to each correlation, a prohibitively expensive computational task that would seem to negate the inherent advantages of using optical instead of electronic correlators.

The alternative to repeated recalculation of the filters would be to precalculate a series of filters that estimate the possible variation in the spectral densi-

The authors are with the Laboratoire Central de Recherches, Domaine de Corbeville, Thomson-CSF, F-91404 Orsay, France.

Received 16 October 1995; revised manuscript received 5 February 1996.
0003-6935/96/203874-06$10.00/0

ties of the image scene. It would then be possible from a single measurement of the spectral density of the image to recall the best approximation filters from memory. In practice, this is an impractical approach because a single filter of size 1024×1024 pixels with 256 gray levels requires (uncompressed) 1 Mbyte of memory, making the storage of a large number of such filters untenable. It is with this problem in mind that we here suggest an alternative architecture for the implementation of optical pattern recognition that solves the problem of storage space without the need for a direct and computationally expensive recalculation of the filters. Our compromise solution is outlined in Fig. 1 and consists of the calculation and storage of a small kernel, say of size 5×5, which when convolved with the original reference object yields an approximation to an OT filter. In this way we reduce the problem of storing numerous large filters to the problem of storing a series of small kernels, from which the approximate OT filters may then be calculated markedly faster than by the complete recalculation of the filters. In this paper we develop a theoretical solution to the calculation of these convolution-kernel-based OT filters and discuss their performance in comparison with conventional OT filters.

By way of introduction we start by recalling briefly the theory underlying the derivation of OT filters, as presented originally in Ref. 5. The main principle consists of identifying analytical expressions that describe certain aspects of filter performance. Thus,

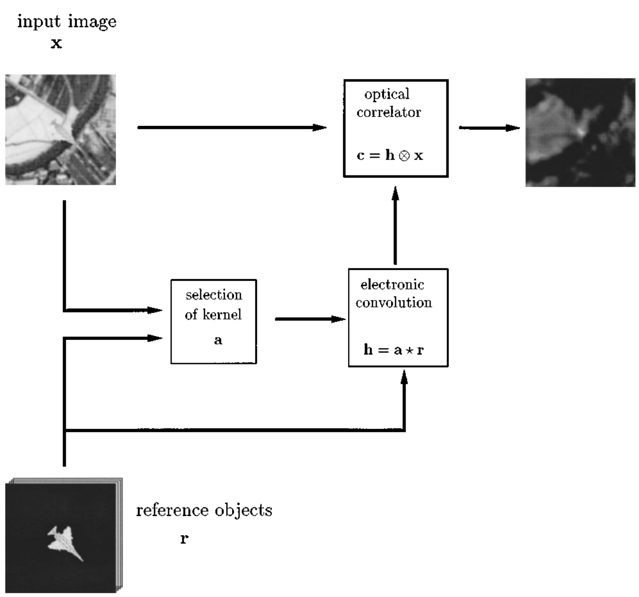

Fig. 1. Schematic of the working principle behind a correlation system that uses convolution-kernel-based OT filters. The power spectrum of the input image, x, is used to select from a set of precalculated kernels, a, those elements that best account for the input image noise spectrum. The convolution-kernel-based OT filters, h, are subsequently calculated electronically when the small kernels are convolved with the reference objects. The input image can now be correlated with the series of convolution-kernel-based OT filters that use an optical correlator.

a good measure of the noise robustness of a filter is the signal-to-noise ratio (SNR), defined as[5]

$$\mathrm{SNR} = \frac{\eta}{\mathrm{MSE}}, \qquad \eta = |h^{\dagger} r|^2, \qquad \mathrm{MSE} = h^{\dagger}\mathbf{S}h, \quad (1)$$

where η is the optical efficiency defined as the modulus squared of the central value of the correlation between the filter h and the reference r, the superscript dagger denotes the complex-conjugate transposition, and MSE denotes the mean-square error or output variance of the correlation peak in the presence of a noise with a covariance matrix $\mathbf{S}$. Note that for clarity all images are represented as one-dimensional vectors.

Another important characteristic in optical correlation is the sharpness of the correlation peak. This may be quantified in terms of the peak-to-correlation-plane energy (PCE), defined as[4,7]

$$\mathrm{PCE} = \frac{\eta}{\mathrm{CPE}}, \qquad \mathrm{CPE} = h^{\dagger}\mathbf{D}h, \quad (2)$$

where CPE is the correlation-plane energy and $\mathbf{D}$ is the covariance matrix of the reference image.

Having defined these three criteria (SNR, PCE, and η) as important for optical correlation, we must now find the class of filters that represents the optimal trade-off between the criteria. We may accomplish this by adopting a Lagrange multiplier method and by simplifying the problem to one of finding the filter that minimizes the following cost function[8]:

$$E(h) = h^{\dagger}\mathbf{B}_{\mu}h - 2\lambda\eta(h), \qquad \mathbf{B}_{\mu} = (1-\mu)\mathbf{S} + \mu\mathbf{D}, \quad (3)$$

where $\mu \in [0, 1]$ and $\lambda \geq 0$. This function leads to the following simple solution[5]:

$$\hat{h}_k^{\mathrm{OT}} = \frac{\hat{r}_k}{[\hat{B}_{\mu}]_k} = \frac{\hat{r}_k}{(1-\mu)\hat{S}_k + \mu\hat{D}_k}, \quad (4)$$

where the caret denotes the Fourier transform of a variable. Note that $\hat{\mathbf{S}}$ and $\hat{\mathbf{D}}$ ($\hat{D}_k = |\hat{r}_k|^2$) are the power spectrums of the noise and the reference images, respectively. In the limits of no trade-off, Eq. (4) reduces, for $\mu = 0$, to the well-known matched filter ($\hat{h}_k = \hat{r}_k/\hat{S}_k$) and, for $\mu = 1$, to the inverse filter ($\hat{h}_k = \hat{r}_k/|\hat{r}_k|^2$). One aspect of OT filters that is worth noting is that they are real valued in direct space. We use this fact to good effect in the following derivation of a convolution-kernel-based OT filter.

Having outlined the theory of OT filters, we now digress slightly to clarify what we mean by the noise contained in Eq. (4). A more rigorous derivation of optimal filters[9] has shown that, although OT filters are optimal *linear* filters, their performance may be improved by the use of nonlinear filtering techniques. More specifically, it has been shown that for the discrimination capabilities and the noise robustness of a filter to be optimized, the power spectrum of the noise must be equal to that of the input-image background, where the background is defined as everything that is not the object of search.[9,10] However, because we assume that there is no *a priori* knowledge of where and how many objects are contained in the image scene, it has been suggested that the best alternative is to estimate the spectral density of the background from the whole image. In other words, for (sub)optimal performance one may, for example, set $\hat{S}_k = |\hat{x}_k|^2$, see Ref. 9, or use other estimations suggested in Ref. 9 or 10. The drawback to nonlinear filtering is clearly that it requires that the filter be recalculated prior to each correlation.

Proceeding now to the derivation of the convolution-kernel-based OT filter, we were inspired by the close visual resemblance between the filter and the reference object in direct space to attempt to write the filter as a convolution of the reference object r and a small kernel, a:

$$h_m = \sum_{p \in V} a_p r_{p+m}, \tag{5}$$

where elements a_p are real numbers. With this equation we have imposed the constraint that the value of a point in a filter h_m should depend only on the weighted average of the neighboring points r_m located within a proximity region V. This region may, for example, be a grid of size 5×5 pixels. Thus, the objective now becomes to deduce the values of the weight elements a that best approximate the OT filters. We may accomplish this in a straightforward manner by inserting Eq. (5) into Eqs. (3). After a few simple algebraic manipulations this leads to

$$E(a) = \sum_p \sum_q a_p a_q f(p - q) - 2\lambda \sum_p a_p g(p), \tag{6}$$

where

$$f = \mathrm{IFT}[\hat{D}\hat{B}_\mu], \qquad g = \mathrm{IFT}[\hat{D}]. \tag{7}$$

Here IFT denotes taking the inverse Fourier transform. Thus $g(0)$ is, for example, the central autocorrelation peak of the reference. Minimizing Eq. (6) with respect to ensemble a yields the optimal solution:

$$a = (\mathbf{G}^{+}\mathbf{F}^{-1}\mathbf{G})^{-1}\mathbf{F}^{-1}\mathbf{G}, \tag{8}$$

where the superscript (-1) denotes inversion. $\mathbf{F}$ and $\mathbf{G}$ are, respectively, the matrix and vector representations of the above functions f and g. Furthermore, the scalar λ has been identified by the requirement that $a^{\dagger}\mathbf{G} = 1$, yielding $\lambda = (\mathbf{G}^{+}\mathbf{F}^{-1}\mathbf{G})^{-1}$. It is worthwhile to note that the analytical solution of Eq. (8) is similar in form to the synthetic discriminant-function solutions derived, for example, in Ref. 13.

To examine the performance of the convolution-kernel-based OT filter, we calculate the PCE versus the SNR curves for filters derived from the trial reference object depicted in Fig. 2(a). We have opted to normalize the PCE and SNR to their maximum possible values attained for the matched and inverse filters, respectively, to better illustrate the filter

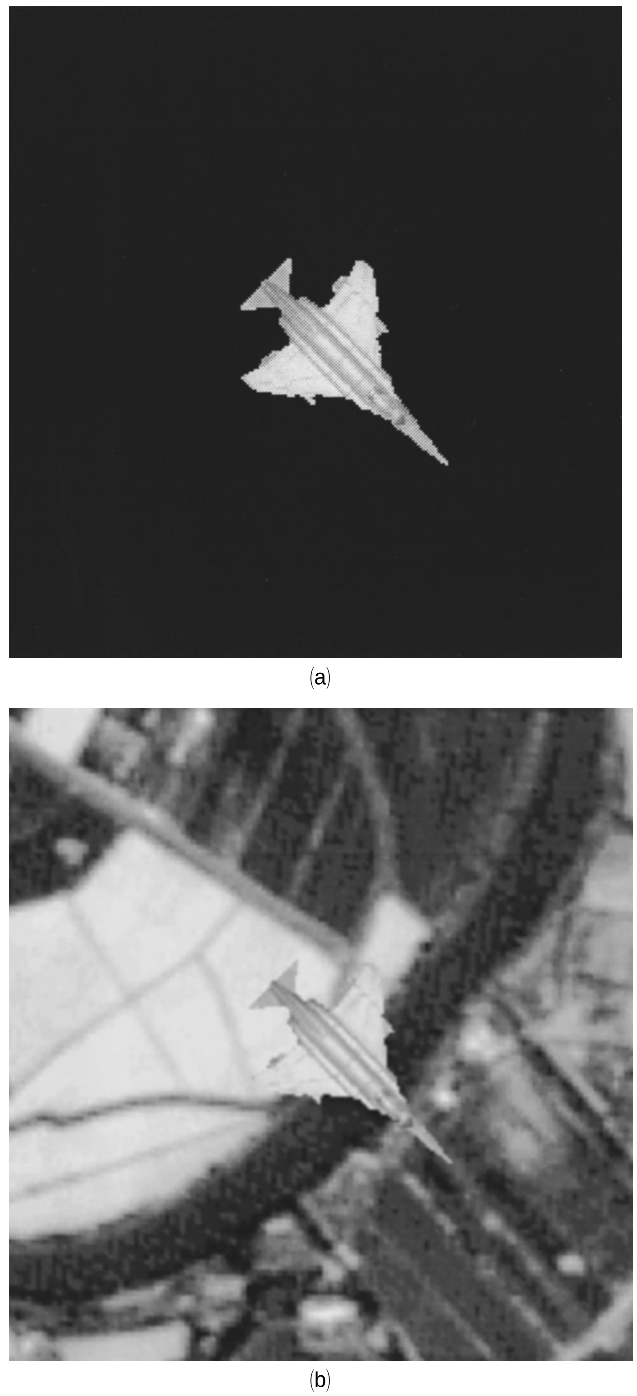

Fig. 2. (a) Trial reference object, i.e., a 256 gray-scale 256×256 image of an airplane. (b) Input image scene, consisting of the object of search (i.e., the airplane) and a complex background.

performance. To start with, we investigate in Fig. 3(a) a case in which the noise is modeled as being white Gaussian. This case is representative of the wideband thermal noise that is inevitably encountered on acquiring a picture electronically. Comparing the optimal characteristics curve (OCC) of the OT filter with those of the convolution-kernel-based filter for kernel sizes of 3×1, 3×3, 5×5, and 7×7 pixels, we clearly see that the convolution-kernel-based filters rapidly approach the limiting performance of the OT–OCC as the size of the kernel is increased. In fact, already on going from 5×5 to 7×7 pixels the return appears almost negligible. We note furthermore that a convolution-kernel-based OT filter derived from a 5×5 kernel represents a marked improvement in object discrimination over that of the matched filter (the lowest point on the OT–OCC curve). At this stage it is worth noting that the calculation of a from Eq. (8) can be sensitive to numerical errors, so it is important to use at least double precision.

In most practical situations the noise is of course not merely white Gaussian but should in addition comprise the spectral density of the input image, as

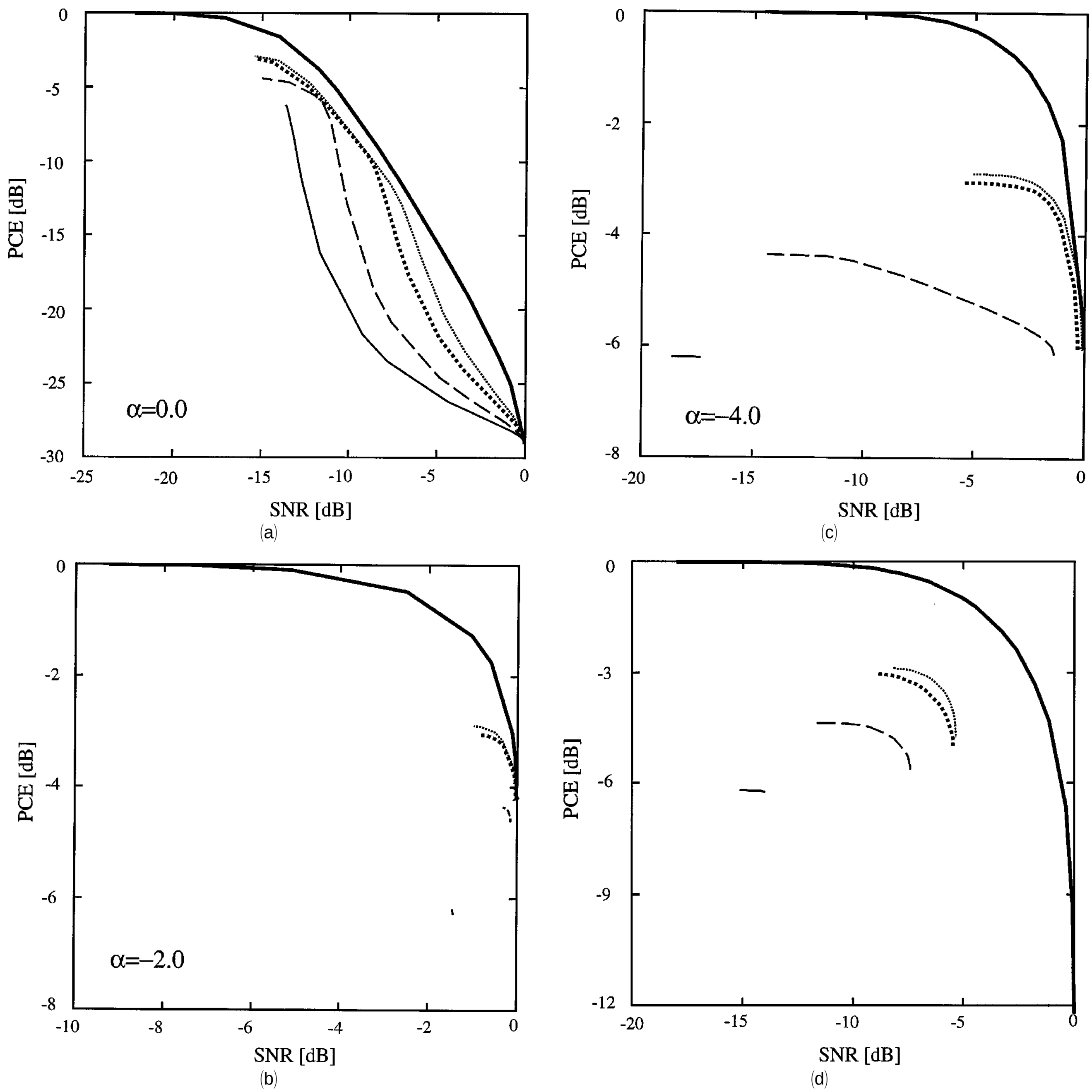

Fig. 3. Normalized PCE as a function of SNR, showing the OCC's of the OT filter (thick solid curve) and of the convolution-kernel-based filter for kernel sizes of 3×1 (thin solid curve), 3×3 (dashed curve), 5×5 (large dotted curve), and 7×7 (small dotted curve). The noise power spectrum is (a) white, (b) $1/f^2$, (c) $1/f^4$, and (d) calculated from the input image in Fig. 2(b).

mentioned above. One option could then be to attempt to estimate this noise power spectrum. Experience has shown that colored noise of the form f^{α} may often be used as a good model, where f is the spectral frequency of the image. Thus, for instance, a spectral density of $1/f^2$ is a relatively good approximation of an airplane against a clear sky, while $1/f^4$ describes quite well the case of an airplane above a cloudy background.[7] In Figs. 3(b) and 3(c) we compare the performances of the filters for these two functions. Again the filter calculated from a 5×5 pixel kernel is seen to be a relatively good approximation of the OT filter. However, we note that in Fig. 3(b) the dynamic range of all the curves is quite small. This is because the spectral density of the noise is quite close to that of the reference object. In fact, when they are identical one cannot talk about a trade-off, because the adjustment of the trade-off parameter μ in Eq. (4) would have no effect.

For situations in which the background is highly structured and in which the power spectrum of the input-image scene changes in an unpredictable manner, it becomes necessary to calculate the power spectrum directly. This would, for example, apply to the case of an airplane flying across a structured landscape, as shown in Fig. 2(b). The comparison of the resulting filters, as illustrated in Fig. 3(d), shows that the convolution-kernel-based filters have a smaller range of trade-off values. However, we should gauge their real performance by comparing them with the classical matched filter ($\hat{h}_k = \hat{r}_k$), which for this case has a SNR and PCE of approximately -51 dB and -29 dB, respectively, showing that the convolution-kernel-based filters offer a significant improvement and are comparatively close to the optimum performance. We should add that when calculating the power spectrum from Fig. 2(b) we have included a small contribution of white noise for two reasons: it represents a good model of the additional recording noise, and it serves to stabilize the filter.[7]

Before continuing let us state that the filters described by Eqs. (4), (5), and (8) are valued from $[-1, 1]$ and therefore cannot be realized directly in incoherent correlators for which the filter may have only positive real values. In effect, it means that the correlation has to be performed twice, once with a filter composed of the positive values, and once with a filter of negative values. We then obtain the final correlation by taking the square of the difference of the individual correlations. This so-called bipolar approach yields exactly the same result as a single correlation with the original filter (see, e.g., Ref. 12).

Having seen how convolution-kernel-based OT filters may offer good performance in terms of noise robustness and object discrimination, let us now turn to the advantages of using them in real optical correlation systems. A $p \times q$ pixel kernel of a convolution-kernel-based filter consists of only $(pq + 1)/2$ independent elements because of the central symmetry of a. This effectively eliminates any problems associated with storing a large number of trade-off filters. However, before the correlation is possible, it is necessary to convolve the reference object with the convolution-kernel-based OT filter kernel. To carry out a rough estimation of how this approach compares with the direct recalculation of the OT filter, we note that an OT filter needs approximately $nm(1 + 3 \log_2 nm)$ operations to perform a fast Fourier transform, filter calculation, and inverse fast Fourier transform for a $n \times m$ pixel filter. In contrast, a convolution-kernel-based OT filter needs roughly an $nmpq$ operation to perform the convolution of a $p \times q$ pixel kernel with an $n \times m$ reference image. For the example from Fig. 1 this translates to a speed increase of a factor 8 for the 5×5 pixel kernel. In practice, however, the increase may be many times higher because the convolution of the reference with such a small kernel can be performed by a specialized electronic processing card when the filter SLM is directly addressed.

In summary, we have presented the derivation of convolution-kernel-based OT filters and have suggested how their use in an optical correlation architecture may solve many of the practical problems that have arisen in terms of speed of correlation and storage of filters. The primary advantage of the proposed solution is that it makes it possible to approximate the performances of both linear and nonlinear correlation filters while much less storage space is used and a shorter recalculation time for the filter is required.

The authors gratefully acknowledge fruitful discussions with P. Chavel. The idea for convolution-kernel-based OT filters was suggested to the authors by P. Réfrégier. The research was supported by the European Communities Human Capital Mobility Program.

References

1. C. H. Huang, G. Gheen, and E. Washwell, "Throughput comparison of optical and digital correlators for automatic target recognition," in *Optical Information Processing Systems and Architectures III,* B. Javidi, ed., Proc. SPIE **1564,** 427–438 (1991).
2. Y. N. Hsu and H. H. Arsenault, "Optical pattern recognition using circular harmonic expansion," Appl. Opt. **21,** 4016–4019 (1982).
3. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23,** 812–816 (1984).
4. B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," Appl. Opt. **29,** 2997–3006 (1990).
5. P. Réfrégier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency," Opt. Lett. **16,** 829–831 (1991).
6. B. V. K. Vijaya Kumar, C. Hendrix, and D. W. Carlson,

"Tradeoffs in the design of correlation filters," in *Optical Pattern Recognition,* J. L. Horner and B. Javidi, eds., Vol. CR40 of SPIE Critical Review Series (Society of Photo-Optical Instrumentation Engineers, Bellingham, Wash., 1992), pp. 191–215.
7. P. Réfrégier and V. Laude, "Critical analysis of filtering techniques for optical pattern recognition: Are the solutions of this inverse problem stable?," in *Euro-American Workshop on Optical Pattern Recognition,* B. Javidi and P. Réfrégier, eds., Proc. SPIE **PM12,** 58–84 (1994).
8. V. Laude and P. Réfrégier, "Multicriteria characterization of optimal Fourier spatial light modulator filters," Appl. Opt. **33,** 4465–4471 (1994).
9. P. Réfrégier, V. Laude, and B. Javidi, "Basic properties of nonlinear global filtering techniques and optimal discriminant solutions," Appl. Opt. **34,** 3915–3923 (1995).
10. L. P. Yaroslavsky, "Optical correlators with $(-k)$th-law nonlinearity: optimal and suboptimal solutions," Appl. Opt. **34,** 3924–3932 (1995).
11. C. S. Weaver and J. W. Goodman, "Technique for optically convolving two functions," Appl. Opt. **5,** 1248–1249 (1966).
12. V. Laude, J. P. Huignard, and P. Réfrégier, "Optical implementation of optimal trade-off bipolar filters for shadow-casting incoherent correlator," in *Optical Implementation of Information Processing,* B. Javidi and J. L. Horner, eds., Proc. SPIE **2565,** 85–95 (1995).
13. D. Casasent, "Unified synthetic discriminant function computation formulation," Appl. Opt. **23,** 1620–1627 (1984).