AD-A258 691

||||||||||||||||||||||||||||

RL-TR-92-208
Final Technical Report
July 1992

ADVANCED CONCEPTS IN DISTORTION-INVARIANT PHASE-ONLY FILTER DESIGN

Carnegie Mellon University

B. V. K. Vijaya Kumar

DTIC
ELECTE
NOV 2 4 1992
S
A
D

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

92-30079

**Rome Laboratory
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700**

92 1: 25 025

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-92-208 has been reviewed and is approved for publication.

APPROVED: *Joseph L. Horner*

JOSEPH L. HORNER
Project Engineer

FOR THE COMMANDER: *Harold Roth*

HAROLD ROTH, Director
Solid State Sciences
Electromagnetics & Reliability Directorate

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL(EROP) Hanscom AFB MA 01731-5000. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE July 1992	3. REPORT TYPE AND DATES COVERED Final May 89 - Feb 92

4. TITLE AND SUBTITLE ADVANCED CONCEPTS IN DISTORTION-INVARIANT PHASE-ONLY FILTER DESIGN	5. FUNDING NUMBERS C - F19628-89-K-0032 PE - 61102F PR - 2305 TA - J7 WU - 36
6. AUTHOR(S) B.V.K. Vijaya Kumar	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Carnegie Mellon University Dept of Electrical & Computer Engineering 5000 Forbes Avenue Pittsburgh PA 15213	8. PERFORMING ORGANIZATION REPORT NUMBER N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Rome Laboratory (EROP) Hanscom AFB MA 01731-5000	10. SPONSORING/MONITORING AGENCY REPORT NUMBER RL-TR-92-208

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Joseph L. Horner/EROP/(617) 377-3841

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
The classical matched filter is analytically optimal for detecting a target in white additive noise, when SNR is defined in the classical way. However, the complex transmittance values pose problems in their practical realization.
The phase-only filter solves this problem but generally adds some noise to the correlation signal. This report studies ways of optimizing the performance of the phase-only filter. This includes the presence of non-white or colored noise in the input scene and detector noise. Formulas are derived and the results tested using actual images of real military armaments.

14. SUBJECT TERMS Optical computing correlator phase-only filters Optical signal processor pattern recognitions	15. NUMBER OF PAGES 130
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT SAR

ABSTRACT

This report summarizes the results obtained during the Contract No. F 19628-89-k-0032 entitled "Advanced concepts in distortion-invariant phase-only filter design." This research focused on improving and evaluating the techniques for designing phase-only and binary phase-only filters.

This report summarizes our contributions in the following areas.

- Fast algorithms for designing phase-only and binary phase-only filters.

- Designing phase-only and binary phase-only synthetic discriminant function filters.

- Characterizing and improving the correlation peak sharpness of various filters.

- Use of 4 phases such as in complex ternary matched filters.

- Trading off various performance measures such as the signal-to-noise ratio and correlation peak sharpness.

- Effects of input noise that is colored.

- Detector noise effects.

Accesion For		
NTIS CRA&I		☑
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distibution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

1. INTRODUCTION

1.1 Motivation

Phase-only filters (POFs) and binary phase-only filters (BPOFs) are extremely attractive in the sense that they provide high Horner efficiencies and are more convenient for implementation on available spatial light modulators (SLMs). However, several important issues must be addressed in designing POFs and BPOFs. These issues include: noise sensitivity of the filters, distortion tolerance, efficient algorithms for filter design, resulting correlation peak sharpness, effect of detector noise and trade-offs between various performance measures. We have addressed many of these issues during this research effort and made many significant contributions. This report summarizes these results.

1.2 Research Contributions

Following important results were established as part of this research effort. More details can be found in subsequent chapters and appendices.

- We have introduced new, efficient algorithms for designing phase-only filters (POFs) and binary phase-only filters (BPOFs).

- We have advanced new algorithms for designing phase-only and binary phase-only distortion-invariant synthetic discriminant function (SDF) filter algorithms.

- We introduced a new performance measure called peak-to-correlation energy (PCE) to characterize the sharpness of correlation peaks.

- We derived algorithms to design POFs, BPOFs and complex ternary matched filters (CTMFs) to maximize PCE, the peak sharpness measure.

- We proposed fractional power filters (FPFs) to illustrate the trade-offs among various performance measures in correlation filter design.

- We extended earlier work of optimum-SNR filter design to include colored input noise.

- We included the role of detector noise in filter design and showed that it leads to filter saturation.

- We proposed multi-criteria optimal binary amplitude phase-only filters capable of providing optimal trade-offs among various measures.

1.3 Overview of the Report

The rest of this report is organized as follows. Chapter 2 provides a brief background on classical matched filters, phase-only filters and binary phase-only filters. Chapter 3 summarizes the results of the fast, efficient algorithms we derived for the design of POFs and BPOFs. Chapter 4 introduces the concept of complex ternary matched filters. Chapter 5 discusses the results of extending phase-only and binary phase-only concepts to synthetic discriminant functions. The effects of colored input noise as well as detector noise are considered in Chapter 6. The new peak sharpness measure and its role in filter design are discussed in Chapter 7. An optimum method for trading off these measures is discussed in Chapter 8. The last chapter provides our conclusions.

This report also includes various appendices which provide more details since each chapter provides only the basic results.

2. BACKGROUND

2.1 Classical Matched Filters

Optical correlator has been the mainstay of optical pattern recognition for many years. In the classical matched filter (CMF), the filter function $H(u,v)$ placed in the frequency plane of a VanderLugt correlator[1] is given by

$$H(u,v) = S^*(u,v) \qquad (2.1)$$

where $S(u,v)$ is the two-dimensional Fourier transform of the reference image $s(x,y)$ and the superscript asterisk indicates the complex conjugate. It is well known[2] that the CMF yields the maximum output signal-to-noise ratio (SNR). The output SNR is defined as

$$SNR \overset{\Delta}{=} \frac{|E\{ c(0,0)\}|^2}{\text{Var}\{ c(0,0)\}} , \qquad (2.2)$$

where $E\{\cdot\}$ and Var $\{\cdot\}$ denote the expected value and variance and $c(0,0)$ denotes the correlation output at the origin. The underlying assumption is that the correlation peak is at the origin.

The classical matched filter has three significant drawbacks.

- The light throughput efficiency[3] (known as the Horner efficiency) is usually small for CMFs since the filter transmittance is less than one at many spatial frequencies.

- Since $H(u,v)$ in eq. (2.1) is complex-valued, it cannot be easily implemented on available devices such as magneto-optic spatial light modulator.[4]

- The CMF is extremely sensitive to distortions in the input scene.

Several remedies have been proposed to alleviate these problems. These include: phase-only filters[5] (POFs) to improve Horner efficiency, binary phase-only filters[6] (BPOFs) to allow use of binary spatial light modulators and synthetic discriminant functions[7] (SDFs) to improve distortion-invariance.

2.2 Phase-only Filters

To overcome the light efficiency problems of CMFs, Horner and Gianino[5] suggested the use of POF.

$$H_{POF}(u,v) = S^*(u,v)/|S(u,v)| \ . \tag{2.3}$$

Here the filter magnitude is 1 at all spatial frequencies and thus the light throughput efficiency is very high. Designing the filter based on only phase (and not the magnitude) of the Fourier transform (FT) of the input is justified on the basis that earlier experiments by Oppenheim and Lim[8] showed that the FT phase appeared to be more important than the FT magnitude in image reconstruction.

One of the problems with the filter in eq. (2.3) is that it is all-pass, i.e., it has unit transmittance at all frequencies. Thus there is no mechanism to control how much of the input noise gets through the filter. In a previous effort, Vijaya Kumar and Bahri[9] suggested the use of a region of support R along with the POF. The set R contains the spatial frequencies that are not set to zero. The combination of the POF with R is termed as the optimal phase-only filter (OPOF) and is given below.

$$H_{OPOF}(u,v) = \begin{cases} H_{POF}(u,v) & \text{for } (u,v) \in R \\ 0 & \text{for } (u,v) \notin R \end{cases} \ . \tag{2.4}$$

Vijaya Kumar and Bahri[9] suggested an algorithm for determining R so that the resulting output SNR is maximized. This is based on the following relationship between SNR and the region of support R.

$$
\begin{aligned}
SNR\ (R) \quad &= \quad \frac{\left| \int \int H(u,v)\ S(u,v)\ dudv \right|^2}{\int \int P_n(u,v)\ |H(u,v)|^2\ dudv} \\[2em]
&= \quad \frac{\left[\int \int_R |S(u,v)|\ dudv \right]^2}{\int \int_R P_n(u,v)\ dudv} \qquad (2.5)
\end{aligned}
$$

where $P_n(u,v)$ is the spectral density of the input noise. When the input noise is white, $P_n(u,v) = N_0$ and the output SNR can be approximated as follows.

$$
\begin{aligned}
SNR\ (R) \quad &= \quad \frac{\left[\int \int_R |S(u,v)|\ dudv \right]^2}{N_0 \int \int_R du\ dv} \\[2em]
&= \quad \frac{1}{N_0 A_R} \left[\int \int_R |S(u,v)|\ dudv \right]^2 \\[1.5em]
&\simeq \quad \frac{1}{N_0 A_R} \left[\sum_{k \in R} |S_k| \right]^2 \qquad (2.6)
\end{aligned}
$$

where A_R is the area of region R and we have used discretization for approximation. It is easy to see from eq. (2.6) that for a fixed A_R, we must include the A_R frequencies with the highest $|S_k|$ values in region R. The optimum value for A_R can be determined only by an exhaustive search.

2.3 Binary Phase-Only Filters

While POFs improve Horner efficiency, they do not lend themselves for implementation on some currently available spatial light modulators such as the magneto-optic SLMs (MOSLMs). For example, MOSLMs can accommodate two phase levels (usually 0 and π radians) and perhaps two magnitude levels (0 and 1). Several methods have been suggested for determining a binary phase-only filter (BPOF) for implementation on such devices. These can be summarized as follows.

$$H_{BPOF}(u,v) = \text{Sgn} \{ Re [S(u,v) e^{j\beta}] \} \qquad (2.7)$$

where $0 \leq \beta \leq \pi/2$ is known as the threshold line angle, $Re[\cdot]$ denotes the real part and $\text{Sgn} \{\cdot\}$ is defined as follows.

$$\text{Sgn} \{x\} = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases} . \qquad (2.8)$$

Several special cases of the BPOF in eq. (2.7) are of interest. When $\beta = 0$, the BPOF is obtained by binarizing the real part of $S(u,v)$ and when $\beta = \pi/2$, the BPOF is obtained by binarizing the imaginary part. When $\beta = \pi/4$, we binarize the sum of the real part and the imaginary part of $S(u,v)$. While the BPOF in eq. (2.7) is certainly convenient for implementation on an MOSLM, it is also an all-pass filter and thus has no noise rejection capability.

Vijaya Kumar and Bahri[10] suggested the use of a region of support R along with the conventional BPOFs. This leads to a filter with three transmittance levels (-1, 0 and +1) rather than just two levels. Kast et al.[11] have recently shown that MOSLMs can be made to operate in a 3-level mode. Filters using these 3 levels of transmittance are also

known as ternary-phase amplitude filters[12] (TPAFs).

Vijaya Kumar and Bahri[10] suggested an algorithm for determining R to maximize the output SNR when the input noise is white. The algorithm is similar to the one discussed in Section 2.2 except for one difference. The difference is that β, the threshold line angle is needed in BPOF design. There does not appear to be a closed form expression for this angle. However, in most simulation studies, $\beta = 0$ (real-part binarization) appears to yield the best output SNR.

2.4 Synthetic Discriminant Functions

The CMF outputs are very sensitive to input image distortions. To improve this sensitivity, Hester and Casasent[7] suggested the use of synthetic discriminant functions (SDFs). In SDFs, the filter is constructed from a *linear* combination of a set of training images. The weights in this linear combination are chosen so that the correlation outputs take on prespecified values (e.g., large values when the input is from a desired class and small values when it is not). To quantify this, let $s_1(x,y), \cdots, s_N(x,y)$ denote the N training images. Then the filter $H(u,v)$ is given as $S^*(u,v)$ where the composite image $s(x,y)$ is given as the linear combination of the N training images.

$$s(x,y) \;=\; \sum_{i=1}^{N} a_i\, s_i\,(x,y) \qquad\qquad (2.9)$$

where the coefficients $\{a_1, a_2, \ldots, a_N\}$ are chosen to satisfy the following constraints on the correlation outputs.

$$\int\int s(x,y)\, s_i(x,y)\, dx dy \;=\; c_i, \qquad i=1, 2, \cdots, N \qquad (2.10)$$

where c_i values are prespecified. By substituting eq. (2.9) in (2.10), we get N linear equations in N unknown. We can thus find $s(x,y)$ and the SDF filter $H(u,v)$.

Several variants of SDFs have been proposed over the last 10 years. However, most of these yield *fully-complex* filter functions and suffer from light efficiency and implementation problems. Recently, Jared and Ennis[13] proposed a relaxation algorithm for obtaining phase-only and binary phase-only SDFs. In this report, we report a new method that we call the successive forcing algorithm.

3. EFFICIENT ALGORITHMS FOR POF AND BPOF DESIGN

3.1 Basic Idea

We discussed in Section 2.2 an algorithm for determining the optimal region of support for POF design and in Section 2.3 an algorithm for determining the optimal region of support for the BPOF design. Those algorithms are already efficient. For example, when dealing with a 32×32 filter array with 1024 pixels, there are a total of $2^{1024} \approx 10^{300}$ possible choices for R. Out of these, one R is the best. In Section 2.2, we stated that when the input noise is white, the optimal R^* (i.e., the region of support leading to maximum SNR) must be of the following form.

$$R^* = \{(u,v) : |S(u,v)| > T\} \tag{3.1}$$

where T is an unknown threshold. This optimal threshold T needs to be determined by an exhaustive search.

Since a 32×32 filter array has 1024 values in it, we need at most 1024 thresholds. Then R^* can be determined by trying out all 1024 thresholds, computing the resulting SNRs and then choosing the threshold that leads to the highest SNR. Thus we need to test only 1024 choices for R instead of 10^{300} choices. We have made additional improvements to speed up the algorithms.

We have proved[14] that if two or more frequencies of the Fourier transforms of the image have equal magnitudes, either all of them or none of them are included in the region of support for the optimal POF. A similar theorem can be stated for BPOFs also. Proofs for this can be found in our *Applied Optics* reprint included as Appendix A of this report.

This theorem provides an even more efficient way of finding the optimal R^* for POFs and BPOFs. For example, suppose we quantize the $|S(u,v)|$ values to 4 bits. That means that there are only 16 distinct magnitudes in $|S(u,v)|$ after quantization. From the above theorem only 16 thresholds need to be verified. This is much faster than the original algorithm. Of course, the quantization in $|S(u,v)|$ leads to sub-optimal filters. We suspect that this suboptimality will be insignificant unless the quantization is very coarse. An efficient algorithm based on this quantization idea is presented in Appendix A.

3.2 Numerical Results

To understand the effectiveness of this algorithm, we tested it on the 32×32 pixel gray-level tank image shown in Fig. 3.1. To compute the filters, we zero-padded this image and performed 64×64 sized discrete Fourier transforms using the FFT algorithm.
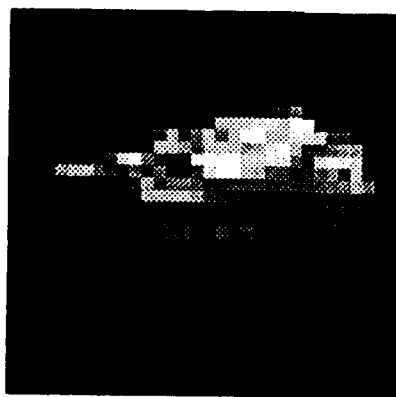

Fig. 3.1. The 32×32 tank image used in the numerical experiments.

First, we compared the efficiencies of the original algorithm and this new algorithm for designing the optimal R^* for POFs. Using the most efficient sorting algorithm, the original algorithm took 10.15 seconds of CPU time on our Vax 11/750 whereas the new algorithm (using 8 bit quantization) took only 0.13 seconds on the same machine. This is a speed-up by a factor of 78. When the FFT size was increased to 128×128, the speed-up factor improved to 130 for the same 8-bit quantization. The price paid for this speed improvement is a small decrease in the output SNR. The old algorithm yielded an output SNR of 18.834585 dB whereas the new one yielded 18.834409 dB. This difference is insignificant.

When the new algorithm was applied to find the optimal regions of support for BPOFs, the CPU times decreased from 1585 seconds (for all 19 ROSs for the 19 choices of β) to 16.8 seconds for a speed-up factor of 95. When FFT sizes were increased to 128×128, this reduction factor went to 980. The SNR loss due to quantization was no worse than 0.00046 dB. These numerical results conclusively prove the superiority of this algorithm.

4. COMPLEX TERNARY MATCHED FILTERS

4.1 Motivation

The CMF's retain full complex information about the image Fourier transforms whereas the POF's and BPOF's discard the magnitude information. In fact, BPOF's retain only one bit of information (-1 or +1) at each spatial frequency. It is important to find out if we can improve the performance of the filters by allowing more bits of representation in the filter domain.

In a previous effort, Vijaya Kumar and Connelly[15] analyzed the effect on the output SNR of quantizing the phase of a POF to N levels. They showed that when $N = 2$ (as in BPOF) the SNR decreased by about 4 dB compared to the POF. On the other hand, for $N \geq 4$, this decrease in SNR ≤ 1 dB. Thus using 4 phase-levels seems like an interesting approach.

In the spirit of using 4 phase levels, Dickey and Hansche[16] suggested quad phase-only filters (QPOF's) which can be implemented using 2 MOSLMs arranged in a detour-phase arrangement. In QPOFs, both devices are used in a BPOF mode and thus no noise control is provided. Complex ternary matched filters[17] (CTMFs) are designed to provide this noise tolerance.

4.2 Basic Formulation

The CTMF $H(u,v)$ is defined as

$$H_{CTMF}(u,v) = H_1(u,v) - j H_2(u,v), \qquad (4.1)$$

where $j = \sqrt{-1}$ and $H_1(u,v)$ and $H_2(u,v)$ are ternary filters, i.e., they can take on three levels of transmittance (-1, 0 and +1). Thus each of this can be implemented using an

MOSLM operated in a ternary mode. The possibility of zero transmittance allows us to control the noise sensitivity.

Letting R_1 and R_2 denote the regions of support for $H_1(u,v)$ and $H_2(u,v)$, respectively, we can write the following expression for the output SNR.

$$SNR_{CTMF} = \frac{\mid \int\int_{R_1} S(u,v)H_1(u,v)\, dudv - j \int\int_{R_2} S(u,v)H_2(u,v)\, dudv \mid^2}{\int\int_{R_1} P_n(u,v)dudv + \int\int_{R_2} P_n(u,v)dudv} \qquad (4.2)$$

Using simple symmetries, we can show that SNR_{CTMF} is optimized (for given R_1 and R_2) by selecting $H_1(u,v)$ and $H_2(u,v)$ as below.

$$H_1(u,v) = \begin{cases} \text{Sgn} \{ Re\ [S(u,v)e^{-j\beta}]\} & \text{for } (u,v) \in R_1 \\ 0 & \text{for } (u,v) \notin R_1 \end{cases} \qquad (4.3)$$

$$H_2(u,v) = \begin{cases} \text{Sgn} \{ Im\ [S(u,v)e^{-j\beta}]\} & \text{for } (u,v) \in R_2 \\ 0 & \text{for } (u,v) \notin R_2 \end{cases} \qquad (4.4)$$

where $Re\ [\cdot]$ and $Im\ [\cdot]$ denote the real part and the imaginary part, respectively, and β is the threshold line angle as discussed in the case of BPOFs. Substituting (4.3) and (4.4) in eq. (4.2), we obtain

$$SNR_{CTMF} = \frac{[\int\int_{R_1} |\hat{S}_R(u,v)|\, dudv + \int\int_{R_2} |\hat{S}_I(u,v)|\, dudv]^2}{\int\int_{R_1} P_n(u,v)\, dudv + \int\int_{R_2} P_n(u,v)\, dudv} \qquad (4.5)$$

where $\hat{S}_R(u,v)$ and $\hat{S}_I(u,v)$ denote the real part and the imaginary part, respectively, of

$[S(u,v) \, e^{-j\beta}]$. For the special case of white noise $P_n(u,v) = N_0$, a constant and optimal R_1 and R_2 can be selected to maximize SNR_{CTMF} by sorting $|\hat{S}_R(u,v)|$ and $|\hat{S}_I(u,v)|$ values.

4.3 Simulation Results

We carried out our computer simulations using the 32×32 tank image shown in Fig. 3.1. Once again, FFTs of size 64×64 were used to avoid getting a circular correlation instead of a linear correlation. The image energy was normalized to 1 and has 60.26% of its energy in the even-part (or equivalently in the real part of its FT) and 39.74% in the odd-part (or equivalently in the imaginary part of its FT). The efficient algorithm discussed in Section 4.2 was applied to this data. The resulting optimal regions of support are shown in Figs. 4.1 and 4.2. In both figures, white areas represent where the filter has nonzero transmittance and black areas represent spatial frequencies that are completely blocked. The center of the array denotes the (0,0) frequency. Fig. 4.1 shows the optimal region of support for $H_1(u,v)$ and Fig. 4.2 shows the optimal region of support for $H_2(u,v)$.
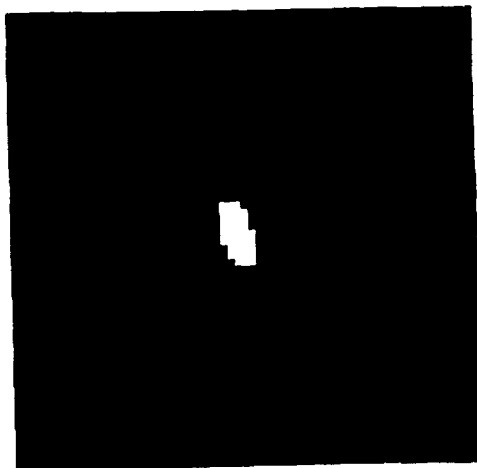
Fig. 4.1. Optimal Region of Support for $H_1(u,v)$ in CTMF.
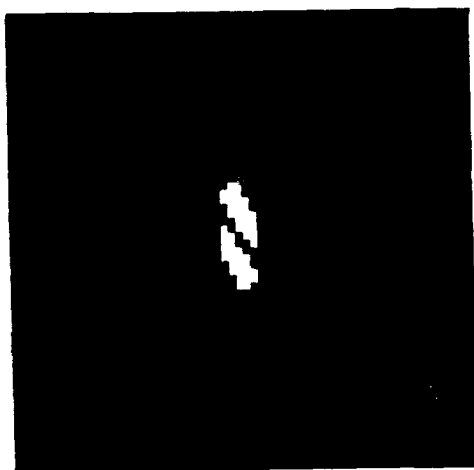

Fig. 4.2. Optimal Region of Support for $H_2(u,v)$ in CTMF.

The noise tolerance of the CTMF was tested by adding zero-mean, Gaussian noise to the input image. Resulting correlation outputs were analyzed and output SNRs were estimated. We show in Table 4.1 the output SNRs for various filters.

Table 4.1

Output SNRS (dB) for Different Filters

Input SNR (dB)	CMF	POF	OPOF	BPOF (Real)	OBPOF (Real)	QPOF	CTMF
-20	10.76	5.69	9.57	4.13	7.63	5.45	8.76
-10	20.64	14.77	19.28	12.23	17.42	14.62	18.67
0	30.60	24.73	29.02	22.07	27.24	24.59	23.63
10	40.59	34.72	38.98	32.02	37.15	34.58	38.64
20	50.58	44.72	48.92	42.00	47.27	44.57	48.63

From this table, we can draw some interesting inferences. Note that using an optimal region of support with POFs improves the output SNR by about 4.5 dB. Similarly going from BPOFs to OBPOFs improves the output SNR by about 5 dB. We show in the last column of Table 4.1 the output SNRs obtained from CTMFs. Note that they are within 2 dB of the highest possible output SNRs (i.e., those achieved by the CMF).

Finally, we show in Fig. 4.3 sample correlation outputs from the CMF, the POF and the CTMF when the input has no noise. Note that the CTMF provides a broad correlation peak. In Fig. 4.4, we show sample correlation outputs using the same filters

when the input is corrupted by additive noise. Note that the CTMF output (like CMF output) is relatively unaffected by input noise whereas the POF output is degraded by intput noise. This noise tolerance in CTMFs is due to the optimization of output SNR in designing CTMF's. More details about this can be found in the reprint in Appendix B.
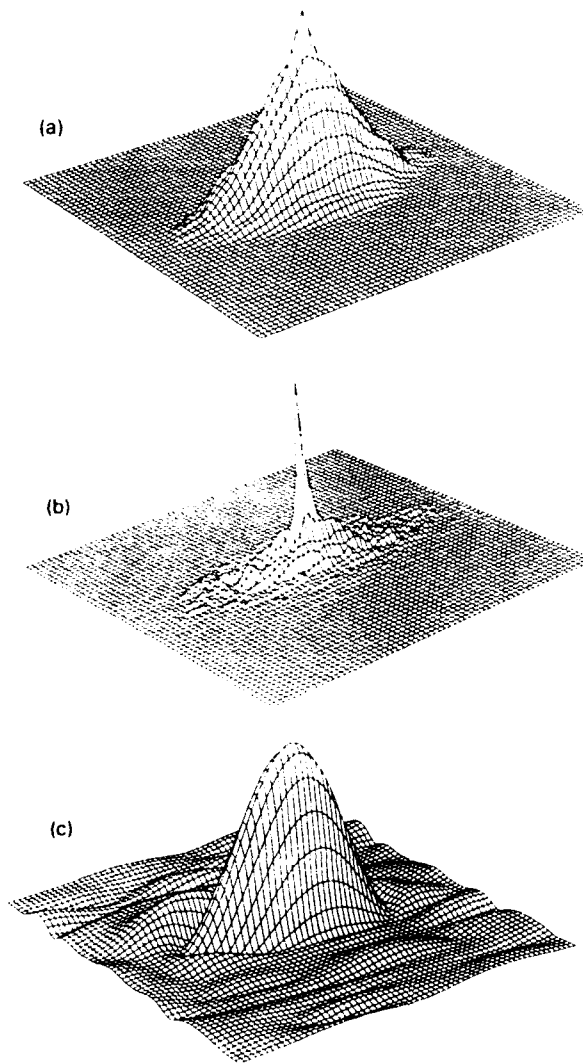
Fig. 4.3   Normalized (peak value set to 1) correlation outputs when the input image has no noise. (a) CMF output, (b) POF output and (c) CTMF output.
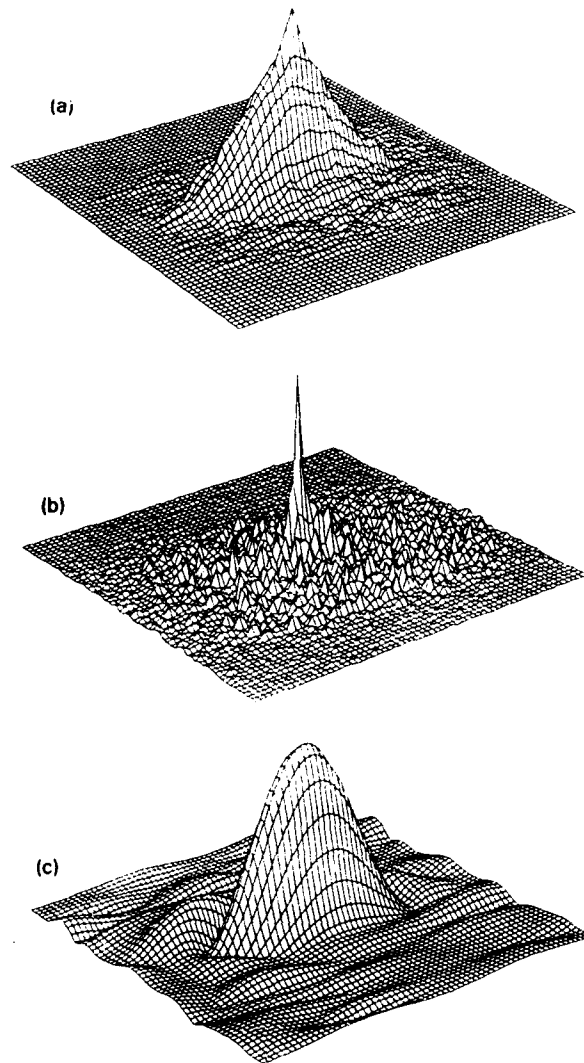
Fig. 4.4   Normalized (peak value set to 1) correlation outputs when the input image has noise such that the input SNR is 0 dB. (a) CMF output, (b) POF output and (c) CTMF output.

5. PHASE-ONLY AND BINARY PHASE-ONLY SYNTHETIC DISCRIMINANT FUNCTIONS

5.1 Motivation

As discussed in Section 2.4, SDFs provide a method to improve the distortion tolerance of the matched filters. Several variations of the basic SDF method have been suggested over the last 10 years. These include: minimum variance SDFs[18], minimum average correlation energy (MACE) SDFs[19] and Gaussian-MACE filters.[20] These and other variants of SDFs improve various attributes. However, they all lead to fully complex filters that are not amenable for implementation on devices with limited modulation.

Jared and Ennis[13] were the first to propose a method to design SDF filters for implementation on limited modulation devices. They suggested an iterative algorithm (called the relaxation algorithm) where the weights of the linear combination are adjusted to simultaneously meet the SDF constraints and to provide phase-only or binary phase-only filter function. In this research effort, we have advanced a new method that we call the successive forcing algorithm[21] (SFA) capable of designing limited-modulation SDFs.

5.2 Successive Forcing Algorithm

This algorithm is most easily explained using the vector notation. Let us discretize the N training images to get d-dimensional column vectors s_1, s_2, ..., s_N. Let X be a $d \times N$ matrix with its i-th column given by s_i. Let us denote the composite image $s(x,y)$ in eq. (2.9) by an equivalent vector s. Then

$$s = a_1 s_1 + a_2 s_2 + \cdots + a_N s_N$$
$$= X a \tag{5.1}$$

where $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T$ is the column vector of weights. The correlation output constraints in eq. (2.10) can be rewritten as

$$\mathbf{X}^T \mathbf{s} = \mathbf{c} \ , \tag{5.2}$$

where $\mathbf{c} = [c_1, c_2 \cdots, c_N]^T$ is the column vector of correlation output constraints.

The usual SDF solution is obtained by substituting eq. (5.1) into eq. (5.2) and solving for $\mathbf{s}$ as below.

$$\mathbf{s}_{SDF} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \ , \tag{5.3}$$

which can be easily seen to satisfy the SDF constraint in eq. (5.2). However, the problem with this is that it is fully complex. The successive forcing algorithm[21] (SFA) designs filters that meet the SDF constraints in eq. (5.2) while meeting the device requirements.

Algorithm:

Step 1:   Start with an arbitrary initial complex coefficient vector $\mathbf{a}$.

Step 2:   Determine the filter vector $\mathbf{s}$ from eq. (5.1).

Step 3:   Force the frequency response of $\mathbf{s}$ to meet device constraints. For example, if we want phase-only SDFs, we set the magnitude of the Fourier transform of $s(x,y)$ to one and leave the phase unchanged. Let this be denoted by $\mathbf{s}'$.

Step 4:   Determine the new constraints satisfied by $\mathbf{s}'$, i.e., $\mathbf{c}' = \mathbf{X}^T \mathbf{s}'$.

Step 5:   If the magnitudes of the components of $\mathbf{c}'$ are close to that in $\mathbf{c}$, stop and exit; otherwise, replace the magnitudes in $\mathbf{c}'$ by magnitudes in $\mathbf{c}$ and leave the phases unchanged. This leads to $\mathbf{c}''$.

Step 6:   Find the next vector $\mathbf{a}$ as $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}''$ and go back to Step 2.

There is no proof that this method will always converge. However, our numerical results are encouraging.

5.3 Numerical Results

To test the SFA, we chose a 2-class problem where one class is the tank image shown in Fig. 3.1 and the other class is the APC image in Fig. 5.1. Extensive simulations were conducted. Here we show some sample results.
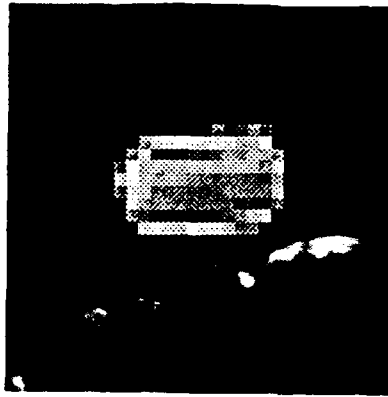

Fig. 5.1. The 32×32 APC image used for the false class.

We used 6 training images (at 60° intervals) from the true class (i.e., tanks) and 6 training images from the false class (i.e., APCs). In Fig. 5.2, we show the correlation outputs (at the origin) when the phase-only version of the conventional SDF is employed. There is very little separation between the curves for the true class and the false class. In comparison, the separation between the true class outputs and the false class outputs for the phase-only SDF designed using the SFA is much better as shown in Fig. 5.3.

The results in Figs. 5.2 and 5.3 show how the phase-only SDFs work with training images. In Fig. 5.4, we show the correlation outputs with all images when the phase-only version of the conventional SDF is employed and in Fig. 5.5 we show the correlation outputs with all images when the phase-only SDFs using SFA are employed.

Note that the number of misclassifications is less in Fig. 5.5 compared to Fig. 5.4. To improve further the performance of phase-only SDFs designed using SFA, we designed it using all 72 available training images (36 from each class). The resulting correlation outputs show significant separation between the two classes. These basic ideas can be extended to binary phase-only SDFs also.
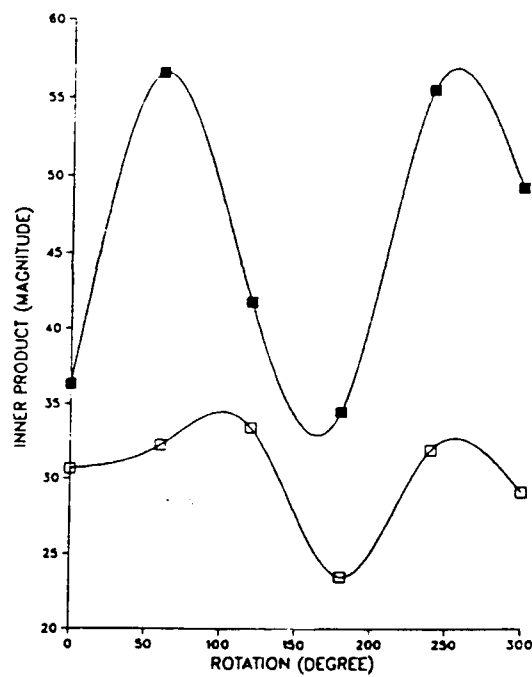
Fig. 5.2   Correlation outputs with the training images for the phase-only version of the conventional SDF.
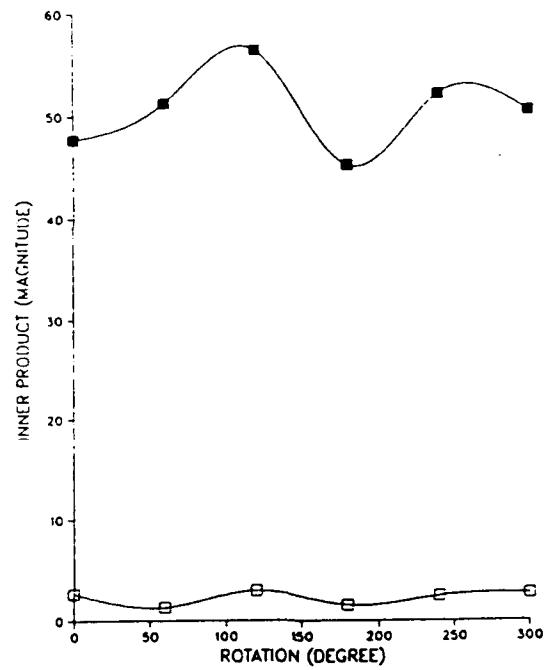
Fig. 5.3   Correlation outputs with the training images for the phase-only SDF designed using SFA.
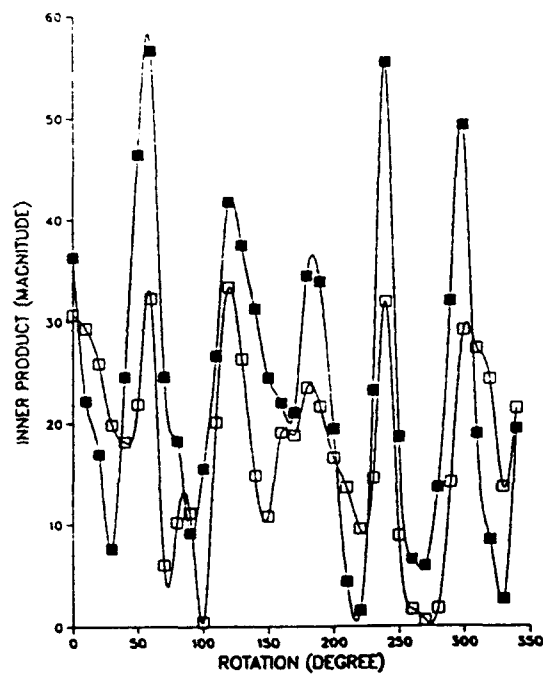
Fig. 5.4   Correlation outputs with all images for the conventional phase-only SDF designed using 12 training images.
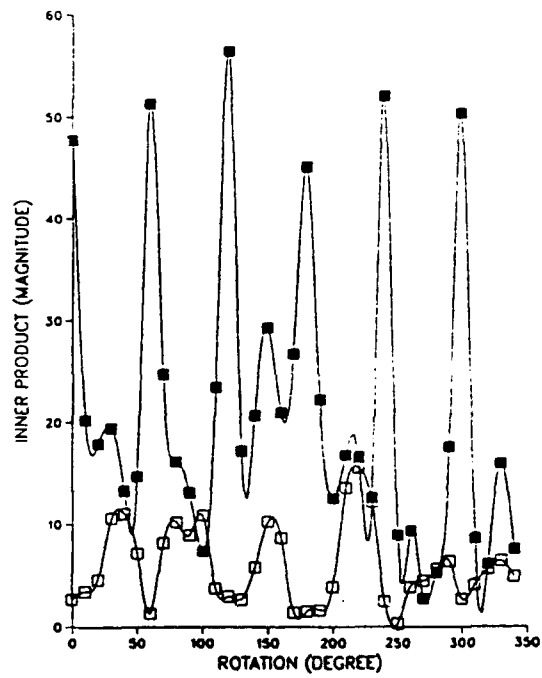
Fig. 5.5   Correlation outputs with all images for the phase-only SDF designed using SFA and 12 training images.
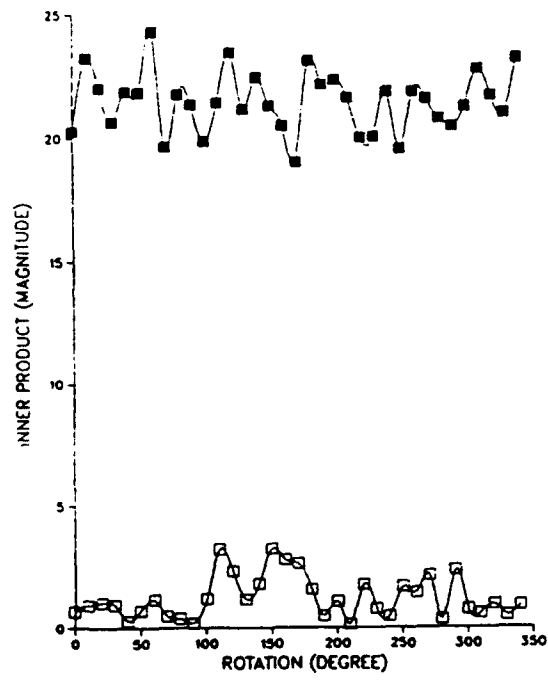
Fig. 5.6   Correlation outputs with all images for the phase-only SDF designed using SFA and 72 training images.

6. EFFECT OF COLORED NOISE AND DETECTOR NOISE

6.1 Background

Much of the previous research in correlation filter design was based on the assumption of white noise being added to the input scene. This may not be appropriate in many situations. For example, often the objects of interest may be in natural backgrounds that can be modeled as sample realizations from random noise. However the spatial spectrum is rarely constant and thus these backgrounds must be modeled as colored noise.

Another important noise source often ignored is the detector placed in the correlation plane. These detectors introduce their own noise prior to any decision scheme. If this detector noise is low or if the light coming through the filter is strong, this is not an issue. But one of the reasons for the popularity of light-efficient filters (such as POFs) is the small amounts of light available in the output plane. Thus detector noise becomes an important issue.

6.2 Colored Noise

When the input noise has a power spectral density $P_n(u,v)$, the resulting output SNR for a filter $H(u,v)$ can be written as

$$SNR = \frac{|\int\int H(u,v)\ S(u,v)\ dudv\ |^2}{\int\int P_n(u,v)\ |H(u,v)|^2\ dudv} . \tag{6.1}$$

For the case of a POF with region of support R, this SNR becomes

$$SNR_{POF} = \frac{[\int\int_R |S(u,v)| \ du \ dv]^2}{\int\int_R P_n(u,v) \ du \ dv} \ .$$ (6.2)

In general, it is not obvious how R should be chosen to maximize SNR_{POF}. We have been able to show[22] that the optimal region of support R^* must be of the following form.

$$R^* = \{ (u,v) : \frac{|S(u,v)|}{P_n(u,v)} \geq T \ \} \ ,$$ (6.3)

where T is an unknown threshold that needs to be determined. Above basic procedure can be extended to BPOFs and CTMFs also. Details are presented in the *Optics Letters* reprint attached as Appendix C.

To illustrate the advantages of using the colored noise formulation, we designed a POF with region of support optimized for a particular "colored" noise. Here the background was used to estimate the spectral density of the colored noise. Fig. 6.1 shows the reference image in zero background and Fig. 6.2 shows the same image in a natural background.
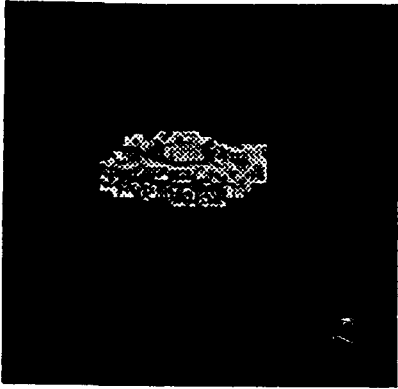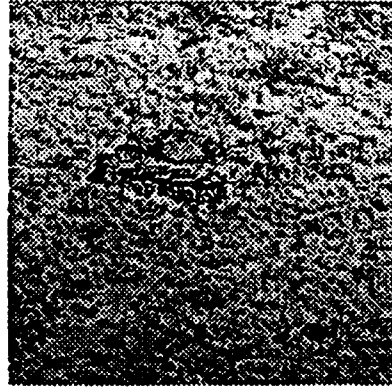
Fig. 6.1. Segmented Reference -
Image

Fig. 6.2. Reference Image in
Background

The background in Fig. 6.2 was used to estimate $P_n(u,v)$. This estimate is shown in

Fig. 6.3. The estimated $P_n(u,v)$ was used with eq. (6.3) to determine the optimal region

of support. Fig. 6.4 shows the optimal ROS under white noise assumption and Fig. 6.5

shows the same using estimated $P_n(u,v)$. Note the significant difference between the
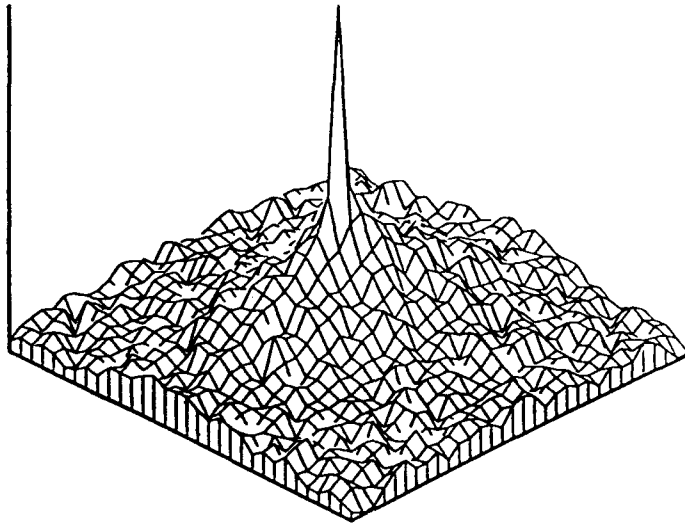
two ROSs.

Fig. 6.3. Es⁺⁺ n ted Power Spectral Density of the Background in Fig. 6.2.
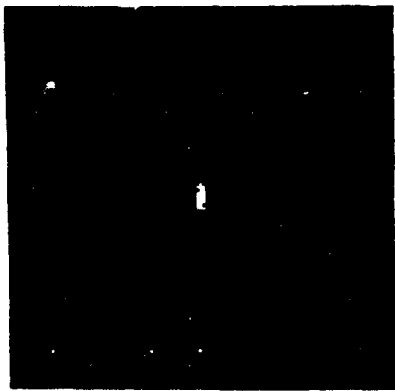

Fig. 6.4. Optimal Region of
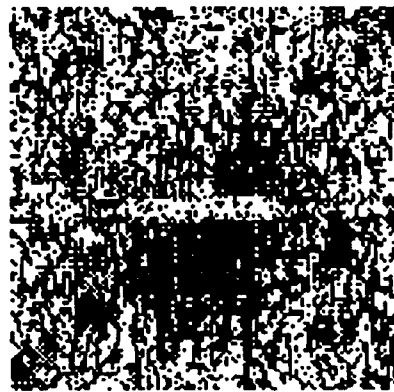Support Under White Noise
Assumption.

Fig. 6.5. Optimal Region of
Support for Estimated Colored
Noise.

More importantly, the correlation outputs must be examined. Figs. 6.6 and 6.7 show

the correlation outputs for ROSs in Figs. 6.4 and 6.5, respectively, when the input

image was the reference in zero background. While both peaks are at correct position,

the peak from the colored noise ROS is much sharper. Figs. 6.8 and 6.9 show the correlation outputs when the input image was the reference in natural background in Fig. 6.2. The difference between the two outputs is dramatic, illustrating the importance of colored noise ROS.
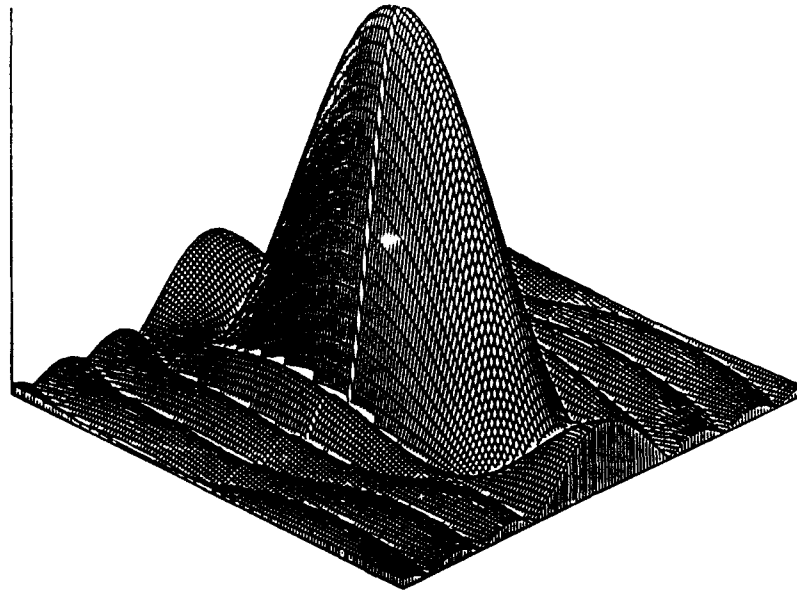

Fig. 6.6.   Correlation output when the reference image in Fig. 6.1 and the ROS in Fig. 6.4 were used.
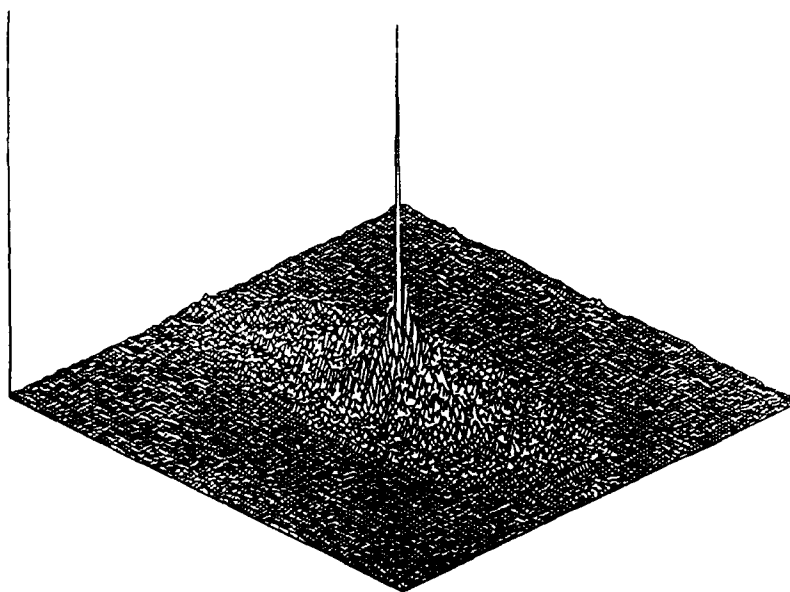
Fig. 6.7.   Correlation output when input is the image in Fig. 6.1 and ROS is the one in Fig. 6.5.
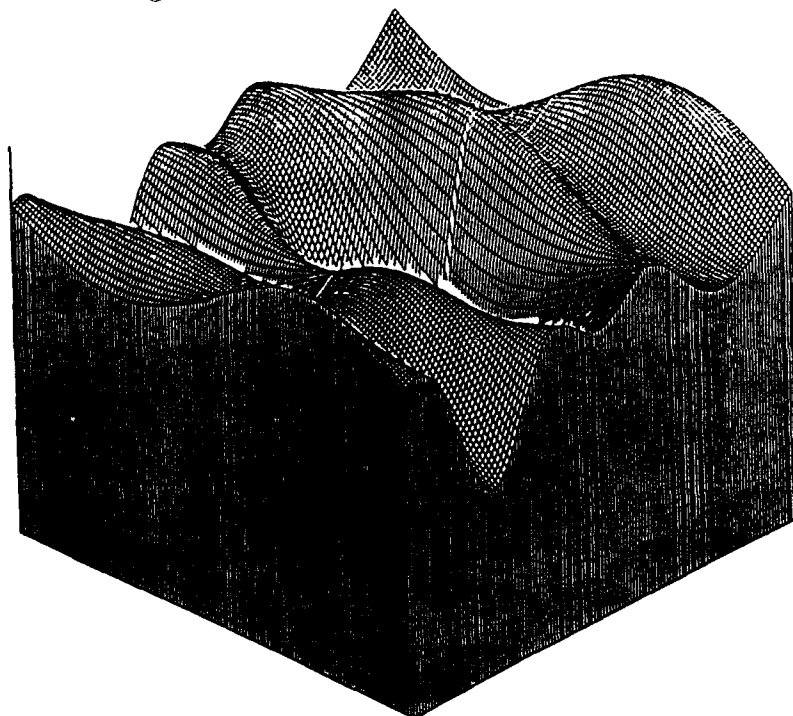
Fig. 6.8.   Correlation output when input is the image in Fig. 6.2 and the ROS is the one in Fig. 6.4.
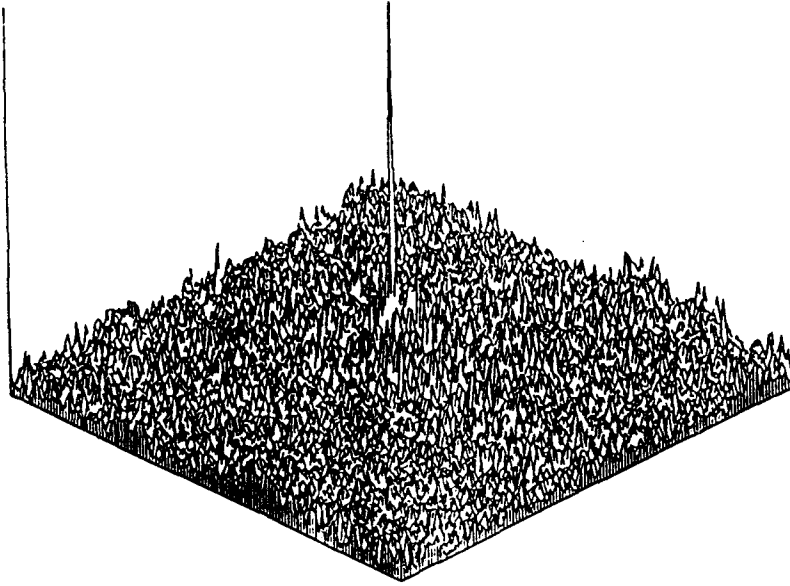
Fig. 6.9.   Correlation output when input is the image in Fig. 6.2 and the ROS is the one in Fig. 6.5.

6.3 Detector Noise

An often ignored aspect of filter design is the detector noise. If there was no detector noise, there would be no need for a light-efficient filter such as the POF. Instead, we can amplify the correlation output by any desired amount and carry out perfect detection. A reasonable model for detector noise is that the observed output is y where

$$y = c(0,0) + n_d,\qquad(6.4)$$

where $c(0,0)$ is the correlation output in the absence of noise and n_d is the detector noise. We assume that n_d is a random variable with mean μ_d and variance σ_d^2. We have shown that the output SNR is then given as follows.

$$SNR_d = \frac{|\int\int H(u,v)\ S(u,v)\ du\,dv\ |^2}{\sigma_d{}^2 + \int\int P_n(u,v)\ |H(u,v)|^2\ du\,dv}$$

(6.5)

When detector noise is zero or small, SNR_d in eq. (6.5) is similar to the usual SNR expression and prior results still hold. But when $\sigma_d{}^2$ is large, maximizing SNR_d is equivalent to maximizing the correlation peak magnitude. Our experiments have shown that when $\sigma_d{}^2$ is large or when the input noise is small, the optimal ROS is large allowing much of the input light to pass. On the other hand, when input noise is large or detector noise is weak, the optimal region of support is small thus preventing much of the input noise. We have derived the rigorous formulae for this. The results are included in *JOSA-A* reprint[23] attached as Appendix D.

7. MAXIMIZING PEAK SHARPNESS

7.1 Motivation

While high output SNRs are certainly needed when using correlation filters, it is also important that the resulting correlation outputs have sharp peaks. Such sharp peaks enable accurate localization of the input target and can improve the discrimination capability of the filters. Prior to our research described in this chapter, there was relatively little research done in designing filters to maximize the peak sharpness. Dickey and Romero[24] showed that the conventional phase-only filters are dually optimal in the sense that they maximize both output SNR and output peak sharpness.

We have proposed a new performance measure called peak-to-correlation energy (PCE). The PCE is defined as below.

$$PCE = \frac{|c(0,0)|^2}{\int \int |c(\tau_x, \tau_y)|^2 d\tau_x \, d\tau_y} \tag{7.1}$$

where $c(\tau_x, \tau_y)$ is the correlation output and $c(0,0)$ is its value at the origin. The numerator in eq. (7.1) is the square of the peak value whereas the denominator yields the total energy in the correlation output. For sharp correlation outputs (e.g., delta functions), the PCE is very high whereas for broad correlation outputs (e.g., constant functions), the PCE is small. Thus designing filters to maximize PCE can lead to filters yielding sharp correlation output peaks.

7.2 Phase-only Filters Maximizing PCE

Using Parseval's theorem, the PCE in eq. (7.1) can be rewritten as

$$PCE = \frac{|\int\int H(u,v)\, S(u,v)\, dudv|^2}{\int\int |H(u,v)|^2\, |S(u,v)|^2\, dudv} \quad . \tag{7.2}$$

If $H(u,v)$ is allowed to be complex, then PCE is maximized by $H(u,v) = 1/S^*(u,v)$, the inverse filter. The inverse filter, of course, suffers from the problems of excessive noise sensitivity.

When $H(u,v)$ is limited to be a phase-only filter with region of support R, the resulting PCE is given as follows.

$$PCE_{POF} = \frac{[\int\int_R |S(u,v)|\, dudv]^2}{\int\int_R |S(u,v)|^2\, dudv} \quad . \tag{7.3}$$

The region of support maximizing the PCE_{POF} in eq. (7.3) was shown by us to be of the following form.

$$R^* = \{\ |S(u,v)| < T\ \} \tag{7.4}$$

where T is an unknown threshold that must be determined by a search. The proof for this is provided in the *Optics Letters* reprint included as Appendix E. Note that the regions of support for maximizing SNR and for maximizing PCE will be different.

7.3 Simulation Results

For testing the advantages of using the optimum-PCE POF, we used the 32×32 missile launcher test image shown in Figure 7.1.
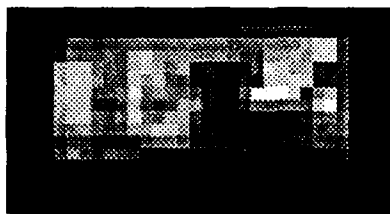
Fig. 7.1. The 32×32 missile launcher image.

As usual, FFTs of size 64×64 were used. The optimum region of support for POFs maximizing the SNR and the PCE are shown in Figs. 7.2 and 7.3, respectively.
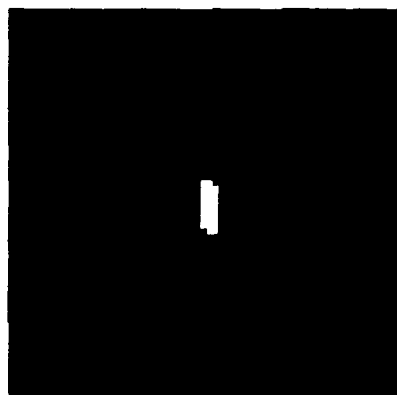
Fig. 7.2. The optimal region of support for the POF maximizing the SNR.
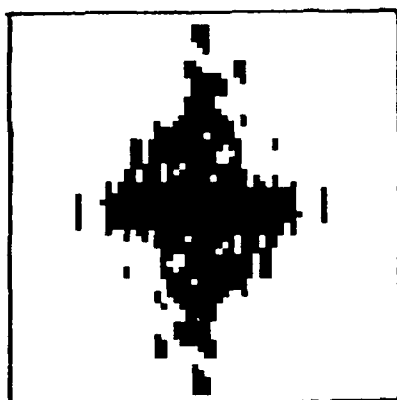
Fig. 7.3. The optimal region of support for the POF maximizing the PCE.

The region of support in Figs. 7.2 and 7.3 are very different. Obviously, the goals of maximizing the SNR are incompatible with maximizing the peak sharpness. In the next chapter, we will explore optimal tradeoffs between the two extremes. In Table 7.1, we show the output SNRs and PCEs for different filters. The input SNR used was 32 dB. Note that optimum-PCE POF provides an increase in PCE of 17 dB over the CMF and of 7 dB over the conventional POF. Figures 7.4 to 7.7 show correlation outputs for the 4 filters. Note that the output from OPCE-POF is the sharpest. We include in Appendix F a reprint that shows how these ideas can be extended to BPOF, CTMF, etc.

43

Table 7.1. SNR and PCE Values for Various Filters

Filter Type	SNR (dB)	PCE (dB)
CMF	68.39	16.00
POF	59.01	26.74
Optimal-SNR POF	66.60	13.71
Optimal-PCE POF	49.67	33.55
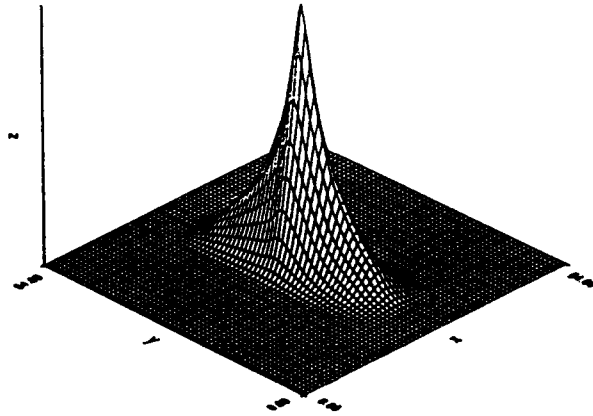
Fig. 7.4. CMF Output

Fig. 7.5. POF Output

Fig. 7.6. Optimum-SNR POF Output
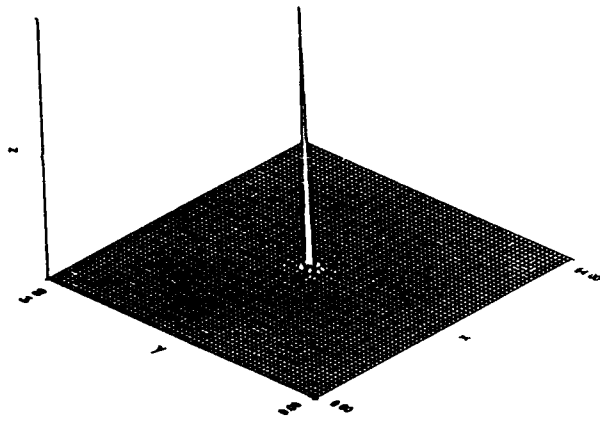
Fig. 7.7. Optimum-PCE POF Output

8. OPTIMUM TRADE-OFF FILTERS

8.1 Motivation

We have seen in Chapter 7 that some of the desired goals in correlation filter design are conflicting. For example, maximizing SNR leads to emphasis on low spatial frequencies whereas maximizing PCEs leads to high-frequency emphasis. It is important to appreciate and understand these tradeoffs and design optimum tradeoff filters. This will help us tailor the filter design to suit particu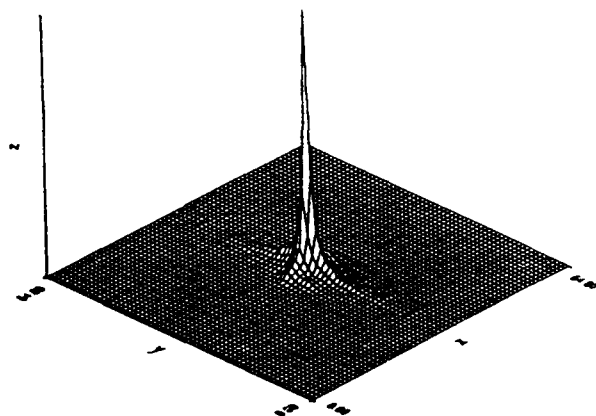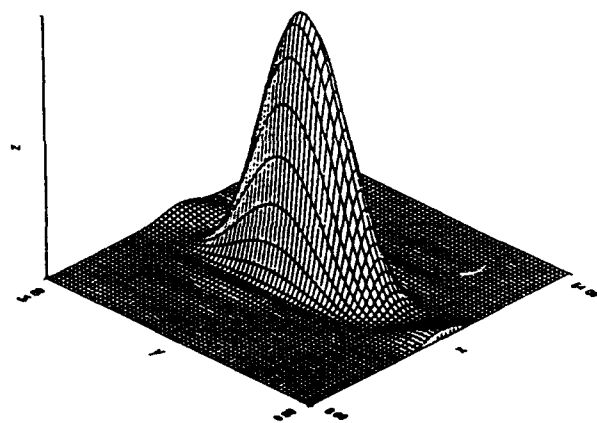lar applications. For example, if the input image is subject to much input noise, it is necessary to optimize the SNR. On the other hand, if the application limits the available power, more light-efficient filters are desired.

While SNR and PCE are useful measures, there are other important measures also. Examples include: Horner efficiency to measure the light throughput capabilities, the accuracy of peak location, discriminability and distortion tolerance. Not all measures can be optimized using one filter. We will present some of the tradeoffs in this chapter.

8.2 Fractional Power Filters

We have seen already that the CMFs maximize the SNR, the inverse filter maximizes the PCE and the POF maximizes the Horner efficiency. We have introduced[26] the following fractional power filters (FPFs) to illustrate the resulting tradeoffs.

$$H_{FPF}(u,v) \;=\; \begin{cases} |S(u,v)|^p \, e^{-j\theta(u,v)} & \text{if } |S(u,v)| \neq 0 \\ \quad\quad 0 & \text{if } |S(u,v)| = 0 \end{cases} \tag{8.1}$$

where p is a real number and $|S(u,v)|$ and $\theta(u,v)$ are the magnitude and phase, respectively, of the Fourier transform of the reference image $s(x,y)$. Here $p = 1$ leads to the CMF, $p = 0$ leads to the POF and $p = -1$ leads to the inverse filter. We can

derive expressions for SNR, PCE and Horner efficiency in terms of p and $|S(u,v)|$. These expressions can be used to display the resulting tradeoffs.

We have used the 32×32 gray level aircraft image shown in Fig. 8.1 for these numerical experiments. This image was placed in a 64×64 array and zero-padded prior to performing a 64×64 FFT. The resulting $|S(u,v)|$ was used to compute SNR, PCE and Horner efficiency as a function of p.
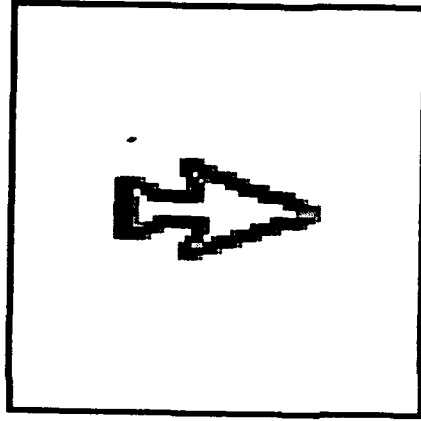
Fig. 8.1. The aircraft used in the numerical experiments.

We show in Fig. 8.2 the three measures (using a dB scale) as a function of p. As expected, the SNR peaks for $p = +1$ (CMF), the PCE peaks for $p = -1$ (inverse filter) and the Horner efficiency peaks for $p = 0$ (the POF). We can also clearly see the resulting tradeoffs. An interesting observation is that the SNR and PCE curves appear to cross at $p = 0$ (the POF). Thus the POF may provide a reasonable compromise for both PCE and SNR while providing maximum Horner efficiency.
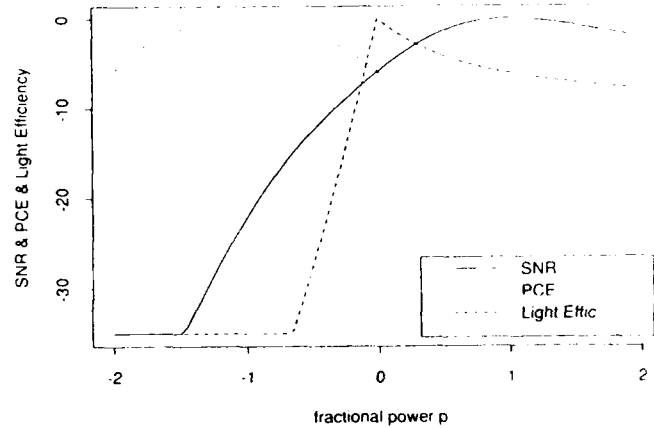
Fig. 8.2. SNR, PCE and Horner Light Efficiency (in dB) as a function of p for the image in Fig. 8.1.

8.3 Optimal Trade-off Filters

Recently, Refregier[27] has derived filters that provide optimum tradeoff among the three measures (SNR, PCE and Horner efficiency). He showed this optimal tradeoff filter (OTF) as given by

$$H(u,v) \;=\; \sigma_\lambda \; \left\{ \; \frac{S(u,v)}{\mu \, P_n(u,v) \, + \, (1-\mu) \, |S(u,v)|^2} \; \right\} \tag{8.2}$$

where $0 \le \mu \le 1$ and the function $\sigma_\lambda \, (\cdot)$ is given as follows.

$$\sigma_\lambda(y) \;=\; \begin{cases} \lambda y & \text{if } |y| \le 1/\lambda \\ e^{-j\psi} & \text{if } |y| > 1/\lambda \end{cases}, \tag{8.3}$$

where ψ is the phase of y. Thus the OTF is a phase-only filter at some frequencies and

a fully complex filter at other frequencies.

When we are interested in designing POFs with optimal tradeoff between SNR and PCE, we can combine the denominators of the two measures to come up with the following compromise performance measure (CPM).

$$CPM = \frac{|\int \int S(u,v)\, H(u,v)\, dudv|^2}{\int \int |H(u,v)|^2\, [P_n(u,v) + \gamma\, |S(u,v)|^2] dudv} \qquad (8.4)$$

where γ denotes the relative emphasis between SNR and PCE. For $\gamma = 0$, CPM is identical to the SNR whereas for very large γ values, CPM is proportional to PCE. For the case of POFs with region of support R, the CPM is given as follows.

$$CPM_{POF} = \frac{[\int \int_R |S(u,v)|\, dudv]^2}{\int \int_R [P_n(u,v) + \gamma\, |S(u,v)|^2]\, dudv} . \qquad (8.5)$$

We have shown[28] that the optimal ROS R^* maximizing this CPM_{POF} is given as follows.

$$R^* = \{ (u,v) : \frac{|S(u,v)|}{P_n(u,v) + \gamma\, |S(u,v)|^2} \geq T \} \qquad (8.6)$$

where T is an unknown threshold to be determined.

We used the binary boundary image shown in Fig. 8.3 for simulation. The resulting tradeoff between SNR and PCE is shown in Fig. 8.4. Similar experiments with BPOFs have also been carried out. Appendix G contains the reprint that details these

tradeoffs.
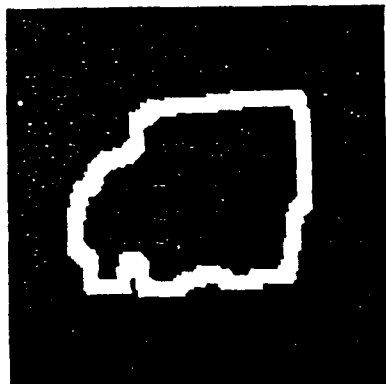
Fig. 8.3. Binary boundary image of a truck.
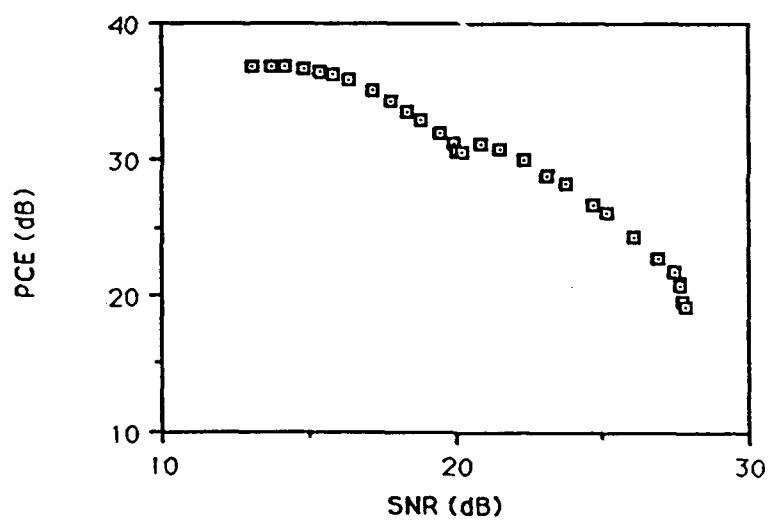
Fig. 8.4. PCE vs. SNR as γ is varied.

9. CONCLUSIONS

Several significant contributions have been made to the area of phase-only and binary phase-only filter design as a result of this research project. We summarize here the major contributions.

- We showed how previous algorithms for designing optimal POFs and BPOFs can be improved to result in a speedup of about two orders of magnitude.

- We introduced a new algorithm known as the successive forcing algorithm (SFA) to design limited modulation (e.g., phase-only and binary phase-only) synthetic discriminant function filters.

- We extended the notion of optimum-SNR POFs and BPOFs to include colored noise in the input scene since colored noise is a better model for the natural background in many input images.

- We designed POFs and BPOFs to maximize the output SNR when detector noise is also present. The presence of detector noise results in the saturation of filter magnitude.

- We introduced complex ternary matched filters (CTMFs) that can be implemented using two ternary devices such as the magneto-optic spatial light modulator (MOSLM).

- We introduced a new performance measure called peak-to-correlation energy

(PCE) to characterize the sharpness of the correlation peak.

- We designed POFs, BPOFs and CTMFs to maximize the resulting PCE.

- We introduced fractional power filters (FPFs) to illustrate the tradeoffs among SNR, PCE and Horner efficiency.

- We designed optimal tradeoff filters (OTFs) capable of providing the optimal tradeoff among various measures.

We believe that these contributions have significantly expanded the capabilities of phase-only and binary phase-only filters thus making them even more attractive for optical implementation.

REFERENCES

(1) A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Trans. on Information Theory*, Vol. 10, 139-145, 1964.

(2) D. O. North, "An analysis of factors which determine signal/noise discriminations in puised-carrier systems," *Proc. of IEEE*, Vol. 51, 1016-1027, 1963.

(3) J. L. Horner, "Light utilization in optical correlator," *Applied Optics*, Vol. 21, 4511-4514, 1982.

(4) W. E. Ross, D. Psaltis and R. H. Anderson, "Two-dimensional magneto-optic spatial light modulator for signal processing," *Optical Engineering*, Vol. 22, 485-490, 1983.

(5) J. L. Horner and P. D. Gianino, "Phase-only matcherd filtering," *Applied Optics*, Vol. 23, 812-816, 1984.

(6) D. Psaltis, E. G. Paek and S. S. Venkatesh, "Optical image correlation with a binary spatial light modulator," *Optical Engineering*, Vol. 23, 698-704, 1984.

(7) C. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition," *Applied Optics*, Vol. 19, 1758-1761, 1980.

(8) A. V. Oppenheim and J. S. Lim, "The importance of phase in signals." *Proc. of IEEE*, Vol. 69, 529-532, 1981.

(9) B. V. K. Vijaya Kumar and Z. Bahri, "Phase-only filters with improved signal-to-noise ratio," *Applied Optics*, Vol. 28, 250-257, 1989.

(10) B. V. K. Vijaya Kumar and Z. Bahri, "Efficient algorithm for designing a ternary-valued filter yielding maximum signal-to-noise ratio," *Applied Optics*, Vol. 28, 1919-1925, 1989.

(11) B. A. Kast, M. K. Giles, S. D. Lindell and D. L. Flannery, "Implementation of a ternary phase-amplitude filter using a magneto-optic spatial light modulator," *Applied Optics*, Vol. 28, 1044-1046, 1989.

(12) D. Flannery, J. Loomis and M. Milkovich, "Design elements of binary phase-olnly correlation filters," *Applied Optics*, Vol. 27, 4231-4235, 1988.

(13) D. Jared and D. Ennis, "Inclusion of filter modulation in synthetic discriminant function construction," *Applied Optics*, Vol. 28, 232-239, 1989.

(14) Z. Bahri and B. V. K. Vijaya Kumar, "Fast algorithm for designing optical phase-only filters (POFs) and binary phase-only filters (BPOFs)," *Applied Optics*, Vol. 29, 2992-2996, 1990.

(15) B. V. K. Vijaya Kumar and J. M. Connelly, "Effects of quantizing the phase in correlation filters," *Proc. of SPIE*, Vol. 1151, 166-173, 1989.

(16) F. M. Dickey and B. D. Hansche, "Quad-phase correlation filters for pattern recognition," *Applied Optics*, Vol. 28, 1611-1613, 1989.

(17) F. M. Dickey, B. V. K. Vijaya Kumar, L. A. Romero and J. M. Connelly, "Complex ternary matched filters yielding high signal-to-noise ratios," *Optical Engineering*, Vol. 29, 994-1001, 1990.

(18) B. V. K. Vijaya Kumar, "Minimum variance synthetic discriminant functions," *JOSA-A*, Vol. 3, 1579-1584, 1986.

(19) A. Mahalanobis, B. V. K. Vijaya Kumar and D. Casasent, "Minimum average correlation energy filters," *Applied Optics*, Vol. 26, 3633-3640, 1987.

(20) D. Casasent, G. Ravichandran and S. Bollapragada, "Gaussian-minimum average correlation energy filters," *Applied Optics*, Vol. 30, 5176-5181, 1991.

(21) Z. Bahri and B. V. K. Vijaya Kumar, "Algorithms for designing phase-only synthetic discriminant functions," *Proc. of SPIE*, Vol. 1151, 138-147, 1989.

(22) B. V. K. Vijaya Kumar and Richard D. Juday, "Design of phase-only, binary phase-only and complex ternary matched filters with increased signal-

to-noise ratios for colored noise," *Optics Letters*, Vol. 16, 1025-1027, 1991.

(23) B. V. K. Vijaya Kumar, Richard D. Juday and P. K. rajan, "Saturated Filters," *JOSA-A*, Vol. 9, 405-412, 1992.

(24) F. M. Dickey and L. A. Romero, "Dual optimality of phase-only filters," *Optics Letters*, Vol. 14, 4-6, 1989.

(25) B. V. K. Vijaya Kumar, W. Shi and C. D. Hendrix, "Phase-only filters with maximally sharp correlation peaks," *Optics Letters*, Vol. 14, 807-809, 1990.

(26) B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," *Applied Optics*, Vol. 29, 2997-3006, 1990.

(27) Ph. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency," *Optics Letters*, Vol. 16, 829-831, 1991.

(28) C. D. Hendrix, B. V. K. Vijaya Kumar, K. T. Stalker, B. A. Kast and R. Shori, "Design and testing of 3-level optimal correlation filters," *Proc. of SPIE*, Vol. 1564, 2-13, 1991.

APPENDIX A

Fast algorithms for designing optical phase-only filters (POFs) and binary phase-only filters (BPOFs)

Zouhir Bahri and B. V. K. Vijaya Kumar

Very efficient suboptimal algorithms for the design of phase-only filters and binary phase-only filters are presented. A reduction of 2- to 3-orders of magnitude in computer time is obtained over previous algorithms. The loss in signal-to-noise ratio is negligible (<0.001 dB).

I. Introduction

Phase-only filters (POFs)[1,2] and binary phase-only filters (BPOFs)[3-7] have received much attention in optical pattern recognition research. While earlier work in the area of POFs and BPOFs has been more or less adhoc, several research groups[6,8,9] have recently started to investigate optimal ways of designing POFs and BPOFs. Most of these efforts have focused on optimizing the output signal-to-noise ratio (SNR). In addition, we have introduced[8,9] algorithms to numerically determine the optimal POFs and BPOFs for the detection of arbitrary reference images corrupted by additive white noise. Both optimal POFs and optimal BPOFs are characterized by regions of support. When a spatial frequency is not included in the region of support, the resulting filters have zero magnitude at that frequency. Thus, a BPOF coupled with a region of support is, strictly speaking, a ternary-valued ($-1, 0$, and $+1$) filter.

Here, we provide very efficient suboptimal algorithms for the design of regions of support for the optimal POFs and BPOFs. These very efficient algorithms result in an impressive reduction in CPU time (2 to 3 orders of magnitude). This is done at the expense of a very small loss in output SNR (<0.001 dB for all cases we tested). This speedup in BPOF design should allow determination of the optimal BPOFs adaptively in real time.

The remainder of this paper is organized as follows. In Sec. II., we provide some background material in which we briefly summarize the earlier algorithms. In Sec. III., we present the new efficient suboptimal algo-

rithm for the design of POFs. Section IV. is analogous to Sec. III. except it is concerned with BPOFs. Finally, Sec. V. summarizes our results.

II. Background

A. Phase-Only Filters

We have shown elsewhere[8] that the OPOFs have the same phase as that of the classical matched filters. The SNR obtainable from the OPOFs can be increased further by introducing a region of support R. This region R determines the spatial frequencies for which the OPOF is not zero. The role of this R is to suppress the frequencies which have little signal power and a lot of noise power and in the process to improve the SNR. Thus, the OPOF, strictly speaking, is not a POF since it accommodates two magnitudes (zero and one). However, even conventional POFs have two magnitudes since the filters are always contained in an aperture. Our earlier research has determined that the optimal R must maximize the following SNR expression:

$$\text{SNR}_{\text{OPOF}} = \frac{\left[\int_R |S(f)| df \right]^2}{\int_R P_n(f) df}. \tag{1}$$

In the above equation $S(f)$ denotes the Fourier transform (FT) of the reference image $s(x)$ and $P_n(f)$ the noise power spectral density. We will use 1-D notation throughout for simplicity, but all our results can be easily extended to higher dimensions. For the case of white noise, the denominator in Eq. (1) depends only on the area of R and not on which exact spatial frequencies are included in this region. Thus, for a given size of R, we must include those spatial frequencies that maximize the numerator of Eq. (1). To do this on a digital computer, we use the discrete notation. To accomplish this, we reorder the samples of the signal Fourier transform as below:

$$|S(1)| \geq |S(2)| \geq \ldots \geq |S(d)|, \tag{2}$$

where d is the number of samples in the signal discrete Fourier transform (DFT) $S(k)$. Let K denote the number of pixels in the region of support. By choosing K pixels corresponding to the first K signal DFT values in Eq. (2), we will maximize the resulting SNR for that choice of K. The corresponding SNR is given by

$$SNR(K) = \alpha \frac{\left[\sum_{i=1}^{K} |S(i)| \right]^2}{K}, \qquad (3)$$

where α is a constant that depends on the sampling interval used (to convert integrals into summations) and the level of the white noise present in the input signal. We numerically evaluate $SNR(K)$ for all possible choices of K and select the one that maximizes it. Notice that without the sorting in Eq. (2), for each value of K, we must try all possible regions of support consisting of K pixels. This is not practical even for small K values.

B. Binary Phase-Only Filters

We have shown elsewhere[9] that the optimal BPOF (OBPOF) takes the form

$$H(f) = I_R(f) \, Sgn[S_r(f) \cos(\theta^*) + S_i(f) \sin(\theta^*)], \qquad (4)$$

where $I_R(f)$ is the indicator function for region R and where $Sgn(x)$ is defined as

$$Sgn(x) = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}. \qquad (5)$$

In the above, θ^* denotes the (yet to be determined) optimal threshold line angle (TLA)[5] and $S_r(f)$, $S_i(f)$ denotes the real (imaginary) part of $S(f)$. To determine the OBPOF, we scan all values of TLA θ (from 0 to $\pi/2$). For each θ, the R that leads to the highest SNR is determined and the corresponding SNR is noted. The TLA that defines the OBPOF (θ^*) corresponds to the largest among the SNRs. To find the optimal region of support for a given TLA, we partition R into R_1 and R_2 defined as

$$R_1 = \{ f \in R : |S_r(f) \cos(\theta)| > |S_i(f) \sin(\theta)| \} \qquad (6)$$

$$R_2 = \{ f \in R : |S_r(f) \cos(\theta)| \leq |S_i(f) \sin(\theta)| \}. \qquad (7)$$

The SNR (for the TLA in question) can then be written as

$$SNR = \frac{\left[\int_{R_1} |S_r(f)| df \right]^2 + \left[\int_{R_2} |S_i(f)| df \right]^2}{\int_{R_1 \cup R_2} P_n(f) df}, \qquad (8)$$

where in the above, we assumed that R_1 and R_2 are even symmetric.

For white noise, we can apply the same idea (in the discrete domain) as in the OPOF case to optimize the SNR in Eq. (8). For a given θ, we define the two regions

$$P_1 = \{ n : |S_r(n) \cos(\theta)| > |S_i(n) \sin(\theta)| \}, \qquad (9)$$

$$P_2 = \{ n : |S_r(n) \cos(\theta)| \leq |S_i(n) \sin(\theta)| \}. \qquad (10)$$

The values $|S_r(n)|$ for $n \in P_1$ and the values $|S_i(n)|$ for $n \in P_2$ are then sorted in descending order. Hence, a discrete approximation for Eq. (8) is

$$SNR = \alpha \frac{\left[\sum_{n=1}^{K_1} |S_r^n| \right]^2 + \left[\sum_{n=1}^{K_2} |S_i^n| \right]^2}{K_1 + K_2}, \qquad (11)$$

where K_1 and K_2 denote the number of pixels in P_1 and P_2, respectively and α is a constant that includes the white noise level and the discretization constants. Here, the superscript n in $|S_r^n|$ and $|S_i^n|$ refers to the nth largest value in the corresponding array. Next, we find that optimal values of K_1, K_2 through a search along the grid (i.e., we fix K_1 and find the optimal value of K_2 and repeat this process for all values of K_1 and choose the best case). This will yield the OBPOF.

A subtle aspect of the above algorithm (not realized earlier) must be pointed out. The θ used in the BPOF form of Eq. (4) must equal the phase of the complex correlations response at the origin when this BPOF is used. When the BPOF of this form, using θ as the TLA, produces the same θ as the phase of the correlation response at origin, we call such a BPOF a feasible filter. In general, there is no guarantee that the BPOFs defined according to Eq. (4) are feasible. However, we have shown recently[10] that when we find the θ^* yielding the best SNR among all TLAs, the corresponding BPOF is indeed a feasible filter and is the OBPOF we are searching for. However, the filters maximizing the SNRs for specific θ values are not necessarily feasible. Thus, the above algorithm works only because we scan all values of θ. Dickey et al.[11] have recently proved the optimality of the above algorithm for a more general filter that they term as the complex ternary matched filter (CTMF).

III. Efficient Suboptimal Algorithm for POF design

A. Analysis

We start the analysis of the new POF algorithm by the following proposition concerning the optimal region of support R.

Proposition 1: If two or more frequencies of the signal Fourier transform have equal magnitudes, either all of them or none of them are included in the optimal region of support.

Proof: It is sufficient to prove the above proposition for two frequencies only. The case of three or more pixels easily follows by deduction. Assume $|S(K_0)| = |S(K_0 + 1)|$ for some $K_0 \geq 2$ (For $K_0 = 1$, it is easy to show that the use of the first two frequencies rather than the first frequency by itself will double the resulting SNR). As always, $|S(K)|$ are sorted as in Eq. (2). Also, $SNR(K_0)$ denotes the SNR obtained including the first K_0 frequencies in R. There are two cases to be considered. The first, is that

$$SNR(K_0) \geq SNR(K_0 - 1). \qquad (12)$$

We now propose to show that in this case SNR $(K_0 + 1)$ $\geq$ SNR(K_0), hence implying that both K_0 and $K_0 + 1$ must be included in the optimal region of support. Let

$$\beta = \sum_{i=1}^{K} |S(i)|. \qquad (13)$$

We will assume from now on that the constant α is equal to one. This will not affect our optimization process. Then using Eq. (13) in Eq. (3), we obtain

$$\mathrm{SNR}(K_0 + 1) - \mathrm{SNR}(K_0)$$

$$= \frac{(\beta + |S(K_0)|)^2}{K_0 + 1} - \frac{\beta^2}{K_0}$$

$$= \frac{1}{K_0(K_0 + 1)} [K_0(|S(K_0)|^2 + 2\beta|S(K_0)|) - \beta^2]. \qquad (14)$$

Similarly, using Eqs. (3), (12), and (13), we obtain the following result:

$$\mathrm{SNR}(K_0) - \mathrm{SNR}(K_0 - 1)$$

$$= \frac{\beta^2}{K_0} - \frac{(\beta - |S(K_0)|)^2}{(K_0 - 1)}$$

$$= [-\beta^2 + K_0(2\beta|S(K_0)| - |S(K_0)|^2)]/[K_0(K_0 - 1)] \geq 0. \qquad (15)$$

Using the above result, we can obtain the following inequality:

$$K_0(2\beta|S(K_0)| + |S(K_0)|^2) - \beta^2 \geq 0. \qquad (16)$$

If we use Eq. (16) in Eq. (14), we obtain the desired result that $\mathrm{SNR}(K_0 + 1)$ is larger than $\mathrm{SNR}(K_0)$. The second case is that

$$\mathrm{SNR}(K_0) < \mathrm{SNR}(K_0 - 1). \qquad (17)$$

In this case, it is not difficult to see that

$$K_{opt} < K_0 \text{ or } K_{opt} \geq K_0 + 1, \qquad (18)$$

indicating that either both pixels K_0 and $K_0 + 1$ or none of the two are included in the optimal support. When we have three or more frequencies with identical magnitudes, it is easy to show that either all of them or none of them are included in the optimal region R. To show this, we apply the above result to all possible pairs of frequencies in this set of equal magnitude frequencies. This completes the proof of the proposition.

The above proposition suggests an improved algorithm. Suppose we quantize the magnitudes $|S(i)|$, $i = 1, \ldots, d$ into N levels. Then, we do not have to compute the SNR for all d pixels. It is enough to compute the SNR for the N quantization levels only, knowing from the above proposition that pixels with the same quantization level will either all or none be included in the optimal region of support. Hence, we have a new, very efficient suboptimal algorithm for the selection of region of support for POF. The efficiency of this new algorithm stems from the fact that it is, in principle, independent of the number of pixels and depends primarily on the number of quantization levels which is expected to be smaller than the number of pixels for practical cases. The suboptimality of this new algorithm is due to the approximation that all pixel magnitudes falling in the same quantization level are equal. This approximation can be made more accurate by increasing N (the number of quantization levels).
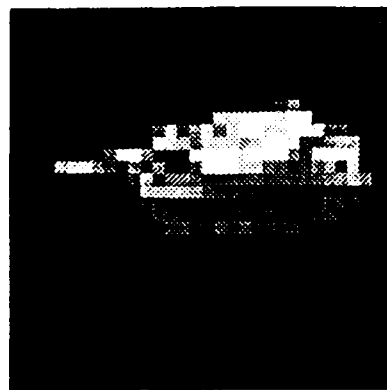

Fig. 1. The 32 × 32 tank image used in the simulations.

B. Algorithm

Below, we list the basic steps of the proposed algorithm.

Step 1: This normalizes and initializes variables. The signal FT magnitudes $|S(n)|$ are normalized by dividing all of them by the largest magnitude.

Step 2: This step quantizes all $|S(n)|$ values to N levels. We also assign tags to frequencies to indicate which quantization level that frequency belongs to.

Step 3: This step searches through the N quantization levels to determine the optimum quantization level.

Step 4: This step determines the optimal region of support using the optimal quantization level determined in Step 3.

C. Numerical Results

We have implemented and tested this new algorithm. In our computer simulations, we used the 32 × 32 pixel gray level tank image shown in Fig. 1. To compute the OPOF, the tank image is zero padded and a 64 × 64 FFT is performed. For the comparison between the old and new algorithms to be fair, we used the most efficient sorting technique (HEAP SORT[12]) to implement the old algorithm. Using $N = 256$ levels, we obtained in 0.13 s the same optimal region of support we got in 10.15 s with the old algorithm. This is a reduction in CPU time by a factor of 78. As already mentioned, we expect this factor to increase even further as the number of frequency pixels increase. To test this, we increased the size of the FFT to 128 × 128 and applied our new algorithm with $N = 256$. It took 0.5 s to find a region of support made of 235 pixels to yield a numerically computed SNR (assuming unit variance white noise) of 76.46116 (= 18.834409 dB). Using the old algorithm, it took 64.6 s to find the optimal region of support consisting of 237 pixels to yield a numerically computed SNR of 76.46427 (= 18.834585 dB). The new algorithm has now achieved a CPU time reduction factor of almost 130! There is a small price though. We have given up ~0.00018 dB in SNR. By increasing N from 256 to 400, we obtained the optimal region of support exactly (i.e., 237 pixels) in practically the same CPU time (0.5 s).

Before moving to the next section, a final assessment of the new algorithm is due. It seems that this newly proposed algorithm has provided us with tremendous savings in computer time (about 2 orders of magnitude) at the expense of a very small loss in SNR ($<2 \times 10^{-4}$ dB). As argued above, we can even get exact results with a saving of at least 1 order of magnitude in CPU time (this being a rather conservative number). It turns out that both the new and old algorithms use the same memory storage, since the d-dimensional array that is used in the old algorithm for the sorting procedure is used in the new algorithm for tagging the pixels (i.e., to which level each pixel belongs). We must mention that all the CPU times presented above do not take into account the time to compute the FFT, get the magnitude of the pixels and normalize them, and the input/output operations. All these computations are common to both algorithms and, hence, will affect both the same way.

IV. Efficient Suboptimal Algorithm for BPOF Design

A. Analysis

As in the POF case, we put forth the following proposition:

Proposition 2: If the real (imaginary) part of two or more pixels in $P_1(P_2)$ are equal in absolute value, either all of them or none of them are included in the optimal $P_1(P_2)$.

Proof: The proof is similar to that of the corresponding proposition in the previous section. The only difference is that the optimization is now performed with respect to two variables K_1 and K_2 rather than with respect to just one variable K_0 as in the case of OPOFs. Hence, the only thing that needs to be verified is that by replacing the discrete version of the SNR expression [given in Eq. (3)] by

$$\text{SNR}(K) = \frac{\left[\sum_{i=1}^{K} |S(i)|\right]^2 + a}{K + b}, \tag{19}$$

where a and b are arbitrary non-negative constants, the conclusion of the first part of the proof of proposition in the previous section is unaffected. Namely, it can be established in exactly the same way as before that if $|S(K_0)| = |S(K_0 + 1)|$ for some $K_0 \geq 2$ and if $\text{SNR}(K_0) \geq \text{SNR}(K_0 - 1)$, then $\text{SNR}(K_0 + 1) \geq \text{SNR}(K_0)$. Here $|S(i)|$ refers to the appropriate real part or the imaginary part of the signal FT. This completes the proof of the proposition.

As before, the idea is to quantize $|S_r(n)|$, $n \in P_1$ and $|S_i(n)|$, $n \in P_2$. By virtue of the above proposition, instead of performing the optimization over the $K_1 \times K_2$ grid of pixels ($K_1(K_2)$ is the total number of pixels in $P_1(P_2)$), we focus only on a $N \times N$ grid (N is the number of quantization levels).

B. Algorithm

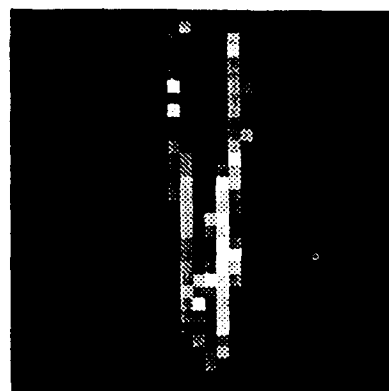Below, we list the basic steps of the efficient suboptimal algorithm:

Fig. 2. The 32 × 32 pliers image used in the simulations.

Step 1: Find the partitions P_1 and P_2.

$M1 = \text{Max}\{|S_r(n)|, n \in P_1\}$,

$M2 = \text{Max}\{|S_i(n)|, n \in P_2\}$,

$M = \text{Max}[M1, M2]$,

$S(n) \leftarrow S(n)/M, n = 1, \ldots, d$.

Step 2: Quantize $|S_r(n)|$, $n \in P_1$ and $|S_i(n)|$, $n \in P_2$. Assign tags to the corresponding pixels.

Step 3: Search through an $N \times N$ grid to get optimal levels L_1 and L_2.

Step 4: Get optimal $P_1(P_2)$ by direct comparison with $L_1(L_2)$.

C. Numerical Results

We have implemented and tested the above algorithm. Using the same 32 × 32 pixel tank image and a 64 × 64 (the tank has been centered) FFT, it took 1585s with the original algorithm to find the optimal region of support. Optimal R was obtained by trying nineteen TLAs, from 0 to 90 in increments of 5°. Applying the new algorithm with $N = 250$, it took only 16.8 s to obtain eighteen of these regions exactly and one region within 2 pixels (60 instead of 58 pixels). The CPU time reduction factor is ~95. By increasing N to 400, all the regions were computed correctly in 36.4 s. This corresponds to a CPU time reduction factor of 44. It can be seen from here that the computing time of the new BPOF algorithm is a faster increasing function of N than in the OPOF case. This is a consequence of searching over a grid rather than along a line.

As before, the CPU time reduction factor becomes more significant as the number of pixels increases. We increased the FFT size to 128 × 128 pixels and focused on the optimal support for the Hartley BPOF (TLA = 45°). It took 1939 s for the old algorithm to find the optimal region of support; whereas, the new algorithm with $N = 256$ took only 1.98 s to find the same answer. This is a time reduction factor of ~980!

We tested this algorithm using several other images (such as a noncentered tank image, centered and noncentered pliers image). The centered pliers image is shown in Fig. 2. Once again, FFTs of size 64 × 64 pixels were used. For $N = 256$, fifty-six out of the seventy-seven images tried resulted in exact regions of

support with an average reduction factor in CPU time of ~77. The worst SNR loss was 0.00046 dB. For $N = 400$, sixty-four out of the seventy-seven regions of support were found exactly with an average CPU time reduction factor of ~35. The worst loss in SNR was well within that of the case with $N = 256$. Once again, all the above timing estimates were done excluding FFTs, inputs/outputs, partitioning, and normalization.

V. Summary

We have presented two very efficient suboptimal techniques for the design of regions of support for POFs and BPOFs. These new algorithms seem to provide a reduction in CPU time from 2 to 3 orders of magnitude. The loss in SNR observed in all cases treated was negligible (<0.001 dB). We believe that these new algorithms will make it possible for optimal POFs and optimal BPOFs to be designed in real time.

References

1. J. L. Horner and P. D. Gianino, "Phase-Only Matched Filtering," Appl. Opt. 23, 812–816 (1984).
2. R. R. Kallman, "Direct Construction of Phase-Only Filters," Appl. Opt. 26, 5200–5201 (1987).
3. J. L. Horner and H. O. Bartelt, "Two-Bit Correlation," Appl. Opt. 24, 2889–2893 (1985).
4. D. Psaltis, E. G. Paek, and S. S. Venkatesh, "Optical Image Correlation with a Binary Spatial Light Modulator," Opt. Eng. 23, 698–704 (1984).
5. D. L. Flannery, J. S. Loomis, and M. E. Milkovich, "Design Elements of Binary Phase-Only Correlation Filters," Appl. Opt. 27, 4231–4235 (1988).
6. M. W. Farn and J. W. Goodman, "Optimal Binary Phase-Only Matched Filters," Appl. Opt. 27, 4431–4437 (1988).
7. F. M. Dickey, K. T. Stalker, and J. J. Mason, "Bandwidth Considerations for Binary Phase-Only Filters," Appl. Opt. 27, 3811–3818 (1988).
8. B. V. K. Vijaya Kumar and Z. Bahri, "Phase-Only Filters With Improved Signal-to-Noise Ratio," Appl. Opt. 28, 250–257 (1989).
9. B. V. K. Vijaya Kumar and Z. Bahri, "Efficient Algorithm for Designing a Ternary Valued Filter Yielding Maximum Signal to Noise Ratio," Appl. Opt. 28, 1919–25 (1989).
10. Z. Bahri, "Phase-Only and Binary Phase-Only Filters for Optical Correlators," Ph.D. Dissertation, Carnegie Mellon University, 1989.
11. F. M. Dickey, B. V. K. Vijaya Kumar, L. A. Romero and J. M. Connelly, "Complex Ternary Matched Filters Yielding High Signal-to-Noise Ratios," Optical Eng. 29, (1989).
12. R. Conway and D. Gries, An Introduction to Programming: A Structured Approach Using PL/I and PL/C (Winthrop Publishers, Cambridge, MA, 1979).

APPENDIX B

Complex ternary matched filters yielding high signal-to-noise ratios

Fred M. Dickey, MEMBER SPIE
Sandia National Laboratories
Albuquerque, New Mexico 87185

B. V. K. Vijaya Kumar, MEMBER SPIE
Carnegie Mellon University
Department of Electrical
  and Computer Engineering
Pittsburgh, Pennsylvania 15213

L. A. Romero
Sandia National Laboratories
Albuquerque, New Mexico 87185

James M. Connelly
Carnegie Mellon University
Department of Electrical
  and Computer Engineering
Pittsburgh, Pennsylvania 15213

Abstract. Complex ternary matched filters (CTMFs) can be implemented optically using detour phase and a single ternary spatial light modulator or a Mach-Zehnder arrangement and two ternary signal light modulators. In this paper, we present the design of CTMFs that yield high signal-to-noise ratios (SNRs) and show with the help of simulation examples that their performance (in the sense of SNRs) is close to that of optimal matched filters.

Subject terms: optical pattern recognition; correlators; matched filtering; phase-only filters; signal-to-noise ratio.

Optical Engineering 29(9), 994-1001 (September 1990).

CONTENTS

1. INTRODUCTION

Although matched spatial filters originally introduced by VanderLugt[1] provide the highest output signal-to-noise ratio (SNR), they have not become practical in many applications due to the complex nature of the spatial filter required. Lately, much research effort has been devoted to methods avoiding the use of complex spatial filters. This research has suggested the use of phase-only filters (POFs),[2-6] binary phase-only filters (BPOFs),[7-10] ternary-valued spatial filters,[11-13] and quad-phase-only filters (QPOFs),[14,15] among others. Research effort has also been directed toward the analysis[16-20] of the performance of these partial-information filters.

The BPOFs are attractive from implementational considerations because of the availability of rapidly programmable binary spatial light modulators (SLMs) such as the magneto-optic SLM.

The SNRs obtainable using BPOFs were bounded[17] to be no worse than 6 dB below the SNRs of POFs and were observed[12] in numerical experiments to be about 4 to 6 dB below the SNRs obtained using the best POFs. However, when compared with the SNR of the classical matched filter, the best BPOF SNRs were still about 4 to 5 dB lower. In this paper, we propose a practical way of using ternary SLMs[13] to obtain SNRs very close to those obtained using the classical matched filter. The basic idea is to use two ternary SLMs, one serving as the real part of a complex function and the other serving as the imaginary part. These filters can be implemented using a Mach-Zehnder arrangement with one ternary SLM in each path. The Mach-Zehnder approach is conceptually simple. Another means of implementation is the detour phase method in Ref. 15, which requires only one SLM. As shown earlier for POFs[6] and BPOFs,[12] it is necessary to mask out certain spatial frequencies in each BPOF. Thus, both the real part and the imaginary part of this complex filter can take on three values ($+1$, -1, and 0) and hence the name complex ternary matched filters (CTMF). This paper presents the basic analysis of CTMFs, discusses the optical implementation, and presents simulation results to illustrate the advantages of CTMFs.

In the next section we provide a brief background to help set up the notation. Then, Sec. 3 introduces CTMFs and describes how the CTMF support function can be selected to yield maximum SNR. An efficient algorithm for designing these optimal support functions is presented in Sec. 4. In Sec. 5 we discuss the optical implementation of CTMFs and discuss some associated issues. The results of our numerical simulations are shown in Sec. 6 to illustrate the advantages of the proposed method.

2. BACKGROUND

The problem under consideration is the detection of a "known" signal/image $s(x)$ in the presence of additive, zero mean, stationary noise $n(x)$. While we will use the 1-D notation throughout for convenience, all of our results are easily generalized to higher dimensions. It is common knowledge that the matched filter provides the highest SNR for this problem. The SNR is defined as

$$SNR = \frac{|E\{C(0)\}|^2}{var\{C(0)\}} ,$$ (1)

where $C(0)$ denotes the filter output at $x = 0$ and $E\{\cdot\}$ and var$\{\cdot\}$ denote the ensemble average and variance, respectively. This SNR measure characterizes the sensitivity of the filter to noise in the input image. The linear invariant filter that maximizes the above SNR is given by

$$H_{MF} = \alpha \frac{S^*(f)}{P_n(f)} = \alpha \frac{|S(f)|}{P_n(f)} \exp[-j\theta_s(f)] ,$$ (2)

where α is an arbitrary complex constant, $P_n(f)$ is the noise power spectral density, $S(f)$ is the Fourier transform (FT) of the reference signal $s(x)$, and $|S(f)|$ and $\theta_s(f)$ denote its magnitude and phase, respectively. For white noise, $P_n(f)$ is a constant independent of f and the matched filter has a transfer function equal to $\alpha S^*(f)$.

It can be easily seen from Eq. (2) that the matched filter is complex valued. This poses two problems. The first problem is that the light throughput efficiency[21] of this filter is low because that the filter magnitude is generally less than unity. The second problem is that most available SLMs cannot accommodate complex-valued functions. A possible solution to the first problem is to employ the phase-only filter, given by

$$H_{POF}(f) = \exp[-j\theta_s(f)] .$$ (3)

The POF given in Eq. (3) yields 100% light efficiency and produces sharp correlation peaks but suffers from being very sensitive to input noise because of its all-pass nature. Vijaya Kumar and Bahri recently[6] introduced the optimal phase-only filter (OPOF) in which they design the POF support function to obtain the maximum SNR.

While the POFs discussed here meet the requirement of higher light efficiency, they still require an SLM capable of representing a continuum of phase values. Since some of the popular SLMs can accommodate only two phase levels (0 and π), much research interest[7-10] has been focused on binary phase-only filters. Early versions of BPOFs employed the binarized versions of either the real part or the imaginary part of $S(f)$, and some later versions used the binarization of the Hartley transform of the signal. As in the case of POFs, it was realized that we must select appropriate support functions[12] even for BPOFs. The added constraint of a support function results in the filters' taking on three values ($+1$, 0, -1) rather than just two values. Kast et al.[13] showed recently that the magneto-optic SLM (MOSLM) can be configured to yield the three desired transmittance values.

Recently, Dickey and Hansche[14] introduced the quad-phase-only filter, defined as

$$H_{QPOF}(f) = sgn[S_R(f)] - jsgn[S_I(f)] ,$$ (4)

where the signal Fourier transform is given by

$$S(f) = S_R(f) + jS_I(f)$$ (5)

and where sgn$[\cdot]$ is defined as

$$sgn[x] = \begin{cases} +1 , & \text{if } x \geq 0 , \\ -1 , & \text{if } x < 0 . \end{cases}$$ (6)

The advantage of using the QPOF in Eq. (4) compared to the use of a BPOF is that (in some sense) we are encoding both the real part and the imaginary part of the matched filter. Dickey and Hansche[14] show that the SNRs obtainable from QPOFs can, in some cases, be 3 dB more than those from the BPOFs designed from Hartley transforms. They also show that the QPOFs can be implemented using two binary SLMs in a Mach-Zehnder arrangement or by using one SLM and the method f detour phase.[15]

3. COMPLEX TERNARY MATCHED FILTERS

An obvious improvement to the quad-phase filters discussed in Sec. 2 is the CTMF:

$$H_{CTMF}(f) = \eta[H_1(f) - jH_2(f)] ,$$ (7)

where both $H_1(f)$ and $H_2(f)$ are ternary filters, i.e., they take on values $+1$, 0, or -1. The arbitrary complex constant η does not affect SNRs and will be taken as 1 without any loss of generality from now on. Note that we are not assuming a priori

that $H_1(f)$ is obtained by binarizing $S_R(f)$ and that $H_2(f)$ is obtained by binarizing $S_I(f)$. In this section, we show that the best choices for $H_1(f)$ and $H_2(f)$ are indeed obtained by binarizing $S_R(f)$ and $S_I(f)$ (or some linear combination of these two) once the support functions are fixed.

Let us denote by R_i the region of spatial frequencies for which $H_i(f)$ is nonzero, i.e., takes on values $+1$ and -1. When the CTMF of Eq. (7) is used in a correlator, there is no guarantee that the correlation output has its maximum value at the origin.[16] However, we consider the output SNR at the origin for our maximization for the following reasons: First, the resultant SNR expression is analytically tractable. Second, the peak will be very close to the origin even though it may not be exactly at the origin. Third, if the correlation peak is not at the origin, then the origin SNR estimates will be conservative, and optimizing these can only imply even higher SNR value somewhere else in the correlation plane. The position of the correlation peak with respect to the origin is determinable and should not be a problem for tracking. The origin SNR of Eq. (1) can be written as

$$\text{SNR}_{\text{CTMF}} = \frac{\left| \int S(f)H(f)df \right|^2}{\int P_n(f)|H(f)|^2 df}$$

$$= \frac{\left| \int S(f)H(f)df \right|^2}{\int P_n(f)[H_1^2(f) + H_2^2(f)]df}$$

$$= \frac{\left| \int S(f)[H_1(f) - jH_2(f)]df \right|^2}{\int_{R_1} P_n(f)df + \int_{R_2} P_n(f)df} . \tag{8}$$

Note that the denominator of the SNR expression in Eq. (8) is fixed once support regions R_1 and R_2 are fixed. Let us now assume that R_1 and R_2 are fixed and determine the signs of $H_1(f)$ and $H_2(f)$ to maximize the numerator of the SNR expression. Toward this end, let us denote the complex value resulting from the integral in the numerator in Eq. (8) as $|C|\exp(j\beta)$, i.e.,

$$\int S(f)[H_1(f) - jH_2(f)]df = |C|\exp(j\beta) . \tag{9}$$

Thus, for a given $S(f)$, every possible filter $H(f)$ has an associated β, the phase of the correlation output at the origin. Thus, maximizing the numerator of Eq. (8) is equivalent to maximizing $|C|$, given as

$$|C| = \int H(f)S(f)\exp(-j\beta)df$$

$$= \int |H(f)| \, |S(f)| \exp\{j[\theta_s(f) - \theta_h(f) - \beta]\}df , \tag{10}$$

where $|H(f)|$ and $|S(f)|$ denote the magnitudes of $H(f)$ and $S(f)$ and $-\theta_h(f)$ and $\theta_s(f)$ denote their phases. Since $|C|$ is real, the imaginary terms on the right-hand side of Eq. (10) cancel out. Equation (10) can thus be written as

$$|C| = \int [\hat{S}_R(f)H_1(f) + \hat{S}_I(f)H_2(f)]df , \tag{11}$$

where $\hat{S}_R(f) = \text{Re}[S(f)\exp(-j\beta)]$ and $\hat{S}_I(f) = \text{Im}[S(f)\exp(-j\beta)]$. Clearly, to maximize Eq. (11) we must choose

$$H_1(f) = \begin{cases} \text{sgn}[\hat{S}_R(f)] , & f \in R_1 , \\ 0 , & \text{otherwise} , \end{cases} \tag{12}$$

$$H_2(f) = \begin{cases} \text{sgn}[\hat{S}_I(f)] , & f \in R_2 , \\ 0 , & \text{otherwise} . \end{cases} \tag{13}$$

Thus, we can rewrite the CTMF yielding the highest SNR as

$$H_{\text{CTMF}}(f) = I_{R_1}(f)\text{sgn}[\hat{S}_R(f)] - jI_{R_2}(f)\text{sgn}[\hat{S}_I(f)]$$

$$= I_{R_1}(f)\text{sgn}[S_R(f)\cos\beta + S_I(f)\sin\beta]$$

$$-jI_{R_2}(f)\text{sgn}[-S_R(f)\sin\beta + S_I(f)\cos\beta] , \tag{14}$$

where support functions I_{R_i} are 1 for f contained in R_i and 0 for f not in R_i. When β is 0 in Eq. (14), $H_1(f)$ binarizes $S_R(f)$ and $H_2(f)$ binarizes $S_I(f)$.

Let the optimal CTMF [maximizing $|C|$ in Eq. (11)] result in a correlation output phase of β^*. The proof of the optimality of $H_{\beta^*}(f)$ is provided in the appendix, where we show that $H_1(f)$ and $H_2(f)$ in Eqs. (12) and (13) result in phase β^* when substituted in Eq. (9). That is, the equations are consistent for the phase angle of the optimal CTMF. For any other β (not corresponding to the maximum), the filter defined by Eqs. (12) and (13) may not result in phase β in Eq. (9). In this case, it can be shown that Eq. (11) is just the real part of the correlation response, which is less than or equal to the magnitude of the correlation response. Thus, our search for optimal CTMFs can be constrained to filters of the form in Eqs. (12) and (13).

The SNR of the CTMF can be obtained from Eqs. (8), (9), (11), and (14) in the form

$$\text{SNR}_{\text{CTMF}} = \frac{\left\{ \int [I_{R_1}(f)|\hat{S}_R(f)| + I_{R_2}(f)|\hat{S}_I(f)|]df \right\}^2}{\int P_n(f)[I_{R_1}(f) + I_{R_2}(f)]df} . \tag{15}$$

It should be noted that the SNR given by Eq. (15) is a function of β and the support function I_{R_1} and I_{R_2}. The problem of optimizing the SNR is then one of searching over the range of admissible β and support regions.

To determine the $H(f) = H_1(f) - jH_2(f)$ that maximizes the SNR in Eq. (8), we can proceed as follows: (1) For each β, find the $H_\beta(f) = H_1(f) - jH_2(f)$ that maximizes $|C|$ in Eq. (11). Let $|C(\beta)|$ be the resulting maximal value. (2) Determine the β^* that maximizes $|C(\beta)|$. The filter function $H_{\beta^*}(f)$ then maximizes the SNR in Eq. (8).

Further, the integrand in the numerator of Eq. (15) can be written as

$$I_{R_1}|S_R\cos\beta + S_I\sin\beta| + I_{R_2}|-S_R\sin\beta + S_I\cos\beta|$$

$$= I_{R_1}|S_R\cos\beta + S_I\sin\beta| + I_{R_2}\left|S_R\cos\left(\beta + \frac{\pi}{2}\right) + S_I\sin\left(\beta + \frac{\pi}{2}\right)\right| . \tag{16}$$

It can be seen from this equation that the range $0 \leq \beta < \pi/2$ covers all possible values of the numerator in Eq. (15). Therefore, the search need only consider this range of values for β.

The maximization of Eq. (15) for general noise spectra can be quite involved. For this reason, we treat the practical and commonly used case of white noise. For white noise $P_n(f) = N_0$, and Eq. (15) becomes

$$\text{SNR}_{\text{CTMF}} = \frac{\left\{ \int [I_{R_1}(f)|\hat{S}_R(f)| + I_{R_2}(f)|\hat{S}_I(f)|]df \right\}^2}{N_0 \int [I_{R_1}(f) + I_{R_2}(f)]df} . \quad (17)$$

We now describe an efficient algorithm to determine the optimal regions of support for fixed β. A search over β completes the optimization of Eq. (17).

4. EFFICIENT ALGORITHM FOR SUPPORT FUNCTION SELECTION

In this section we describe an efficient algorithm to select the support functions I_{R_1} and I_{R_2} to maximize SNR_{CTMF} in Eq. (17). To enable the use of a digital computer, we discretize the SNR_{CTMF} expression as

$$\text{SNR}_{\text{CTMF}} = \alpha \frac{\left[\sum_{i \in R_1} |\hat{S}_R(i)| + \sum_{j \in R_2} |\hat{S}_I(j)| \right]^2}{N_0(K_1 + K_2)} , \quad (18)$$

where α is a constant that depends on the sampling interval Δf, K_i denotes the number of samples in R_i, $i = 1,2$, and $\hat{S}_R(i)$ and $\hat{S}_I(j)$ in Eq. (18) correspond to $\hat{S}_R(i\Delta f)$ and $\hat{S}_I(j\Delta f)$. The constant α does not affect the selection of the best R_1 and R_2.

Note that the denominator of Eq. (18) depends only on the areas K_1 and K_2 of the support functions and not on which particular frequencies are included in R_1 and R_2. Thus, for a given K_1 and K_2 we must include the $K = (K_1 + K_2)$ highest values among $|\hat{S}_R(i)|$ and $|\hat{S}_I(i)|$ for $1 \leq i \leq N$. Using these, we can find the highest possible SNR for given K. Then we vary K over all possible values and select the best SNR_{CTMF}.

Let $\hat{S}(i) = \hat{S}R(i) + j\hat{S}_I(i)$, $1 \leq i \leq N$, denote the N-point discrete Fourier transform (DFT). Then the following algorithm outlines the basic steps in identifying the optimal support functions R_1 and R_2.

Step 1. Sort $|\hat{S}_R(i)|$ and $|\hat{S}_I(i)|$ values as

$$|\hat{S}^1| \geq |\hat{S}^2| \geq \ldots \geq |\hat{S}^{2N}| , \quad (19)$$

where $|\hat{S}^i|$ denotes the ith largest value among $|\hat{S}_R(i)|$ and $|\hat{S}_I(i)|$ for $1 \leq i \leq N$.

Step 2. Compute the following running sum:

$$R(K) = \sum_{i=1}^{K} |\hat{S}^i| . \quad (20)$$

Step 3. Increment K from 1 to $2N$ in steps of 1. Then

$$\text{SNR}_{\text{MAX}}(K) = \alpha \frac{R^2(k)}{N_0 K} . \quad (21)$$

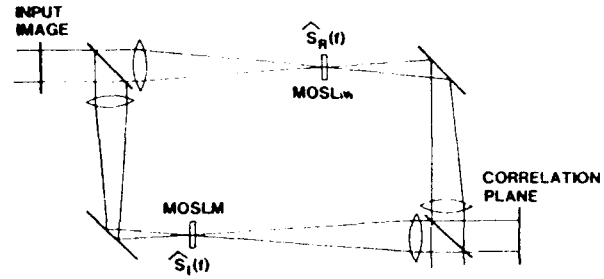*Step 4.* Determine the best $\text{SNR}_{\text{MAX}}(K)$ over all K.

Fig. 1. Mach-Zehnder realization of the CTMF.

Once K is determined, the CTMF is uniquely determined because K_1 and K_2 can be determined from the sorted list in Eq. (19). The real part $H_1(f)$ will be equal to $\text{sgn}[\hat{S}_R(f)]$ for the spatial frequencies with the K_1 highest $|\hat{S}_R(i)|$ values and $H_2(f)$ will be equal to $\text{sgn}[\hat{S}_I(f)]$ for the frequencies with the K_2 highest $|\hat{S}_I(i)|$ values.

5. OPTICAL IMPLEMENTATION

The CTMF of Eq. (7) can be implemented using the architectures suggested earlier,[14,15] for the QPOF. We show the more illustrative Mach-Zehnder architecture in Fig. 1 schematically. This is basically a two-path system. In each path, we do spatial filtering using a magneto-optic spatial light modulator. The MOSLM in the top path implements $H_1(f)$, whereas the one in the bottom path implements $H_2(f)$. We can obtain the CTMF $H_1(f) - jH_2(f)$ either by placing a 90° phase shifter in the bottom path or by making sure that the two paths differ in their length by $\lambda/4$, where λ is the wavelength of the coherent source. The light waveforms from the two paths are then interferometrically detected to yield the output correlation. Other types of filters can be obtained by using the path length difference and the fraction of light in the two legs as additional degrees of freedom. We can replace the two paths in Fig. 1 by a single path if we can get a SLM capable of yielding nine possible complex transmittances $(0, 1, -1, j, -j, 1+j, 1-j, -1+j,$ and $-1-j)$. However, this is not possible with real-time devices. The detour phase method, discussed in detail in Ref. 15, can be used to implement the CTMF in a single device. In this approach, alternate MOSLM elements (in one dimension) are used to represent $\hat{S}_R$ and $\hat{S}_I$, respectively, with a corresponding loss in space-bandwidth product. The detour phase implementation is mechanically simpler and more stable.[15]

5.1. Peak bifurcation

A problem often noticed[9] in the use of BPOFs is that the correlation outputs had two peaks instead of one. Such extraneous peaks must be eliminated before we can use the correlator for target location. Some explanations for this phenomenon of "double peaks" have been put forth and methods have have been proposed for its reduction.[17,22] We believe[17,23] that the reason for producing double peaks with BPOFs is the following. When designing BPOFs, we must first convert the complex matched filter $S_R(f) - jS_I(f)$ into a real function $[\cos\beta S_R(f) + \sin\beta S_I(f)]$ before binarizing it to obtain a BPOF. Here β refers to the threshold line angle available as a variable in BPOF design.[3,4] This conversion of complex functions to real functions imposes an artificial symmetry on the impulse response of the BPOF that can cause the peak bifurcation in the correlation plane. On the
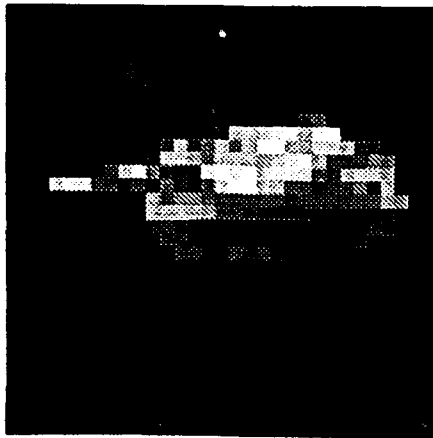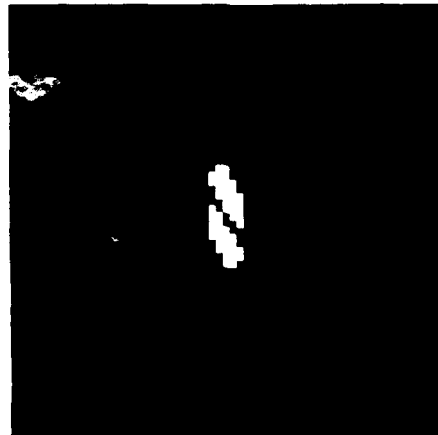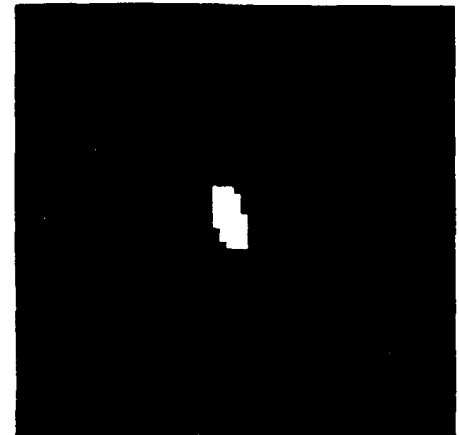
(a)

(b)

Fig. 2. 32 × 32 tank image used in the simulations.

Fig. 3. Support functions (over 64 × 64 arrays) for sgn[$\hat{S}_R(f)$] and sgn[$\hat{S}_I(f)$] obtained for β = 0 CTMF. Black is opaque; white is transparent: (a) Imaginary mask; (b) real mask.

TABLE I. Output SNRs from various filters.

Input SNR (dB)	Matched Filter	POF	OPOF	BPOF-Re	OBPOF-Re	BPOF-Im	OBPOF-Im	QPOF	CTMF
-20	10.76	5.69	9.37	4.13	7.63	4.02	5.15	5.45	8.76
-10	20.64	14.77	19.28	12.23	17.42	10.64	15.11	14.62	18.67
0	30.60	24.73	29.02	22.07	27.24	20.76	25.11	24.59	28.63
10	40.59	34.72	38.98	32.02	37.15	30.79	35.10	34.58	38.64
20	50.58	44.72	48.92	42.00	47.27	40.81	45.10	44.57	48.63

other hand, the CTMF proposed here retains the complex nature of the original matched filter and will break the simple symmetry of the BPOF. Thus, we do not expect to see double peaks in the correlation plane. We must, however, caution that the filter function can have higher-order symmetries.

6. SIMULATION RESULTS

We carried out computer simulations to investigate the advantages of using CTMFs. As the reference image of interest, we chose the tank image shown in Fig. 2. This image is of size 32 × 32, with each pixel having 8 bits of gray-scale resolution. This image was placed in a 64 × 64 array prior to carrying out a 64 × 64 fast Fourier transform (FFT) to determine |$\hat{S}_R(K)$| and |$\hat{S}_I(K)$| values for $1 \leq K \leq 64^2$. The origin of the 64 × 64 array was made to coincide with the centroid of the image in Fig. 2 so that the reference image is centered at the origin. We used zero padding rather than padding by average value as was done elsewhere[24] since we want to get linear correlation, not circular correlation. The image energy is normalized to 1 and has 60.26% of this energy in the even part and the remaining 39.74% in odd part.

The efficient algorithm outlined in Sec. 5 was then applied to the 64 × 64 array obtained after the FFT was carried out. The optimal support functions obtained for sgn[$\hat{S}_R(f)$] and sgn[$\hat{S}_I(f)$] are shown in Fig. 3 for $\beta = 0$. Here, the transparent regions indicate the spatial frequencies for which the filters are nonzero. These figures are plotted such that the center of the array cor-

responds to the (0,0) spatial frequency. Note from Fig. 3 that mostly low spatial frequencies (which have most signal energy) are being allowed through. This can cause two problems. First, the resulting light efficiency will be reduced. However, since the open regions in the frequency planes correspond to regions in which the signal has most energy, the reduction should not be too great. Second, this filter may not do a good job of discrimination because signal high frequencies are being totally suppressed. However, once we realize that CTMF was designed to yield the highest SNR, we must accept the deficiencies that come with the best SNR. However, if best discrimination between two signals $S_1(x)$ and $S_2(x)$ is of interest, we must match our filter to $S(x) = |S_1(x) - S_2(x)|$ and investigate its discrimination performance.

To investigate the sensitivity of CTMF to noise in the input, we added zero-mean, white Gaussian noise of variance σ^2 to the reference image. This noise was generated using standard random number generators. Since the input image of size 32 × 32 has a total energy of 1, each pixel on the average has energy (1/1024). When the noise pixels have a variance of σ^2, the input SNR is defined as

$$SNR_{in} = 10\log_{10}\left(\frac{1}{1024\sigma^2}\right) \tag{24}$$

In Table I, we show the output SNRs obtained using various filters for five different input SNR values. The output SNRs were calculated by collecting the correlation output values (at

TABLE II. CTMF output SNRs as a function of β (input SNR = 0 dB).

β (deg.)	SNR (dB)	β (deg.)	SNR (dB)
0	28.63	50	28.75
5	28.65	55	28.69
10	28.69	60	28.63
15	28.70	65	28.52
20	28.55	70	28.55
25	28.52	75	28.70
30	28.63	80	28.69
35	28.69	85	28.65
40	28.75	90	28.63
45	28.81		

the origin) for 200 noise realizations with the same variance. The output SNR is then given as the ratio of the average of the 200 values to their standard deviation.

In Table I we list the output SNRs obtained using the matched filter, the conventional phase-only filter, the optimal phase-only filter, optimized and unoptimized binary phase-only filters (BPOFs) using the real part as well as the imaginary part of $S(f)$, the quad-phase-only filter, and the proposed complex ternary matched filter with $β = 0$. Obviously, the MF provides the highest output SNRs among all of the filters. Also, there is a 10 dB improvement in the output SNR for all filters with a 10 dB improvement in the input SNR. These results indicate a processing gain of about 30 dB for the matched filter. This agrees with the fact that the image has unit energy and noise variance of (1/1024) (for $SNR_{IN} = 0$ dB), which should result in an output SNR of 1/(1/1024) = 1024 or about 30 dB.

The conventional POF provides output SNRs about 5 to 6 dB lower than those of the matched filter, whereas the optimized POF yields SNRs about 1.5 dB below those of the matched filter. Between the two BPOFs, the BPOF using the real part performs better, perhaps due to the fact that the signal has more energy in its even part than in its odd part. Note that the optimized BPOF using the real part yields output SNRs about 4 to 5 dB below those of the MF. The QPOF results in output SNRs about 6 dB below those of the MF. Finally, the CTMF (for $β = 0$) proposed in this paper yields SNRs about 2 dB below those of the matched filter. Thus, the SNRs obtainable from a CTMF (which requires two binary SLMs) seem to be very close to the SNRs obtainable from optimal POFs (which require a SLM capable of accommodating a continuum of phase values).

The algorithm was next used to compute the output SNR for increments of β of 5° and an input SNR of 0 dB. The results are presented in Table II. From the table, it can be seen that the maximum SNR is obtained for $β = 45°$. However, the peak value exceeds that for $β = 0$, the natural binarization, by only 0.18 dB in this case. We expect that the $β = 0$ binarization will generally give near-optimum SNRs for real object functions. It should be noted that as discussed following Eq. (14), Eqs. (12) and (13) define a filter for arbitrary β and the numerator of Eq. (15) is just the real part of the filter response. Thus, the values in Table II are conservative and the maximum value is the actual value of the optimal filter response.

The SNRs discussed so far are concerned with only one point in the correlation plane (namely, its origin). In Fig. 4, we show the output correlation planes (of size 64 × 64) obtained using MFs, POFs, and CTMFs when there is no noise in the input
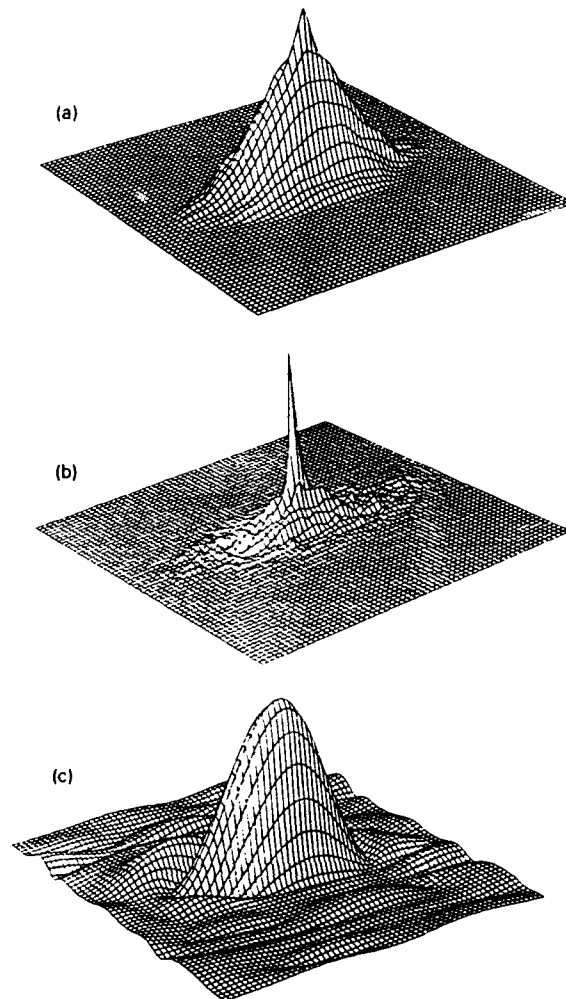
Fig. 4. Complete normalized (peak value set to 1) 64 × 64 correlation outputs for no noise in the input. (a) Matched filter; (b) conventional POF; (c) CTMF.

plane. All correlation peaks are normalized to the same height in these plots since the absolute values are not important. Note that the conventional POF yields the sharpest correlation peaks, whereas the CTMF yields correlation peaks that are broad. In Fig. 5, we show the three correlation plots when the input SNR is 0 dB. The conventional POF exhibits most variation in the correlation plane, whereas the CTMF seems to achieve tolerance to input noise by allowing for broad correlation peaks. Also, notice that there are no double peaks in the output correlation plane when we use the CTMF. To contrast this, we show in Fig. 6 the output correlation when the filter is the optimized BPOF (from the imaginary part) and when there is no input noise. This clearly displays two equally high correlation peaks. These simulations have clearly demonstrated the advantages of the CTMF.

7. CONCLUSIONS

In this paper, we introduced a new spatial filter, the complex ternary matched filter, and presented the theoretical analysis required for maximizing its output signal-to-noise ratio. We also
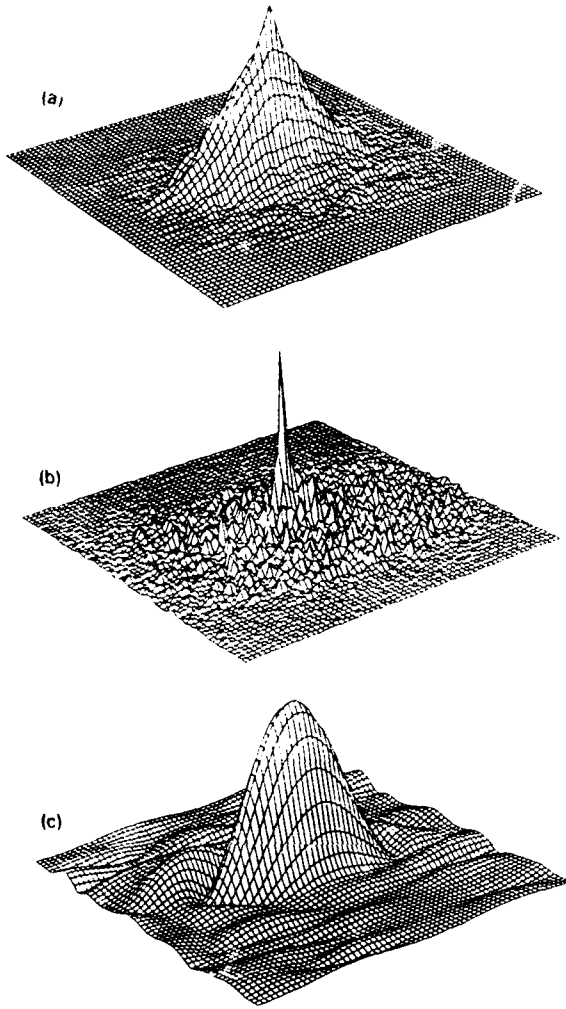
(a)

(b)

(c)

Fig. 5. Complete normalized 64×64 correlation outputs for input SNR = 0 dB. (a) Matched filter; (b) conventional POF; (c) CTMF.

presented an efficient algorithm for determining the CTMF regions of support that lead to the best output SNR. With the help of simulation examples, we have shown that for the particular case studied, the CTMF yields SNRs that are within 2 dB of the SNRs provided by the matched filter. Also, the CTMF does not exhibit symmetry effects such as double peaking.

Based on this, we believe that the CTMF provides a practical method for obtaining SNRs close to those of the matched filters while using only ternary programmable spatial light modulators in the filter plane. While it may be of theoretical interest to go beyond using two binary SLMs, we feel that our results indicate that the CTMFs offer very close to the best performance obtainable and thus any additional improvements may not warrant the complexity.
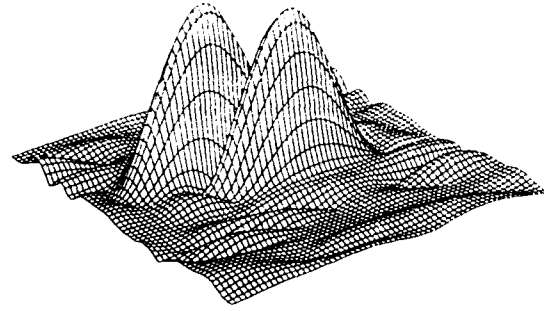
8. ACKNOWLEDGMENTS

Fig. 6. Complete normalized 64×64 correlation output when there is no input noise and filter is an optimized imaginary part BPOF.

9. APPENDIX

For fixed R_1 and R_2, the denominator of the SNR expression in Eq. (8) is unaffected by the choice of $H(f) = H_1(f) - jH_2(f)$. Thus, we focus only on maximizing the numerator, or equivalently, the expression for $|C|$ in Eq. (11). Let β^* correspond to the filter $H_\beta^*(f)$ that maximizes $|C(\beta)|$ and let the resulting maximum be $|C(\beta^*)|$. Then,

$$|C(\beta^*)| = \int [\hat{S}_{\beta^* R}(f)H_{\beta^* 1}(f) + \hat{S}_{\beta^* I}(f)H_{\beta^* 2}(f)]df , \qquad (A1)$$

where the subscripts β^* are included to denote the explicit dependence of $\hat{S}(f)$ and $H_\beta^*(f)$ on it. It is obvious from Eqs. (12) and (13) that the following filters maximize the right-hand side of Eq. (A1):

$$H_1(f) = \begin{cases} \text{sgn}[\hat{S}_{\beta^* R}(f)] , & \text{for } f \in R_1 , \\ 0 , & \text{otherwise} , \end{cases} \qquad (A2)$$

$$H_2(f) = \begin{cases} \text{sgn}[\hat{S}_{\beta^* I}(f)] , & \text{for } f \in R_2 , \\ 0 , & \text{otherwise} . \end{cases} \qquad (A3)$$

For the filters in Eqs. (A2) and (A3) to be valid, optimal choices, we must verify that they result in angle β^* when substituted in Eq. (9). To prove that this is true, let us assume that the filters result in angle $\beta_1 \neq \beta^*$ when we use Eq. (9). Let $\Delta = \beta_1 - \beta^*$ and note that, by definition,

$$\text{Re}[\exp(-j\beta^*) \int S(f)H_{\beta^*}(f)df] = |C(\beta^*)| , \qquad (A4)$$

$$\exp(-j\beta^*) \int S(f)H_{\beta^*}(f)df = |\hat{C}|\exp(+j\Delta) . \qquad (A5)$$

The left-hand side of Eq. (A4) is the real part of the complex quantity in the left-hand side of Eq. (A5). Thus, if $\Delta \neq 0$, we have

$$|\hat{C}| > |C(\beta^*)| . \qquad (A6)$$

This is a contradiction since we are assuming that $|C(\beta^*)|$ is the maximum possible value in the numerator of Eq. (9), but if Eq. (A6) holds, we have a larger value.

10. REFERENCES

1. A. Vander' .t, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory 10, 139–145 (1964).
2. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. 23, 812–816 (1984).
3. R. Kallman, "Optimal low noise phase-only and binary phase-only optical correlation filters for threshold detectors," Appl. Opt. 25, 4216–4217 (1986).
4. R. R. Kallman, "Direct construction of phase-only filters," Appl. Opt. 26, 5200–5201 (1987).
5. J. Rosen and J. Shamir, "Distortion invariant pattern recognition with phase filters," Appl. Opt. 26, 2315–2319 (1987).
6. B. V. K. Vijaya Kumar and Z. Bahri, "Phase-only filters with improved signal-to-noise ratios," Appl. Opt. 28, 250–257 (1989).
7. D. Psaltis, E. G. Paek, and S. S. Venkatesh, "Optical image correlation with a binary spatial light modulator," Opt. Eng. 23, 698–704 (1984).
8. J. L. Horner and J. Leger, "Pattern recognition with binary phase-only filters," Appl. Opt. 24, 609–611 (1985).
9. D. M. Cottrell, R. A. Lilly, J. A. Davis, and T. Day, "Optical correlator performance of binary phase-only filters using Fourier and Hartley transforms," Appl. Opt. 26, 3755–3761 (1987).
10. M. W. Farn and J. W. Goodman, "Optimal binary phase-only matched filters," Appl. Opt. 27, 4431–4437 (1988).
11. D. L. Flannery, J. S. Loomis, and M. E. Milkovich, "Transform-ratio ternary phase-amplitude filter formulation for improved correlation discrimination," Appl. Opt. 27, 4079–4083 (1988).
12. B. V. K. Vijaya Kumar and Z. Bahri, "Efficient algorithm for designing ternary-valued filter yielding maximum signal-to-noise ratio," Appl. Opt. 28, 1919–1925 (1989).
13. B. A. Kast, M. K. Giles, S. D. Lindell, and D. L. Flannery, "Implementation of ternary phase-amplitude filters using a magneto-optic spatial light modulator," Appl. Opt. 28, 1044–1046 (1989).
14. F. M. Dickey and B. D. Hansche, "Quad-phase correlation filters for pattern recognition," Appl. Opt. 28, 1611–1613 (1989).
15. B. D. Hansche, J. J. Mason, and F. M. Dickey, "Quad-phase-only filter implementation," Appl. Opt. 28, 4840–4844 (1989).
16. F. M. Dickey, K. T. Stalker, and J. J. Mason, "Bandwidth considerations for binary phase-only filters," Appl. Opt. 27, 3811–3818 (1988).
17. F. M. Dickey, J. J. Mason, and K. T. Stalker, "Analysis of binarized Hartley phase-only filter performance with respect to stochastic noise," Opt. Eng. 28(1), 8–13 (1989).
18. F. M. Dickey and L. A. Romero, "Dual optimality of the phase-only filters," Opt. Lett. 14, 4–5 (1989).
19. M. W. Farn and J. W. Goodman, "Tightest possible bounds on the performance of phase-only, binary phase-only, and N-ary phase-only matched filters," JOSA A, submitted (Jan. 1989).
20. J. Knopp and R. T. Wingerter, "Comparison of binary and continuous phase-only filters with a restricted phase range," in Aerospace Pattern Recognition, Proc. SPIE 1098, 272–280 (1989).
21. J. L. Horner, "Light utilization in optical correlators," Appl. Opt. 21, 4511–4514 (1982).
22. T. H. Barnes, K. Matsuda, and N. Ooyama, "Reduction of false correlations with binary phase-only filters," Appl. Opt. 27, 3785 (1988).
23. B. V. K. Vijaya Kumar, "Signal-to-noise ratio loss in correlators using real filters," Appl. Opt. 28, 3287–88 (1989).
24. J. M. Florence, "Design considerations for phase-only correlation filters," in Optical Information Processing Systems and Architectures, Proc. SPIE 1151, 195–202 (1990).

Fred M. Dickey received his BS and MS degrees in electrical engineering from the University of Missouri-Rolla in 1964 and 1965 and a Ph.D. in electrical engineering from the University of Kansas in 1975. He has worked for the U.S. Naval Research Laboratory, Southwest Research Institute, Boeing Military Airplane Company, Wichita State University, Dynalectron Corp. and Research and Development Associates and is presently a member of the technical staff at Sandia National Laboratories. His research interests include Fourier optics, optical sensors, optical image and data processing, and optical computing. He has three patents and several technical publications. Dr. Dickey is a member of the Eta Kappa Nu, Tau Beta Pi, OSA, IEEE, and SPIE.

B. V. K. Vijaya Kumar: Biography and photograph appear with the special issue guest editorial in this issue.

Louis Romero received his BS degree from CalTech in 1975, his MS degree from Brown in 1977, and his Ph.D. from CalTech in 1982. All of his degrees are in applied mathematics. He is currently in the Numerical Mathematics Division at Sandia National Laboratories. His interests include fluid mechanics, stability and bifurcation theory, numerical methods, and signal analysis.

James M. Connelly received his BSEE degree from the Pennsylvania State University in 1986 and his MSEE degree from Carnegie Mellon University in 1988. He will receive his Ph.D. from Carnegie Mellon in the fall of 1990. His research topics include synthetic discriminant filter, binary phase only filters, and real time 2-D correlation using 1-D devices. He is a member of IEEE, Tau Beta Phi, and Eta Kappa Nu.

APPENDIX C

Design of phase-only, binary phase-only, and complex ternary matched filters with increased signal-to-noise ratios for colored noise

B. V. K. Vijaya Kumar

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Richard D. Juday

Tracking and Communications Division, NASA Johnson Space Center, Houston, Texas 77058

An algorithm is provided for treating nonwhite additive noise in determining regions of support for phase-only filters, binary phase-only filters, and complex ternary matched filters. It is analytically shown to be optimal in the signal-to-noise ratio sense. It extends earlier research that assumed white noise.

In signal detection in the presence of additive noise, using any known spectral characteristics of the noise is important. It is not enough to have a large response to the signal, if that sole criterion produces a large response to noise also. In this Letter we will consider the signal-to-noise ratio (SNR) as the measure to be optimized. We explicity accommodate colored noise in the filter design.

A practical filter must have a physical implementation, and the available spatial light modulators (SLM's) are restricted in the complex transmittances that they can express. In that context filter optimization is then done with the knowledge of noise spectrum, signal spectrum, and the operational limitations of the SLM. Some subsets of the global problem have been explored. For example, without regard to SLM limitations and if the noise is white, the classical matched filter[1] is known to optimize the SNR. However, this filter requires the use of a complex-valued transmittance in the filter plane, although extant programmable SLM's do not access full regions in the complex plane.

There has been growing interest in replacing classical matched filters by partial information filters that are less demanding of the SLM. For the most part they have not taken noise into account (whether white or nonwhite). They have instead concentrated on the SLM limitations. Primary examples include phase-only filters[2] (POF's), binary POF's[3] (BPOF's), and quad POF's.[4] Since these filters are essentially all pass filters, all the input noise comes through unattenuated, thus deteriorating the output SNR.

Recently we have begun to optimize filters with respect to additive input noise, in the general context of SLM's that are limited to expressing the filters of Refs. 2–4. We suggested that, by making the filter magnitude zero at some spatial frequencies, we can significantly improve the SNR. The set of frequencies for which the filter magnitude is not zero is known as the region of support (ROS). We have previously derived algorithms to determine the optimal ROS's (in the sense of maximizing the SNR with white noise) for POF's,[5] BPOF's,[6] and a modification of quad POF's that we call the complex ternary matched filters[7] (CTMF's). In this Letter we derive algorithms for determining the optimal ROS's when the noise is colored.

We will use one-dimensional notation for convenience. However, all our results can be easily generalized to higher dimensions. Let $s(x)$ denote the target signal and let $n(x)$ be a sample realization from a zero-mean noise process. Then the input signal is assumed to be given by $s(x) + n(x)$. Let $S(u)$ be the one-dimensional Fourier transform of $s(x)$ and let $H(u)$ denote the filter being used in the frequency plane of the optical correlator. Since the input $s(x)$ is centered, we assume that the output peak will appear at the origin, and we will denote this as $c(0)$. Strictly speaking, the correlation is guaranteed to peak at the origin only for (POF's) and classical matched filters for which the filter phase completely cancels the phase of the Fourier transform of $s(x)$. For BPOF's and CTMF's, the phases do not necessarily cancel, and the peak may not be at the origin. In such cases the SNR defined below is smaller than the maximum possible:

$$\text{SNR} = |E[c(0)]|^2/\text{var}[c(0)]$$

$$= \left| \int S(u)H(u)du \right|^2 \bigg/ \int P_n(u)|H(u)|^2 du, \quad (1)$$

where $E[\cdot]$ and $\text{var}[\cdot]$ denote the expected value and the variance and $P_n(u)$ denotes the power spectral density of the noise. When a ROS R is included in $H(u)$, we can rewrite the SNR as

$$\text{SNR} = \left| \int_R S(u)H(u)du \right|^2 \bigg/ \int_R P_n(u)|H(u)|^2 du. \quad (2)$$

We have previously[5,7] considered the case of white noise $[P_n(u)$ is a constant] and showed how R can be determined to maximize SNR. Recently Flannery[8] proposed a heuristic, iterative method for choosing R to maximize SNR when $P_n(u)$ is not a constant. In this Letter we will derive a rigorous, noniterative algorithm to determine optimal R (with the understanding that, for BPOF's and CTMF's, the SNR at the origin is not necessarily the highest SNR in the correlation plane) for the three types of filter (POF, BPOF, and CTMF).

First, let us consider POF's. It is straightforward to show that $H(u)$ must equal $\exp[-j\phi(u)]$ for all u in R, where $|S(u)|$ and $\phi(u)$ are the magnitude and the phase, respectively, of $S(u)$. Using this in Eq. (2), we can write the maximum SNR obtainable from a POF as

$$\text{SNR}_{\text{POF}}(R) = \left[\int_R |S(u)|du\right]^2 \bigg/ \int_R P_n(u)du. \tag{3}$$

In order to carry out the ROS selection on a digital computer, we discretize the above expression as

$$\text{SNR}_{\text{POF}}(R) = \left(\sum_{k\in R} |S_k|\right)^2 \bigg/ \sum_{k\in R} P_{nk}, \tag{4}$$

where $S_k = S(k\Delta u)$ and $P_{nk} = P_n(k\Delta u)$ with Δu being the sampling interval. Also, we have omitted some constant terms in Eq. (4) because they do not affect the selection of the optimal ROS. For colored noise, the following theorem proved elsewhere[9] becomes useful for the determination of the optimal R:

Theorem: Let $x_i, y_i > 0$ for $i = 1, 2, \ldots, N$. Let R denote a subset of integers $\{1, 2, \ldots, N\}$. Let the subscripts be such that

$$z_i = x_i/y_i, \qquad i = 1, 2, \ldots, N \tag{5}$$

form a descending sequence; i.e.,

$$z_1 \geq z_2 \geq z_3 \geq \ldots \geq z_N > 0. \tag{6}$$

Then, if an integer n is included in the optimal subset R^* maximizing the ratio

$$\eta(R) = \left(\sum_{i\in R} x_i\right)^2 \bigg/ \sum_{i\in R} y_i, \tag{7}$$

all integers $k \leq n$ must also be included in R^*.

Comparing Eqs. (4) and (7), we see that $\eta(R) = \text{SNR}_{\text{POF}}(R)$ if $x_i = |S_i|$ and $y_i = P_{ni}$. This suggests that the optimal ROS must be of the following form:

$$R^*_{\text{POF}} = \{i : (|S_i|/P_{ni}) \geq T\}, \tag{8}$$

where T is a threshold to be determined. We will explain the resulting algorithm with the help of an example Suppose that we perform a 64 × 64 fast Fourier transform on the reference image in order to obtain 4096 $|S_i|$ values. Then we divide these by 4096 P_{ni} values. These 4096 ratios are then sorted. Different thresholds T are used, the resulting R^*_{POF} and SNR_{POF} are computed, and the highest SNR among these is selected to yield R^*. Our theorem above proves that the optimal R must be of the form

of Eq. (8). We will need to test at most 4096 thresholds in this example. Usually, we will need fewer than 4096 thresholds. Suppose that we use eight bits to represent the ratios $|S_i|/P_{ni}, i = 1, 2, \ldots, 4096$. Then we need to test only 256 thresholds since there are only 256 different ratios possible. If we use more bits for representing these ratios, we will need more thresholds.

Similar results can be derived for BPOF's and CTMF's also. The BPOF with a ROS is really a three-valued filter. It has been shown elsewhere[10] that the optimal BPOF must be of the following form:

$$H(u) = \begin{cases} \text{sgn}[S_{\beta R}(u)] & u \in R \\ 0 & u \notin R \end{cases}, \tag{9}$$

where $\text{sgn}[x]$ is $+1$ if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. Also $S_{\beta R}(u)$ is the real part of

$$S_\beta(u) = S(u)e^{-j\beta}. \tag{10}$$

The β in Eq. (10) is the same as the phase of the complex $c(0)$ at the origin of the correlation. The resulting SNR is given as follows:

$$\text{SNR}_{\text{BPOF}}(R) = \left[\int_R |S_{\beta R}(u)|du\right]^2 \bigg/ \int_R P_n(u)du. \tag{11}$$

Comparing Eq. (11) with Eq. (3), we see that they are similar. Thus, using $|S_{\beta R}(u)|$ in place of $|S(u)|$ in the algorithm presented above, we can obtain the optimum ROS. However, unlike the POF case, this must be repeated for all β values (in the range 0–$\pi/2$), and the best $\text{SNR}_{\text{BPOF}}(R)$ must be used. It can be shown[7] that the resulting BPOF does indeed produce a consistent correlation phase at the output.

Finally, we consider CTMF's defined as

$$H_{\text{CTMF}}(u) = H_1(u) - jH_2(u), \tag{12}$$

where both $H_1(u)$ and $H_2(u)$ can take on three values $(+1, 0, \text{and} -1)$. The optimal CTMF is given[7] by

$$H_1(u) = \begin{cases} \text{sgn}[S_{\beta R}(u)] & u \in R_1 \\ 0 & u \notin R_1 \end{cases}, \tag{13}$$

$$H_2(u) = \begin{cases} \text{sgn}[S_{\beta I}(u)] & u \in R_2 \\ 0 & u \notin R_2 \end{cases}. \tag{14}$$

Here $S_{\beta R}(u)$ and $S_{\beta I}(u)$ are the real part and the imaginary part, respectively, of $S_\beta(u)$ defined in Eq. (10). Note also that the CTMF is characterized by two ROS's, R_1 for the real part and R_2 for the imaginary part. The resulting SNR is given as follows:

$$\text{SNR}_{\text{CTMF}}(R_1, R_2) = \frac{\left[\int_{R_1} |S_{\beta R}(u)|du + \int_{R_2} |S_{\beta I}(u)|du\right]^2}{\int_{R_1} P_n(u)du + \int_{R_2} P_n(u)du}. \tag{15}$$

It may not be obvious initially how Eq. (15) can be made to look like Eq. (7). However, it is possible.

For this effect, let us define the following sequences of length $2N$ (8192 in our example).

$$x_i = |S_{\mu R}(i\Delta u)|,$$

$$x_{i+N} = |S_{\mu I}(i\Delta u)|,$$

$$y_i = P_n(i\Delta u),$$

$$y_{i+N} = P_n(i\Delta u), \qquad i = 1, 2, \ldots, N. \qquad (16)$$

Using these definitions, we can determine R^*_{CTMF} by

$$R^*_{CTMF} = \{i : (x_i/y_i) \geq T\}, \qquad (17)$$

where T is a threshold to be determined. Since the index i in Eqs. (16) can take on values from 1 to $2N$, all integers from 1 to N in R^*_{CTMF} correspond to R^*_1, and the remaining integers from $(N + 1)$ to $2N$ in R^*_{CTMF} correspond to R^*_2. This completes the determination of the optimal ROS for the CTMF.

In summary, we have presented rigorous algorithms for designing the optimal ROS's for POF's, BPOF's, and CTMF's in the presence of colored noise. These algorithms involve the sorting of ratios (of signal Fourier-transform-related quantities to noise power spectral densities) and the testing of all possible thresholds (the number of thresholds depends on the number of bits used to express the ratios) to find the best ROS.

References

1. A. VanderLugt, IEEE Trans. Inf. Theory **IT-10**, 139 (1964).
2. J. L. Horner and P. D. Gianino, Appl. Opt. **21**, 812 (1984).
3. D. Psaltis, E. G. Paek, and S. S. Venkatesh, Opt. Eng. **23**, 698 (1984).
4. F. M. Dickey and B. D. Hansche, Appl. Opt. **28**, 1611 (1989).
5. B. V. K. Vijaya Kumar and Z. Bahri, Appl. Opt. **28**, 250 (1989).
6. B. V. K. Vijaya Kumar and Z. Bahri, Appl. Opt. **28**, 1919 (1989).
7. F. M. Dickey, B. V. K. Vijaya Kumar, L. A. Romero, and J. M. Connelly, Opt. Eng. **29**, 994 (1990).
8. D. Flannery, Proc. Soc. Photo-Opt. Instrum. Eng. **1297**, 194 (1990).
9. B. V. K. Vijaya Kumar, W. Shi, and C. Hendrix, Opt. Comput. Process. **1**, 29 (1991).
10. M. W. Farn and J. W. Goodman, Appl. Opt. **27**, 4431 (1988).

APPENDIX D

Saturated filters

B. V. K. Vijaya Kumar

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Richard D. Juday

Johnson Space Center, Houston, Texas 77058

P. Karivaratha Rajan

Tennessee Technological University, Cookeville, Tennessee 38505

Most previous research into the design of correlation filters considered only input noise and filter spatial light modulators (SLM's) of an implicitly assumed infinite contrast ratio. We introduce a signal-to-noise ratio that also includes correlation-detector noise and finite contrast SLM's. Filters maximizing this signal-to-noise ratio exhibit saturation at some frequencies and are called saturated filters. We accommodate SLM's whose amplitude has a finite maximum and a nonzero minimum. We give algorithms for optimum saturated complex- and real-valued filters. Previous results are reproduced as various limiting cases. The phase-only filter and the binary phase-only filter are limiting cases for large detector noise with, respectively, complex and real modulators.

1. INTRODUCTION

Noise is always present in the correlation-detection process, even if its major component is no more than the quantization of the detected correlation. Nonetheless, there is little to be found in the literature regarding the effects of that detection noise on the performance of an optical correlator. All else being equivalent and if all noise were additive in the input plane, multiplying the filter by a constant would have no effect on the relative amounts of signal and noise in the correlation plane. (We will say that filters have the same shape if they differ from each other only by a possibly complex constant factor. The magnitude of that factor will be called their relative size.) Increasing the size of a filter seems likely to produce an improvement in the detected correlation. Thus it is common practice in optical-filter computation to scale a calculated filter so that its maximum magnitude is just at the saturation level. (We will say that a filter is saturated if magnitude limitations have changed the shape of the filter.) The shape of the marginally saturated filter is not modified; and if there is no detection noise, the signal-to-noise ratio (SNR) is not affected. In the presence of detector noise, though, the SNR can be improved on by getting an increase in the power to the correlation detector. The price is a loss of spectral matching between the filter and the signal. We adapt our metric of filter optimization to include the noise in correlation detection, and we quantitatively balance the loss of signal-detection selectivity with the increase in power to the correlation detector.

An ideal passive-filter spatial light modulator (SLM) would be controllable so as to transmit any specified magnitude between zero and unity. In this paper we scale the noise in the correlation-detection plane so that we may consider the maximum transmittance as unity; this will have no effect, of course, on the zero at the other end of the control range. In reality, programmable SLM's have finite dynamic range, so there is a finite minimum transmittance. The practical effect is that all the frequency plane passes noise that is additive in the input plane, if only at that minimum transmittance. The effect on the SNR is obviously detrimental. We include minimum transmittance when calculating optimum saturated filters.

A phase-only filter (POF) is fully saturated *ab initio*. The POF was introduced by Horner and Gianino[1] to improve the light-throughput efficiency of classical matched filters. However, the all-pass nature of the conventional POF's results in extreme noise sensitivity.[2] One way to improve the noise tolerance is to set the filter magnitude at select frequencies to zero. An algorithm for determining the region of support (i.e., the set of frequencies for which the filter magnitude is nonzero) for these optimal POF's was presented by Vijaya Kumar and Bahri for continuous[3] and binarized[4] POF's with additive white input noise. More recently, we presented theory[5] and simulation,[6] including additive colored input noise, in determining the optimal regions of support for POF's, binary POF's, and complex ternary matched filters.[7] For simple detection problems, (i.e., when we need to detect the presence or absence of a *known* reference signal in additive noise), the POF optimal region of support was observed typically to consist of mostly low spatial frequencies. Such small areas of support end up blocking much of the light that is incident upon the frequency plane, thus leading once again to a low level of light throughput. Owing to the optimal POF's passing predominantly low frequencies when recognizing objects containing high frequencies, the correlation peak typically is broad

Vijaya Kumar *et al.*

Low light throughput is not an issue if the detectors placed in the correlation-output plane have zero noise and or arbitrarily large gain. When detectors have limited gain and introduce noise, we must ensure that we send as much light to the detector as possible while providing tolerance to noise in the input scene. For the already saturated POF's, this necessitates increasing the area of the region of support beyond what is the optimal choice from solely an input-noise-tolerance consideration. For other filters (complex,[8] real,[9] coupled[10,11]) that take on zero values without separate consideration of a region of support, the equivalent is to drive more of their area into saturation. In those actions the spectral shape of the filter is changed, with an adverse effect on the SNR if only the input noise is considered. In this paper we will discuss making the trade-off between tolerance to input noise and tolerance to detector noise. We explicitly include the fact that a minimum exists for the realized filter amplitude, which also changes the shape of the filter from the calculated ideal.

The rest of this paper is organized as follows. In Section 2 we introduce a simple SNR model that includes detector-noise effects. In Section 3 we derive analytical expressions for the optimal filters. We show algorithms for calculating the filters, including for determining the portions that are driven into either high or low saturation. In Section 4 we show the limiting forms of the filters for arbitrarily large and small noise and for arbitrarily large and small contrast ratio. Section 5 is a discussion. We do not present numerical simulations, since we show analytically that our filters are optimal with respect to our stated measure of the SNR and because the SNR benefits of the saturation will vary from case to case. Our algorithms will permit an investigator to determine any benefit for his own situation.

2. DETECTOR NOISE IN SIGNAL-TO-NOISE RATIO

Let $s(x)$ denote the reference image (we will use one-dimensional notation for simplicity), and let $S(f)$ denote its Fourier transform. Let $H(f)$ denote the complex transmittance of the filter placed in the frequency plane of a frequency-plane correlator. In the absence of input noise, the resulting correlation output at the origin is given by

$$c(0) = \int S(f)H(f)df, \qquad (1)$$

where the limits of integration are those implied by the bandwidths of $S(f)$ or $H(f)$ (whichever leads to the tighter limit). These implicit integration limits will be observed throughout the paper.

A model for possible uncertainties in the input is the additive noise $n(x)$. We model this as a sample realization from a wide-sense stationary random process with mean μ_n and power spectral density $P_n(f)$. The additive noise $n(x)$ in the input leads to randomness in the output $c(0)$. We can easily show that, when signal $s(x)$ is present in the input,

$$E\{c(0)\} = \mu_n H(0) + \int S(f)H(f)df, \qquad (2)$$

$$\text{var}\{c(0)\} = \int P_n(f)|H(f)|^2 df. \qquad (3)$$

When the input to the correlator is only noise $n(x)$, the output $c(0)$ will be a random variable with mean $\mu_n H(0)$ and the same variance as in Eq. (3). For good detection, we need to separate the two means as much as possible while keeping the variance small. A convenient measure for this is the SNR that is defined below, and we will shortly present our motivation for using it:

$$\text{SNR} = \frac{\left| \int S(f)H(f)df \right|^2}{\int P_n(f)|H(f)|^2 df}. \qquad (4)$$

In the past, optimal choices of $H(f)$ [in the sense of maximizing the SNR in Eq. (4)] have been found within the constraint that the filter be phase only,[3] binary phase only,[4] or real.[9] Since a constant factor times the magnitude of H does not affect the SNR in Eq. (4), these optimizations did not consider the finite nature of modulator amplitude. The numerator has been maximized if magnitude and phase are cross coupled.[10,11]

When the correlation output $c(0)$ is detected by a photodetector, several things happen. Detectors respond only to $|c|^2$, thus ignoring all phase information. Also, the detectors introduce a gain and some noise. An accurate model for detector noise is complicated and must include the signal-dependent nature of detector noise. Instead, we use the following simple model for y, the detector output:

$$y = c(0) + n_d. \qquad (5)$$

In this detector-noise model, we assumed that the detector gain is unity and that the detector noise n_d is additive, without loss of generality. The noise n_d is assumed to have mean μ_d and variance σ_d^2. We assume that the detector-noise characteristics have included within them the scale factors that are appropriate to the correlator whose optimal filter is being calculated. For example, the power of the correlator's coherent light source might directly multiply the amplitude of the encoded signal transform, so that increasing the light source power would increase the SNR if σ_d^2 included only local detector noise. Instead of treating those considerations explicitly, we wrap them into the definition of detector noise.

The additive-detector-noise assumption is somewhat questionable. However, it makes the analysis tractable and helps us to illustrate our main point (i.e., that we must trade off input-noise tolerance for detector-noise tolerance).

We will now find the mean and the variance of y in Eq. (5) for the two possible input cases. We will assume throughout that the input noise and the detection noise are statistically independent.

When the input contains only noise $n(x)$ (hypothesis H_0), the mean and the variance of y are as below. Note that H_0 and $H(0)$ have different meanings. The former is the null hypothesis, and the latter is the filter at zero frequency:

$$E\{y|H_0\} = \mu_d + E\{c(0)|H_0\}$$

$$= \mu_d + \mu_n H(0), \tag{6}$$

$$\text{var}\{y|H_0\} = \sigma_d^2 + \text{var}\{c(0)|H_0\}$$

$$= \sigma_d^2 + \int P_n(f)|H(f)|^2 df. \tag{7}$$

When the input contains signal $s(x)$ corrupted by additive noise $n(x)$ (hypothesis H_1), the mean and the variance of y are

$$E\{y|H_1\} = \mu_d + \mu_n H(0) + \int S(f)H(f)df, \tag{8}$$

$$\text{var}\{y|H_1\} = \sigma_d^2 + \int P_n(f)|H(f)|^2 df. \tag{9}$$

Using the statistics in Eqs. (6)–(9), we can express the SNR in the presence of detector noise as

$$\text{SNR} \equiv \frac{|E\{y|H_1\} - E\{y|H_0\}|^2}{1/2(\text{var}\{y|H_1\} + \text{var}\{y|H_0\})}$$

$$= \frac{\left|\int S(f)H(f)df\right|^2}{\sigma_d^2 + \int P_n(f)|H(f)|^2 df}. \tag{10}$$

The only difference between the SNR's in Eqs. (4) and (10) is the extra σ_d^2 in the denominator of Eq. (10). However, this makes the optimal filter choices for the two SNR's different. When σ_d^2 is small [compared with $\int P_n|H(f)|^2 df$], the two SNR's are identical, and previous optimal filters will still be optimal. In another limiting case, when σ_d^2 is large [such that we consider only the σ_d^2 term in the denominator of Eq. (10)], the SNR is proportional to $|E\{c(0)\}|^2$, and we must simply maximize the correlation value at the center. In Section 3 we provide expressions for $H(f)$ that maximize the SNR in Eq. (10) for intermediate cases.

3. ANALYTICAL EXPRESSIONS

We established earlier the equivalence of the discrete and the continuous formulations for the optimal filters [see Eqs. (16) and (28) of Ref. 9], and for convenience we now switch to the discrete representation. $H(f)$ becomes H_k, $P_n(f)$ becomes P_{nk}, the region of integration becomes the obvious summation range, etc.

We consider filters realized on two different types of SLM: fully complex and real. In this section we will determine the set of H_k's that maximizes the SNR in Eq. (10). We will use the amplitude A_k and the phase ϕ_k of the reference-signal transform S_k:

$$S_k = A_k \exp(j\phi_k), \quad 0 \le \phi_k < 2\pi, \quad 0 \le A_k. \tag{11}$$

We will use the magnitude M_k and the phase θ_k of the filter H_k:

$$H_k = M_k \exp(j\theta_k), \quad 0 \le \theta_k < 2\pi, \quad 0 \le M_k \le 1. \tag{12}$$

Even before they are expressly known, we will have repeated use for the amplitude, B, and the phase, β, of the central value of the filtered signal. We will refer to β as the output phase.

$$B \exp(j\beta) = \sum_k S_k H_k, \quad 0 \le \beta < 2\pi, \quad 0 \le B. \tag{13}$$

In these terms, if we let Δf be a uniform sampling interval in the frequency plane, the SNR becomes

$$\text{SNR} = \frac{\Delta f^2 \left|\sum_k S_k H_k\right|^2}{\sigma_d^2 + \Delta f \sum_k P_{nk} M_k^2} = \frac{B^2}{(\sigma_d^2/\Delta f^2) + \sum_k (P_{nk}/\Delta f) M_k^2}. \tag{14}$$

From here on, we will assume that the detector noise and the input noise power spectral density are normalized to the sampling frequency interval as indicated in Eq. (14). In determining the filters, we will explicitly use

$$0 \le D_{\min} \le M_k \le D_{\max} \le 1, \tag{15}$$

expressing the finite limits of the amplitude (*not* the intensity) transmission of a passive filter. One sees that the value of $D_{\max}$ in Eq. (15) can be included with the previously mentioned scale factors in the definition of detector-noise variance. Thus we will use the following equation for the filter-magnitude limitation:

$$\rho \le M_k \le 1, \tag{16}$$

where

$$\rho = \frac{D_{\min}}{D_{\max}}. \tag{17}$$

In this equation ρ is $D_{\min}/D_{\max}$, the reciprocal of the amplitude-contrast ratio for the filter SLM.

Our approach to finding the optimum SNR begins with specifications that produce stationary values for the SNR. If we have a filter that produces a maximum SNR and if we make small changes in the unsaturated portions of the filter, no first-order change in the SNR occurs. The method is explained more fully in our earlier work, where variational calculus[10,11] and partial differential equations[9] are used. The statement of stationarity leads to a search on only a limited set of SNR's, among which exists the globally optimum one.

A. Full Complex Filter

Let us allow H_k to take on any complex value on or within the unit circle. The phase and the amplitude of the filter may be independently determined. Clearly the numerator in Eq. (14) is maximized by matching the phases of reference-signal transform and filter (i.e., their sum is a constant), and the denominator is not affected by the choice of filter phase. So our first action to maximize the complex filter's SNR in Eq. (14) is to choose its phase by

$$\theta_k + \phi_k = \text{const}. \tag{18}$$

Vijaya Kumar et al.

The complex filter's amplitude remains to be found. The selection of the filter's phase by Eq. (18) leaves the now-real-valued equation to optimize for the complex filter's SNR.

$$\text{SNR} = \frac{\left(\sum_k A_k M_k\right)^2}{\sigma_d^2 + \sum_k P_{nk} M_k^2}.$$
(19)

We must choose the set $\{M_k\}$ to maximize the SNR subject to $\rho \leq M_k \leq 1$. If we are examining the optimality of the SNR by the choice of M at the mth frequency, then it must be the case either that M_m is an extremum of its allowed values or that the SNR is stationary with respect to the local choice of M_m. The condition of stationarity is

$$\frac{\partial \text{SNR}}{\partial M_m} = 0.$$
(20)

If this condition produces a value $\rho \leq M_m \leq 1$, we accept it. Otherwise one extreme value or the other is necessary. When we take the partial derivative in Eq. (19),

$$\frac{\partial \text{SNR}}{\partial M_m} =$$

$$\frac{\left(\sigma_d^2 + \sum_k P_{nk} M_k^2\right) 2\left(\sum_k M_k A_k\right) A_m - \left(\sum_k M_k A_k\right)^2 2 M_m P_{nm}}{\left(\sigma_d^2 + \sum_k P_{nk} M_k^2\right)^2}.$$
(21)

and the derivative will be zero if

$$M_m = \frac{A_m}{P_{nm}} \left[\frac{\sigma_d^2 + \sum_k P_{nk} M_k^2}{\sum_k M_k A_k} \right].$$
(22)

The term in parentheses is a constant with respect to the frequency index, since that index is summed over. Thus we may express it as a constant, G. Then,

$$M_m = \frac{A_m}{P_{nm}} G,$$
(23)

and we have defined the optimizing gain, G. In other words, for those frequencies at which the filter is not saturated, the magnitude of the filter is proportional to A/P_n. The gain will cause saturation of the filter at some frequencies. It is crucial to note, however, that the definition of G explicitly includes the participation of the saturated frequencies; it is an equation, not an expression of proportionality. We do not have the freedom to multiply a filter by an arbitrary scalar, as may often have been done in the past, to get a calculated filter somehow to fit within the physical limitations of the modulator on which the filter is to be expressed. If

$$G \frac{A_m}{P_{nm}} > 1,$$
(24)

then it may be easily verified from Eq. (21) that

$$\frac{\partial \text{SNR}}{\partial M_m} > 0 \quad \text{in} \quad 0 \leq M_m \leq 1,$$
(25)

and thus the SNR reaches its maximum (by choice of M_m) at $M_m = 1$. If, conversely,

$$G \frac{A_m}{P_{nm}} < \rho,$$
(26)

then, similarly,

$$\frac{\partial \text{SNR}}{\partial M_m} < 0 \quad \text{in} \quad \rho \leq M_m \leq 1,$$
(27)

and the SNR is maximized by $M_m = \rho$.

Defining a symbol for the saturated value allows a shorter form to be used for M_m. Presuming that $0 \leq a \leq b$,

$$\{\{x\}\}_a^b \equiv \text{sgn}(x) \times \max[a, \min(b, |x|)].$$
(28)

We may now write

$$M_m = \left\{\left\{ G \frac{A_n}{P_{nm}} \right\}\right\}_\rho^1,$$
(29)

where G is yet to be determined. For any candidate value of G, Eq. (14) expresses the realized SNR. We will shortly present the scheme to search on G for the globally optimized filter. For now the important point is that, where the optimal filter is *not* saturated, its magnitude is proportional to A/P_n (the phase has already been given). That the unsaturated portion of the optimal filter holds its shape is certainly a significant point, and it is not obvious.

We now know everything except the globally optimizing value of G. We can obviously limit the search between the values that would just saturate the largest and the smallest values of A/P_n. We now develop our search strategy. We show one method guaranteed to find the globally best gain, and we show another method that may reduce the size of the search.

In finding the globally best gain, we have two routes that we may pursue. The first uses the knowledge that the optimizing gain will just saturate the filter at a particular value of A/P_n. We will call the filter gain selected according to that criterion G_a. The second route requires that, once a gain is known, Eq. (23) also be met if the SNR is an extremum. We will call the gain that is calculated from that equation once a filter is selected G_b.

We use a previous result[4] that showed that, when a filter is optimized by choice of the region of support, all the frequencies having equal values of A/P_n are included as a block. Thus we first sort the N frequencies so that

$$\frac{A_1}{P_{n1}} \geq \dots \geq \frac{A_r}{P_{nr}} > \frac{A_{r+1}}{P_{n,r+1}} \geq \dots \geq \frac{A_{i-1}}{P_{n,i-1}} > \frac{A_i}{P_{ni}} \geq \dots \geq \frac{A_N}{P_{nN}}.$$
(30)

For the moment, r and i are arbitrary, except that each is adjacent to an inequality in the sequence, as indicated. We now pick the rth frequency as the one to be just saturated high; we denote by $G_a = G_a(r)$ the gain that just saturates the rth frequency:

$$G_a = \frac{P_{nr}}{A_r}.$$
(31)

The region R_1 is those frequencies that saturate high (i.e.,

for which $k \leq r$). Similarly R_2 is those frequencies for which saturation does not occur ($r < k < i$), and R_3 is those frequencies that saturate low ($k \geq i$). The ratio ρ is seen to induce the relationship between r and i; i is the smallest index value such that

$$\rho \frac{A_r}{P_{nr}} > \frac{A_i}{P_{ni}}. \tag{32}$$

Of course, if the signal amplitude is so high at all frequencies that the condition in inequality (32) is not met, then no frequencies are saturated low, and R_3 is the null set.

We therefore have the following algorithm for the optimal saturated complex filter:

Given A_k, ϕ_k, P_{nk}, ρ, and σ_d^2,

(1) Form the sorted A_k/P_{nk}; suppose that this takes on L levels.

(2) For each l in $1 \leq l \leq L$, calculate the just-high-saturating G_l as

$$G_l = \frac{P_{nl}}{A_l}. \tag{33}$$

Set the M_{lk} according to

$$M_{lk} = \left\{\left\{ G_l \frac{A_k}{P_{nk}} \right\}\right\}_\rho^1. \tag{34}$$

Calculate the SNR_l according to

$$SNR(G_l) = \frac{\left[\sum_{R_1} A_k + G_l \sum_{R_2} (A_k^2/P_{nk}) + \rho \sum_{R_3} A_k \right]^2}{\sigma_d^2 + \sum_{R_1} P_{nk} + G_l^2 \sum_{R_2} (A_k^2/P_{nk}) + \rho^2 \sum_{R_3} P_{nk}}, \tag{35}$$

where

R_1 is a set of frequencies for which $M(f) = 1$,
R_2 is a set of frequencies for which $\rho < M(f) < 1$,
R_3 is a set of frequencies for which $M(f) = \rho$.

(3) Among the L SNR's, pick the highest. Equations (33) and (34) then give $M_k{}^*$, the amplitude of the optimal constrained complex filter.

(4) Within an arbitrary phase constant, the optimal filter is then $H_k{}^* = M_k{}^* \exp(-j\phi_k)$.

(5) Obviously, in the continuous case, A/P_n may take on a continuum of values. Evaluate the SNR from assuming a sufficiently fine discretization, A_k/P_{nk}, to exhibit its functional effect on the SNR, and determine the optimizing threshold value.

If the gain $G = G_a$ is to produce the optimum SNR, then the requirement on gain shown in Eq. (23) must be met. We denote by G_b the gain that meets Eq. (23) for the gain G_a currently under consideration; if $G_a = G_b$, we have a candidate for the gain of the globally optimized filter. Explicitly showing the partition of saturated and unsaturated filter values, the value G_b that meets the stationary SNR requirement is

$$G_r = \frac{\sigma_d^2 + \sum_R P_{nk} + G_b^2 \sum_{R_2} (A_k^2/P_{nk}) + \rho^2 \sum_R P_{nk}}{\sum_{R_1} A_k + G_b \sum_{R_2} (A_k^2/P_{nk}) + \rho \sum_R A_k}. \tag{36}$$

which easily reduces to

$$G_r = \frac{\sigma_d^2 + \sum_R P_{nk} + \rho^2 \sum_R P_{nk}}{\sum_R A_k + \rho \sum_{R_3} A_k}. \tag{37}$$

There may be a benefit to calculating the optimum filter on the basis of the behavior of G_b as a function of G_a; see the discussion in Section 5. G_b is explicitly needed in a consistency check on the output phase β for the optimum real filter, as follows.

B. Real Filter

As in the case of the complex filter, we assume that the detector noise is so scaled that the maximum amplitude of the filter is unity. From Eq. (12), we can express a real filter on $-1 \leq H \leq 1$ by taking the phase of the filter, θ, to be zero or π and its amplitude to be $\rho \leq M \leq 1$. In calculating the optimal complex filter, in the frequency plane we locally matched the filter's phase to that of the reference object. The real filter does not have that flexibility of phase adjustment. In calculating the optimal real filter, we will see that the equivalent of the complex filter's phase matching is taking the projection between the signal transform's local phase and the output phase of the filtered signal, β, from Eq. (13). In addition to discovering an optimizing gain, G, we must discover the correct angle β. Our development is similar to that in Ref. 9, where more details are presented.

Equation (14) gives the SNR. In the present case H_k is real; thus we do not locally match phase, and our SNR becomes

$$SNR = \frac{\left[\sum_k A_k \exp(+j\phi_k)H_k \right]\left[\sum_l A_l \exp(-\phi_l)H_l \right]}{\sigma_d^2 + \sum_k P_{nk}H_k^2}$$

$$= \frac{B^2}{\sigma_d^2 + \sum_k P_{nk}H_k^2}, \tag{38}$$

where we have used the definition of B and β in Eq. (13). Taking the partial derivative of SNR with respect to H_m, the filter value at the mth frequency, and setting it to zero, we have

$$2H_m P_{nm}B^2 = \left(\sigma_d^2 + \sum_k P_{nk}H_k^2\right)B[A_m \exp j(\phi_m - \beta)$$

$$+ A_m \exp j(\beta - \phi_m)]. \tag{39}$$

This reduces to the form

$$H_m = \left\{\left\{ G \frac{A_m}{P_{nm}} \cos(\phi_m - \beta) \right\}\right\}_\rho^1, \tag{40}$$

in which G is defined similarly as before,

$$G = \frac{\sigma_d^2 + \sum_k P_{nk}H_k^2}{\left| \sum_l A_l \exp(j\phi_l)H_l \right|}. \tag{41}$$

Just as for the complex filter, the two expressions for G

must be consistent. A significant difference arises, however. In Eq. (40) we see both the output phase, β, and the gain, G. As we will see, the saturation effected by a given value of G interacts with the value of β in a manner different from that in Ref. 9.

Again, we begin by sorting frequencies according to expression (30) and search on the L different gains that saturate at the L distinct levels in the sequence. $G_a = G_a(l)$ is taken as the gain that will marginally saturate the lth distinct value of A_k/P_{nk}:

$$G_a = \frac{P_{nl}}{A_l}. \tag{42}$$

As in inequality (32), the chosen value of gain partitions the ordered frequencies into R_1 (saturated high), R_2 (unsaturated), and R_3 (saturated low). Inserting the gain into Eq. (41), we obtain

Then the numerator in Eq. (45) becomes

$$\sum_{R_1} A_k \sin(\phi_k - \beta) + \rho \sum_{R_3} (\text{similar})$$

$$= \sum_{R_1} A_k \frac{1}{2j} \{\exp[j(\phi_k - \beta)] - \exp[-j(\phi_k - \beta)]\}$$

$$+ \rho \sum_{R_3} (\text{similar})$$

$$= \frac{\exp(-j\beta)}{2j} \sum_{R_1} A_k \exp(+j\phi_k) - \frac{\exp(+j\beta)}{2j}$$

$$\times \sum_{R_1} A_k \exp(-j\phi_k) + \rho \sum_{R_3} (\text{similar})$$

$$= \frac{1}{2j} [\exp(-j\beta)E \exp(+j\epsilon) - \exp(+j\beta)E \exp(-j\epsilon)]$$

$$= E \sin(\epsilon - \beta). \tag{50}$$

$$G_b = \frac{\sigma_d^2 + \sum_{R_1} P_{nk} + G_b^2 \sum_{R_2} (A_k^2/P_{nk}) \cos^2(\phi_k - \beta) + \rho^2 \sum_{R_3} P_{nk}}{\exp(-j\beta)\left[\sum_{R_1} A_k \exp(j\phi_k) + G_b \sum_{R_2} (A_k^2/P_{nk})\exp(j\phi_k)\cos(\phi_k - \beta) + \rho \sum_{R_3} A_k \exp(j\phi_k)\right]}, \tag{43}$$

in which we have attained the absolute value in the denominator of Eq. (41) by multiplying by $\exp(-j\beta)$. With some effort we can solve for G_b as a function of β. First we carry the $\exp(-j\beta)$ into the other terms in the denominator. Then, cross multiplying and equating the real parts results in

$$G_b = \frac{\sigma_d^2 + \sum_{R_1} P_{nk} + \rho^2 \sum_{R_3} P_{nk}}{\sum_{R_1} A_k \cos(\phi_k - \beta) + \rho \sum_{R_3} A_k \cos(\phi_k - \beta)}. \tag{44}$$

Equating the imaginary parts produces another equation for G_b:

$$G_b = -\frac{\sum_{R_1} A_k \sin(\phi_k - \beta) + \rho \sum_{R_3} A_k \sin(\phi_k - \beta)}{\sum_{R_2} (A_k^2/P_{nk})\sin(\phi_k - \beta)\cos(\phi_k - \beta)}. \tag{45}$$

We in turn equate these expressions for G_b. By applying well-known trigonometric identities

$$\sin x \cos x = \frac{\exp(+jx) - \exp(-jx)}{2j},$$

$$\sin x \cos x = \frac{\sin 2x}{2}, \tag{46}$$

etc., we will isolate β in two expressions, one for $\sin(2\beta)$ and the other for $\cos(2\beta)$. Their ratio gives $\tan(2\beta)$. First we will require some definitions. We define F^2, E, ϵ, Q^2, and q as functions of the gain G_a and the output phase β being investigated. Recall that G_a and β induce the partition of frequencies into R_1, R_2, and R_3. Let

$$F^2 \equiv \sigma_d^2 + \sum_{R_1} P_{nk} + \rho^2 \sum_{R_3} P_{nk}, \tag{47}$$

$$E \exp(j\epsilon) \equiv \sum_{R_1} A_k \exp(j\phi_k) + \rho \sum_{R_3} A_k \exp(j\phi_k), \tag{48}$$

$$Q^2 \exp(j2q) \equiv \sum_{R_2} \frac{[A_k \exp(j\phi_k)]^2}{P_{nk}}. \tag{49}$$

In like manner, the denominator of Eq. (44) becomes

$$\sum_{R_1} A_k \cos(\phi_k - \beta) + \rho \sum_{R_3} A_k \cos(\phi_k - \beta)$$

$$= E \cos(\epsilon - \beta), \tag{51}$$

and the denominator of Eq. (45) becomes

$$\sum_{R_2} \frac{A^2}{P_n} \sin(\phi_k - \beta)\cos(\phi_k - \beta)$$

$$= \sum_{R_2} \frac{A^2}{P_n} \frac{1}{2} \sin(2\phi_k - 2\beta)$$

$$= \frac{1}{2}\left[\frac{1}{2j} \exp(-2j\beta) \sum_{R_2} \frac{A^2}{P_n} \exp(+2j\phi_k)\right.$$

$$\left. - \frac{1}{2j} \exp(+2j\beta) \sum_{R_2} \frac{A^2}{P_n} \exp(-2j\phi_k)\right]$$

$$= \frac{1}{2} Q^2 \sin(2q - 2\beta). \tag{52}$$

Equating the two expressions for G_b, we have

$$\frac{F^2}{E \cos(\epsilon - \beta)} = -\frac{E \sin(\epsilon - \beta)}{\frac{1}{2} Q^2 \sin(2q - 2\beta)}. \tag{53}$$

Using the trigonometric identities from Eq. (46) and doing some minor algebra, we then have

$$\sin(2\beta)[E^2 \cos(2\epsilon) + F^2 Q^2 \cos(2q)]$$

$$= \cos(2\beta)[E^2 \sin(2\epsilon) + F^2 Q^2 \sin(2q)], \tag{54}$$

and we have this expression for β:

$$\tan 2\beta = \frac{E^2 \sin 2\epsilon + F^2 Q^2 \sin 2q}{E^2 \cos 2\epsilon + F^2 Q^2 \cos 2q}. \tag{55}$$

Unfortunately, although to superficial appearance β has been separated as we accomplished in Ref. 9, we cannot use this equation to calculate β. The difficulty is that β appears in Eq. (40). Its influence follows into the defini-

tions for F, E, etc., in Eqs. (47)–(49) and thence into its own expression in Eq. (55). The utility of Eq. (55) is as a consistency check; the optimum SNR saturated filter will produce the same value for β as was originally used in Eq. (40). [Equation (55) also serves in the limiting analyses in Section 4.] The equivalent consistency check for the gain is that $G_c = G_b$. We are ineluctably led to a two-dimensional search on G and β for the optimum SNR real filter. Except for searching on two parameters rather than one, the order-N algorithm for the optimum real saturated filter is virtually identical to that for the matched filter, with the substitution of the cosine term for phase matching. Having candidate values for G and β, we calculate filters from Eq. (40) and then the SNR from Eq. (38).

We may be able to simplify the search for the optimum value of G by observing G_b from Eq. (44) or (45). When $G_b = G_a$ we have an extremum SNR. See the discussion in Section 5.

4. LIMITING FORMS

Complex; small detector noise, infinite contrast. Consider the effect on Eq. (19) if $\sigma_d^2 = 0$ and $\rho = 0$. The equation becomes

$$\text{SNR} = \frac{\left(\sum_k A_k M_k\right)^2}{\sum_k P_{nk} M_k^2}. \tag{56}$$

Setting to zero the partial derivative of the SNR with respect to M_m,

$$M_m = \left\{ \left\{ \frac{A_m}{P_{nm}} G \right\} \right\}_0^1, \tag{57}$$

we may choose any G, such that

$$G \le \min_k \left[\frac{P_{nk}}{A_k} \right], \tag{58}$$

to hold the shape of the filter. In that case, we have (within an arbitrary phase constant)

$$H_m = G \frac{A_m}{P_{nm}} \exp(-j\phi_m), \tag{59}$$

the classical matched filter.

Complex; small detector noise, finite contrast. If $\rho \ne 0$ but $\sigma_d^2 = 0$, Eq. (19) becomes

$$M_m = \left\{ \left\{ \frac{A_m}{P_{nm}} G \right\} \right\}_\rho^1, \tag{60}$$

where G is from Eq. (37) with $\sigma_d^2 = 0$. As before, we are assured of the optimum SNR if we try all G's having values

$$G_a(r) = \frac{P_{nr}}{A_r}, \tag{61}$$

and finding $G_b = G_a$ may be a shortcut. This is a new result.

Complex; large detector noise, any contrast. Equation (19) becomes

$$\text{SNR} = \frac{\left(\sum_k A_k M_k\right)^2}{\sigma_d^2}, \tag{62}$$

which is maximized by choosing $M_m = 1$ for all k. By Eq. (18) we have

$$H_k = \exp[-j(\phi_k + \text{const.})], \tag{63}$$

the POF.

Complex; any detector noise, unity contrast. Equation (29) shows that all filter magnitudes are unity, and Eq. (18) produces the POF.

Real; no detector noise, infinite contrast. Equation (40) is met by using any choice of G such that

$$G < \min_k \left(\frac{P_{nk}}{A_k} \right). \tag{64}$$

This is the case we developed earlier.[9]

Real; infinite detector noise, any contrast. Consider Eqs. (40) and (41). As σ_d^2 becomes large, it dominates in Eq. (41) and causes M_m, the magnitude of H_m, to saturate high. The sign of H_m is the same as the sign of $\cos(\phi_k - \beta)$. We have a bipolar filter regardless of the value of ρ. The polarization occurs with respect to the output phase β, which we can obtain from Eqs. (40), (41), and (47)–(55) by the following observations. As σ_d^2 becomes large, F^2 becomes large. The first thought is that β would tend to $q \pm n\pi/2$ ($n = 0, 1, 2, \dots$) from Eq. (55):

$$\tan 2\beta = \frac{E^2 \sin 2\epsilon + F^2 Q^2 \sin 2q}{E^2 \cos 2\epsilon + F^2 Q^2 \cos 2q} \xrightarrow[(F^2 \to \infty)]{} \tan 2q. \tag{65}$$

However, as σ_d^2 grows, regions R_2 and R_3 shrink to the null set and Q^2 becomes zero short of an infinite value for σ_d^2. Thus $F^2 Q^2$ also becomes zero short of an infinite value for σ_d^2. Since R_2 and R_3 become null, R_1 becomes the universe of frequencies, Ω, and Eq. (48) becomes

$$E \exp(j\epsilon) \xrightarrow[(\sigma_d^2 \to \infty)]{} \sum_\Omega A_k \exp(j\phi_k). \tag{66}$$

Then we see that the correct limiting value for β is ϵ, the phase of the signal averaged without regard to input noise:

$$\tan 2\beta \xrightarrow[(\sigma_d^2 \to \infty)]{} \tan 2\epsilon \tag{67}$$

or

$$\beta \xrightarrow[(\sigma_d^2 \to \infty)]{} \arg\left[\sum_\Omega A_k \exp(j\phi_k) \right] \pm n\frac{\pi}{2}, \quad n = 0, 1, 2, \dots. \tag{68}$$

The extremum solutions for which n is odd produce minima; compare with the discussion in Ref. 9 in the vicinity of its Eq. (21). Within an inconsequential factor of ± 1, then, we choose $n = 0$, and our limiting-case real filter is

$$H_m = \text{sgn}[\cos(\phi_m - \beta)]. \tag{69}$$

5. DISCUSSION

Practical and programmable SLM's have finite contrast ratios. The effect of a finite contrast ratio is that regions of the frequency plane where little signal energy exists but where noise power is appreciable can nevertheless not be turned entirely off. Most objects that are likely signals to be detected by correlation do not distribute their energy widely in the frequency plane. This situation lends force to our accommodating the modulator's finite contrast ratio when optimizing the SNR of the filter.

The effect of considering detector noise is to push the magnitude of a filter upward into saturation. The shape of the magnitude of the filter is then not matched to that of the signal. One wishes to send more processed light through the filter to dominate the detection noise. We have presented analytic algorithms that balance the SNR loss owing to deteriorated shape matching against increased total-light throughput of a filter in such a fashion as to maximize our stated SNR. We showed that the unrestricted attempt to increase light throughput (viz., the POF) is optimal only in certain limiting conditions.

In the case of the real filter, we have shown that we include the projection of local phase onto the phase of the filtered signal [see the cosine term in Eq. (40)]. We showed that we are generally unable to obtain the value of that filtered phase before calculating the filter itself, as had been done earlier,[9] so a two-dimensional search is necessary.

There is a pleasing symmetry in the forms of Eqs. (31), (37), and (44). G_a is determined as the local noise-to-signal ratio in the frequency plane, and G_b is the global noise-to-signal ratio (including the detector noise) for the noise power and the signal amplitude passed by the saturated parts of the filter. One has the sense that, in retaining the matched filter's shape, the unsaturated portion of the filter contributes the most that it can to noise reduction and that, when we find the correct levels at which to saturate the filter, we have balanced the local and the global noise tolerance. One has the same sense with the real filter, in which we locally weight a filter magnitude according to how well the reference-signal transform's phase lines up with the phase of the reference object as finally filtered. In each of those two cases we do local calculation of filter values so that the global SNR is maximized.

We gave algorithms that are guaranteed to produce the globally optimum saturated filters. The algorithms run in order-N time, where N is the number of frequencies at which to obtain filter values. We suspect (but have not yet shown) that the problem is reducible to at most an order-$\log_2 N$ search. Our reasoning is as follows.

At most N values of G_a are possible. The optimum SNR achievable is stated . a function of G_a without the need to calculate the full-filter values and then the SNR from Eq. (14). The sums in Eq. (35) do not have to be wholly recalculated for each candidate value of G as frequency elements move from one region to another and the sums can be appropriately adjusted. Next we showed how G is calculable straightaway without the necessity of going through the computation of the SNR. The globally optimum choice of G_a is characterized by $G = G_b$, so that equality holds for the G_a under consideration. We have a candidate for the optimum filter. We thus may eliminate the filter and the SNR calculations during the search by observing G_b as a function of G_a. Certainly G is a monotonic function of r (the index of the first high-saturated frequency). If, as we suspect, G_a has the opposite monotonicity, then a binary search on r for $G_b = G_a$ suffices, and we have at most an order-$\log_2 N$ search. We further suspect that, in most practical instances, good behavior of G_a and G_b will further reduce the number of different saturating thresholds that we must evaluate.

ACKNOWLEDGMENTS

REFERENCES

1. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. 23, 812–816 (1984).
2. F. M. Dickey, K. T. Stalker, and J. J. Mason. "Bandwidth considerations for binary phase-only filters," Appl. Opt. 27, 3811–3818 (1988).
3. B. V. K. Vijaya Kumar and Z. Bahri, "Phase-only filters with improved signal-to-noise ratio," Appl. Opt. 28, 250–257 (1989).
4. B. V. K. Vijaya Kumar and Z. Bahri, "Efficient algorithm for designing a ternary-valued filter yielding maximum signal-to-noise ratio," Appl. Opt. 28, 1919–1925 (1989).
5. B. V. K. Vijaya Kumar and R. D. Juday, "Design of phase-only, binary phase-only, and complex ternary matched filters with increased signal-to-noise ratios for colored noise," Opt. Lett. 16, 1025–1027 (1991).
6. B. V. K. Vijaya Kumar, V. Liang, and R. D. Juday, "Optimal phase-only correlation filters in colored scene noise," in Computer and Optically Generated Holographic Optics, I. Cindrich, ed., Proc. Soc. Photo-Opt. Instrum. Eng. 1555, 138–145 (1991).
7. F. M. Dickey, B. V. K. Vijaya Kumar, L. A. Romero, and J. M. Connelly, "Complex ternary matched filters yielding high signal-to-noise ratios," Opt. Eng. 29, 994–1001 (1990).
8. A. VanderLugt, "Signal detection by complex spatial filtering," IEEE Trans. Inf. Theory IT-10, 139–145 (1964).
9. R. D. Juday, B. V. K. Vijaya Kumar, and P. K. Rajan, "Optimal real correlation filters," Appl. Opt. 30, 520–522 (1991).
10. R. D. Juday, "Optical correlation with a cross-coupled spatial light modulator," in Spatial Light Modulators and Applications, Vol. 8 of 1988 OSA Technical Digest Series (Optical Society of America, Washington, D.C., 1988), pp. 238–241.
11. R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross coupling," Appl. Opt. 28, 4865–4869 (1989).

APPENDIX E

Phase-only filters with maximally sharp correlation peaks

B. V. K. Vijaya Kumar, Wei Shi, and Charles Hendrix

Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

A performance measure called the peak-to-correlation energy is used to characterize the sharpness of correlation outputs. This measure is then used to determine the phase-only filter with maximally sharp correlation peaks.

Phase-only filters (POF's), originally introduced by Horner and Gianino,[1] have been the focus of much recent research. We were able to prove[2] that the POF's introduced by Horner and Gianino are indeed optimal in the sense that they yield the maximum signal-to-noise ratio (SNR) of all unit-modulus filters. Also, we can increase the output SNR (by ~3.5 dB in the cases that we tested[2]) by allowing the filter magnitude to be zero at some frequencies. However, the SNR measures the tolerance of the filter to noise in the input image and does not measure the sharpness of the resulting correlation output peaks.

Several measures have been proposed to characterize the correlation-peak sharpness. Although these measures are good at distinguishing sharp peaks from broad peaks, their formulations preclude us from theoretical analyses. The notable exception is the research of Dickey and Romero,[3] who use the special properties of prolate spheroidal functions to show that the POF's introduced by Horner and Gianino are also optimal from considerations of peak sharpness.

In this Letter we introduce a peak sharpness measure that we call peak-to-correlation energy (PCE). The PCE turns out to be a special case of the more general peak sharpness measure introduced in Ref. 3. However, the PCE is much easier to use than the measure described in Ref. 3. First we show, using a simple proof, that the conventional POF maximizes the PCE among all unit-modulus filters. Dickey and Romero[3] proved a similar result, using analysis involving prolate spheroidal functions. We then show that selecting a proper region of support for POF's can increase the PCE significantly.

Let $s(x)$ denote the input image (we use one-dimensional notation for convenience) and $S(f)$ denote its Fourier transform. Let $H(f)$ denote the frequency response of the correlation filter being used. The resulting output, $c(\tau)$, is the inverse Fourier transform of the product $H(f)S(f)$ and is given by

$$c(\tau) = \int_{-\infty}^{\infty} H(f)S(f)\exp(j2\pi f\tau)df. \tag{1}$$

Ideally, $c(\tau)$ should have a large value at $\tau = 0$ and small values elsewhere. A convenient characterization of this attribute is the PCE measure, defined as

$$PCE \triangleq \frac{|c(0)|^2}{\int |c(\tau)|^2 d\tau}, \tag{2}$$

where the limits of integration are from $-\infty$ to $+\infty$. It is easy to see that PCE attains its smallest value of zero when $c(\tau) = c(0)$, a constant for all τ. Conversely, the PCE approaches infinity as $c(\tau)$ approaches a delta function. Thus, as desired, larger PCE values appear to imply sharper correlation peaks. However, the major reason for using the PCE over other available peak sharpness measures is its analytical convenience. Using Parseval's theorem, we can rewrite the PCE measure as

$$PCE = \alpha \frac{\left|\int S(f)H(f)df\right|^2}{\int |S(f)|^2|H(f)|^2 df}, \tag{3}$$

where α is a constant. It is easy to verify that the PCE is maximized if $H(f) = K/S(f)$, an inverse filter. Thus it yields the expected result that the inverse filter produces the sharpest correlation peaks.

Next let us consider all unit-modulus filters, i.e., all filters with $|H(f)| = 1$ for all f. Let $\Phi_S(f)$ and $\Phi_H(f)$ denote the phases of $S(f)$ and $H(f)$, respectively. Then the PCE measured is given by

$$PCE = \alpha \frac{\left|\int |S(f)|\exp\{j[\Phi_S(f) + \Phi_H(f)]\}df\right|^2}{\int |S(f)|^2 df}. \tag{4}$$

The denominator in Eq. (4) is a constant for a given $s(x)$, and the numerator is maximized by using

$$\Phi_H(f) = -\Phi_S(f) + \phi, \tag{5}$$

where ϕ is any constant. Thus the conventional POF yields the maximum PCE among all unit-modulus filters. This result was proved in Ref. 3 by using a more general measure, which required the use of a more complicated analysis involving prolate spheroidal functions. The resulting PCE_{POF} is given by

$$PCE_{POF} = \alpha \frac{\left[\int |S(f)|df\right]^2}{\int |S(f)|^2 df}. \qquad (6)$$

It is also reassuring to note that, like classical matched filters, conventional POF's yield correlation peaks at origin in the absence of noise. The PCE value for the classical matched filter[4] (CMF) is given by

$$PCE_{CMF} = \alpha \frac{\left[\int |S(f)|^2 df\right]^2}{\int |S(f)|^4 df}. \qquad (7)$$

It is interesting to see whether the POF always yields sharper correlation peaks than the matched filter. Toward that end, we form the ratio

$$\eta \triangleq \frac{PCE_{POF}}{PCE_{CMF}}$$

$$= \frac{\left[\int |S(f)|df\right]^2}{\int |S(f)|^2 df} \frac{\int |S(f)|^4 df}{\left[\int |S(f)|^2 df\right]^2}. \qquad (8)$$

It would have been satisfying to show that $\eta \geq 1$ for any $|S(f)|$. However, the first ratio in Eq. (8) is always greater than or equal to 1, whereas the second ratio is always less than or equal to 1. Thus the product may or may not be greater than 1. In the two images that we tested (an image of a missile launcher and an image of an aircraft), η was greater than 1.

As in the case of optimizing[2] SNR's, we can obtain higher PCE values by allowing for some of the filter frequencies to have zero magnitude. Let R denote the set of frequencies for which the filter has unit magnitude. Thus, for frequencies in R, the phase is given by Eq. (5), and the resulting PCE is given as follows:

$$PCE(R) = \alpha \frac{\left[\int_R |S(f)|df\right]^2}{\int_R |S(f)|^2 df}. \qquad (9)$$

We can show that the support region R resulting in the highest PCE can be obtained by including all those frequencies for which $|S(f)|$ is below a threshold T. We provide a proof for this statement in another paper.[5] This threshold T is unknown beforehand. Thus we try many possible T values and find the corresponding support regions and PCE values. Of these, the T value yielding the highest PCE is selected. This yields the optimal PCE POF.

We investigated the advantage of using the optimal PCE POF by using the 32 × 32 image shown in Fig. 1. To ensure that our computer simulations used linear correlations rather than circular correlations, we padded the 32 × 32 image with additional zeros to obtain a 64 × 64 array. A fast Fourier transform of size 64 × 64

is used to obtain the necessary signal Fourier-transform magnitudes. These are then used to determine the optimal support regions. For comparison purposes, we determined the region of support for the optimal SNR POF that maximizes[2] the output SNR. The resulting regions of support are shown in Fig. 2. In both parts of the figure the center represents the origin (zero frequency), the dark regions are where the POF is blocked, and the white regions represent places where the POF is transparent. Note from Fig. 2 that the SNR is optimized by permitting low frequencies for this image, whereas PCE is maximized by blocking low frequencies for this image. The intersection between the two regions of support appears to be empty. Thus optimizing SNR and PCE appear to be conflicting goals.
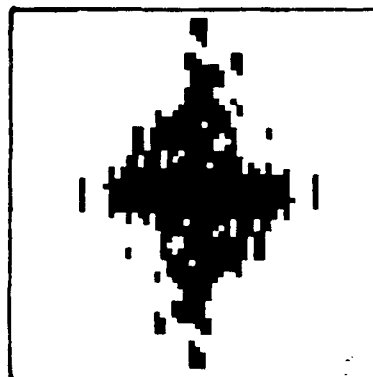
The input image energy is such that a noise variance of 1 at the input is equivalent to having an input SNR of 32.27 dB. The resulting output SNR and PCE values for the CMF,[4] the conventional POF,[1] the opti-

Fig. 1. 32 × 32 image used for simulations.

(a)

(b)

Fig. 2. Regions of support yielding maximal SNR (a) and PCE (b).

Table 1. SNR and PCE Values for Various Filters[a]

Filter Type	SNR (dB)	PCE (dB)
CMF	68.39	16.00
POF	59.01	26.74
Optimal POF	66.60	13.71
Optimal PCE	49.67	33.55

[a] Input SNR, 32.37 dB.

mal SNR POF,[2] and the optimal PCE POF[5] are shown in Table 1. Note that the CMF permits magnitude and phase variations, whereas the other three filters have magnitudes that are either zero or unity. For this image, the conventional POF yields SNR's that are 10 dB lower than those obtained with the classical matched filters. The optimal SNR POF is ~2 dB lower in SNR than the CMF. The optimal PCE POF is almost 19 dB lower than the CMF. However, in comparing PCE values we note that the optimal PCE is the best (33.55 dB), the conventional POF the next best (26.74 dB), the classicial matched filter the next (16.00 dB), and the optimal POF the worst (13.71 dB). We assumed α to be 1 in obtaining the above numbers. By selecting the region of support optimally, we im-

proved the PCE value by almost 7 dB. Another interesting observation is that the classical matched filter outperforms the optimal POF in both SNR and PCE measures. That is perhaps due to permitting both magnitude and phase in the filter plane. Comparing optimal POF and optimal PCE entries, we note that the optimal PCE results in ~20 dB more PCE but 17 dB less SNR. Thus it appears that we have to trade off noise tolerance for peak sharpness.

References

1. J. L. Horner and P. D. Gianino, Appl. Opt. **23**, 812 (1984).
2. B. V. K. Vijaya Kumar and Z. Bahri, Appl. Opt. **28**, 250 (1989).
3. F. M. Dickey and L. A. Romero, Opt. Lett. **14**, 4 (1989).
4. A VanderLugt, IEEE Trans. Inf. Theory **IT-10**, 139 (1964).
5. B. V. K. Vijaya Kumar, C. Hendrix, and W. Shi, "An algorithm for designing phase-only filters with maximally sharp correlation peaks," Proc. Soc. Photo-Opt. Instrum. Eng. **1296** (to be published).

APPENDIX F

Partial information correlation filters with maximally sharp correlation peaks

B. V. K. Vijaya Kumar, Wei Shi and Charles Hendrix

Algorithms are derived for designing three partial information filters (phase-only filters, binary phase-only filters and complex ternary matched filters) that result in maximally sharp correlation peaks. The peak sharpness measure used is the peak-to-correlation energy (PCE), which is the ratio of the square of the correlation peak to the total energy in the correlation output. Several simulation examples are provided to enable a comparison of the correlations obtained by maximizing the PCE and by maximizing the signal-to-noise ratio (SNR).

1. Introduction

While classical matched filters (CMFs) [1] provide the highest output signal-to-noise ratio (SNR), they have not become practical mainly because of the following problems.

(a) CMFs are extremely sensitive to small changes in the reference images [2];

(b) CMFs require spatial light modulators (SLMs) capable of representing complex-valued frequency responses [3];

(c) CMFs attenuate the light because the filter magnitude is less than unity at many frequencies and this results in low light throughput efficiency [4].

Several remedies have been proposed to alleviate these problems. Composite filters [5–7] are suggested for improving tolerance to distortions and phase-only filters (POFs) [8] are proposed to increase light throughput efficiency. Also binary phase-only filters (BPOFs) [9] and complex ternary matched filters (CTMFs) [10] have been proposed as filters suitable for implementation on real-time SLMs such as the magneto-optic SLM (MOSLM) [11].

POF, BPOF, and CTMF share the common feature that all of them attempt to approximate the complex-valued CMF. In that sense, they are *partial information filters*. How well these partial information filters do the job depends on the figure of merit we employ. In the past, we have focused on the maximization of the

output SNR [10, 12, 13]. This led to filters that are maximally robust to input noise.

However, another desirable attribute of the correlation filters is that they result in sharp correlation peaks. In this paper, we use a recently introduced [14] peak sharpness measure called peak-to-correlation energy (PCE) to design POFs, BPOFs, and CTMFs resulting in maximally sharp correlation peaks. We also include some numerical results to illustrate the advantages of using such optimized partial information filters.

The rest of this paper is organized as follows. The next section sets up the notation and provides the necessary background. In section 3, we derive and describe algorithms for designing partial information filters that maximize PCE. These filters are tested and results are presented in section 4. Our conclusions are provided in section 5.

2. Background

Let $s(x)$ denote the signal/image which is being searched for and let $S(u)$ denote its appropriate (1-D for signals and 2-D for images) Fourier transform (FT). We will use 1-D notation throughout for convenience, but all our results apply to higher dimensions also.

2.1. Classical matched filter

The classical matched filter (CMF) for detecting the signal $s(x)$ is given by

$$H_{CMF}(u) = S^*(u)$$
$$= |S(u)| e^{-j\phi(u)}, \qquad (1)$$

where the superscript asterisk denotes the conjugation and $|S(u)|$ and $\phi(u)$ are the magnitude and the phase, respectively, of $S(u)$. It is well known that the CMF yields the highest possible output signal-to-noise ratio SNR), where SNR is defined as below.

$$ \text{SNR} = \frac{|E\{c(0)\}|^2}{\text{var}\{c(0)\}}, \qquad (2) $$

where $c(0)$ denotes the correlation output at the origin in the absence of noise, autocorrelation peaks at the origin) and $E\{\cdot\}$ and $\text{var}\{\cdot\}$ denote the expected value and the variance, respectively.

The CMF in equation (1) is complex-valued and thus requires two spatial light modulators (one for magnitude and one for phase) or use of computer-generated holograms (not attractive for real-time applications). Thus, much effort has been devoted to designing spatial filters suitable for implementation on currently available spatial light modulators (SLMs). Such filters include phase-only filters (POFs), binary phase-only filters (BPOFs) and complex ternary matched filters (CTMFs).

2.2. Phase-only filters

The POFs originally introduced by Horner and Gianino [8] are given by

$$ H_{\text{POF}}(u) = e^{-j\phi(u)}. \qquad (3) $$

Because of their unit magnitude, POFs provide 100% light throughput efficiency [4]. However, this also means that all the input noise comes through unattenuated because of the all-pass nature of the POF [15]. Vijaya Kumar and Bahri [12] showed that the output SNR can be improved if we allow certain filter frequencies to have zero magnitude. Such phase-only filters (which maximize SNR) were called optimal phase-only filters (OPOFs) and can be expressed as below.

$$ H_{\text{OPOF}}(u) = \begin{cases} e^{-j\phi(u)}, & \text{if } u \in R, \\ 0, & \text{if } u \notin R, \end{cases} \qquad (4) $$

where R (called the region of support) indicates which frequencies have unit magnitude in the filter. Vijaya Kumar and Bahri [12] provide an algorithm for determining this region of support.

2.3. Binary phase-only filters

Phase-only filters cannot be implemented on available SLMs such as the magneto-optic SLM (MOSLM) [11] which can be used as either a two-level (-1 and $+1$) or

a three-level ($-1, 0$ and $+1$) device [16]. For this reason, binary POFs (BPOFs) of the following form received much attention.

$$ H_{\text{BPOF}}(u) = \text{sgn}\,[S_R(u)\cos\beta + S_I(u)\sin\beta], \qquad (5) $$

where $S_R(u)$ and $S_I(u)$ are the real part and the imaginary part, respectively, of $S(u)$, $0 \le \beta \le (\pi/2)$ and

$$ \text{sgn}\,[x] = \begin{cases} +1, & \text{if } x \ge 0, \\ -1, & \text{if } x < 0. \end{cases} \qquad (6) $$

Many of the BPOFs proposed are special cases of the general BPOF in equation (5). Using $\beta = 0$ yields the binarization of the real part [9], $\beta = \pi/2$ yields the binarization of the imaginary part [17] and $\beta = \pi/4$ yields the binarization of the sum of the real part and the imaginary part [18] (this sum is the same as the Hartley transform [19] of the original signal). As in the case of POFs, we can increase the achievable output SNR by using the optimal BPOF (OBPOF) defined below.

$$ H_{\text{OBPOF}}(u) = \begin{cases} H_{\text{BPOF}}(u), & \text{if } u \in R, \\ 0, & \text{if } u \notin R. \end{cases} \qquad (7) $$

Vijaya Kumar and Bahri [13] proposed an efficient algorithm for finding R so that the output SNR is maximized. In the examples tested, using an optimal R improved output SNR by about 8 dB, whereas using an optimal β in equation (5) improved the output SNR by less than 2 dB. The OBPOF in equation (7) requires a device capable of 3 levels ($-1, 0$ and $+1$) and thus should be called a ternary filter.

2.4. Complex ternary matched filter

The BPOFs in equations (5) and (7) attempt to represent a *complex* frequency response with a binary or ternary *real* function. This results in certain unwanted artifacts such as peak bifurcation. One solution proposed to avoid these is the quad phase-only filter (QPOF) [20], which accomodates four phase values ($+\pi/4, -\pi/4, +3\pi/4$ and $-3\pi/4$) and unit magnitude. It has been shown [21] that these QPOFs can be implemented with use of only one MOSLM. A powerful generalization of the QPOF is the complex ternary matched filter (CTMF) [10] defined below.

$$ H_{\text{CTMF}}(u) = H_1(u) - jH_2(u), \qquad (8) $$

where $H_i(u)$, $i = 1, 2$ is a ternary ($-1, 0$ and $+1$) filter. Thus each $H_i(u)$ can be considered as a BPOF with its

own region of support R_i. A simple algorithm was presented [10] to identify these regions of support. In the examples tested, CTMF was seen to yield SNRs greater than those obtained using BPOFs, QPOFs, and even POFs. In fact, the CTMF SNR was within 2 dB of the highest possible SNR (that of the CMF).

2.5. *Peak-to-correlation energy*

Much of our past emphasis in designing partial information correlation filters was in obtaining highest possible output SNR, that is in achieving maximum tolerance to additive noise in the input. This is certainly needed. But another desirable attribute of the correlation filters is that they produce sharp peaks in the output correlation plane. Such sharp peaks reduce the false alarm probability by decreasing the probability that the correlation output has larger values elsewhere than at a location corresponding to the true target location.

Several measures have been suggested to quantify the sharpness of the correlation peak. Most of these measures are not conducive to optimization and thus do not help us in designing regions of support to maximize the peak sharpness. To help us in determining the optimal regions of support, we introduced [14] the following peak-to-correlation energy (PCE) measure.

$$PCE \doteq \frac{|c(0)|^2}{\int |c(\tau)|^2 \, d\tau},$$ (9)

where $c(\tau)$ denotes the correlation output. Obviously, PCE is the ratio of the square of the correlation peak (assuming that it occurs at the origin) to the total energy in the correlation plane. For broad outputs such as $c(\tau) = $ constant for all τ, PCE approaches zero. On the other hand, PCE approaches infinity when $c(\tau)$ is a sharp function such as the delta function. This measure was inspired by our earlier success in designing minimum average correlation energy (MACE) filters [22]. PCE is a special case of a more general peak sharpness measure introduced by Dickey and Romero [23], but is a more convenient measure for optimization purposes. For example, it is fairly easy to show that

$$PCE = \frac{|\int S(u)H(u) \, du|^2}{\int |S(u)|^2 |H(u)|^2 \, du}.$$ (10)

It is also easy to show that PCE in equation (10) is maximized if we choose $H(u) = a/S(u)$ for any constant a. This confirms our expectation that the inverse filter (IF) yields the sharpest possible output or equivalently the highest PCE value. Of course, we will not pursue IF

any further for two reasons. First, it is extremely sensitive to noise. Second, its frequency response requires both magnitude and phase variations and thus does not constitute a partial information filter. In the next section, we discuss how various partial information filters can be designed to maximize PCE.

3. Peak-to-correlation energy (PCE) maximization

In this section, we discuss how POFs, BPOFs and CTMFs can be designed to maximize the PCE in equation (10). In all three cases, we allow for filter frequencies to have zero magnitude, that is there are regions of support R associated with these filters. We present efficient algorithms to identify these regions of support that use a proposition proved in appendix 1. Throughout this section, we assume that the input is noise-free since our focus is on obtaining sharp correlation peaks by use of partial information filters.

3.1. *Phase-only filters*

Let us consider phase-only filters of the following form.

$$H(u) = \begin{cases} e^{-j\theta(u)}, & \text{if } u \in R, \\ 0, & \text{if } u \notin R. \end{cases}$$ (11)

Here we have deliberately chosen $-\theta(u)$ instead of $-\phi(u)$ as phase of $H(u)$. The resulting PCE is given as below.

$$PCE = \frac{|\int_R |S(u)| \exp j[\phi(u) - \theta(u)] \, du|^2}{\int_R |S(u)|^2 \, du}.$$ (12)

For a given R, PCE in equation (12) is maximized if and only if

$$\theta(u) = \phi(u) + \rho,$$ (13)

where ρ is any constant. From now on, we will assume without loss of generality that $\rho = 0$. This confirms the earlier results [23] which indicated that the conventional POF is optimal from noise tolerance as well as peak sharpness considerations. When we substitute equation (13) in equation (12), we obtain

$$PCE_{POF} = \frac{|\int_R |S(u)| \, du|^2}{\int_R |S(u)|^2 \, du}.$$ (14)

The next task is to choose R such that PCE_{POF} is maximized with respect to R.

To determine R^*, the region of support that maximizes PCE_{POF} in equation (14), we form the discrete version of equation (14) as below.

$$PCE_{POF}(R) = a \frac{(\Sigma_{i \in R} S_i)^2}{\Sigma_{i \in R} S_i^2}, \qquad (15)$$

where a is a constant that depends on the sampling interval Δu and where

$$S_i = |S(i\Delta u)|. \qquad (16)$$

We will assume from now on that $a = 1$. We will also assume that $1 \le i \le N$. Our objective is to choose the integer subset R from the integer set $\{1, 2, \ldots, N\}$ so that $PCE_{POF}(R)$ in equation (15) is maximized.

The proposition stated and proved in appendix 1 is useful here. Note that $\eta(R)$ in equation (A 1) is identical to the $PCE_{POF}(R)$ in equation (15) if $x_i = S_i$ and $y_i = S_i^2$, $i = 1, 2, \ldots, N$. Then $z_i = x_i/y_i$ equals $1/S_i$ for $i = 1, 2, \ldots, N$. According to the proposition in appendix 1, the optimal subset R^* must include all z_i values larger than a threshold. Since $z_i = 1/S_i$, the optimal region of support R^* for the POF must be as below.

$$R^*_{POF} = \{i : S_i \le T\}, \qquad (17)$$

where T is a threshold to be determined. This suggests the following algorithm for determining the optimal region of support.

POF algorithm

1. Perform the Fourier transform on the reference signal/image and sample it to obtain N samples. Let S_i denote the magnitudes. Let there be M distinct magnitudes $T_1 < T_2 < \cdots < T_M$ in the set $\{S_1, S_2, \ldots, S_N\}$.

2. Construct the region of support $R^*_{POF}(T_i)$ as below for $i = 1, 2, \ldots, M$.

$$R^*_{POF}(T_i) = \{k : S_k \le T_i\}. \qquad (18)$$

The corresponding PCE_{POF} is given by

$$PCE_{POF}(T_i) = \frac{(\Sigma_{k \in R^*_{POF}(T_i)} S_k)^2}{\Sigma_{k \in R^*_{POF}(T_i)} S_k^2}. \qquad (19)$$

3. Choose the highest value from $PCE_{POF}(T_1)$, ..., $PCE_{POF}(T_M)$. Use the corresponding $R^*_{POF}(T_i)$ as the region of support

This algorithm is fairly simple to implement. It basically requires M steps when the number of distinct FT magnitudes is M. Even when the image size or its FT size is large (for example 512 by 512), we can reduce M to reasonable values (such as 256) by quantizing the spectral magnitudes to a few bits (for example 8). In section 4, we will show numerical results obtained using this algorithm.

3.2. *Binary phase-only filters*

Let us consider general BPOFs of the form given below.

$$H(u) = \begin{cases} -1 \text{ or } +1, & \text{if } u \in R, \\ 0, & \text{if } u \notin R. \end{cases} \qquad (20)$$

Strictly speaking, this is a ternary valued filter [24], but we will refer to it as a BPOF with a region of support. Using equation (20) in equation (10), we obtain

$$PCE = \frac{|\int_R S(u) H(u) \, du|^2}{\int_R |S(u)|^2 \, du}, \qquad (21)$$

where the denominator is seen to depend only on the choice of R. Thus, for a fixed R, maximizing PCE in equation (21) is equivalent to maximizing $|c(0)|$, where

$$c(0) = \int_R S(u) H(u) \, du. \qquad (22)$$

If $c(0)$ has magnitude $|c(0)|$ and phase β, that is

$$c(0) = |c(0)| e^{i\beta}, \qquad (23)$$

we can write $|c(0)|$ as below.

$$|c(0)| = \int_R S(u) e^{-i\beta} H(u) \, du$$

$$= \int_R S_\beta(u) H(u) \, du, \qquad (24)$$

where

$$S_\beta(u) = S(u) e^{-i\beta}. \qquad (25)$$

Before we discuss how $|c(0)|$ in equation (24) can be maximized by choosing proper $H(u)$, it is important to reiterate that β is the phase of the correlation output (at the origin) when $H(u)$ is used as the filter. The following discussion about maximizing $|c(0)|$ is the same as the one found in reference [10] and is repeated here for convenience. Since $H(u)$ and $|c(0)|$ are real, equation

(24) is equivalent to the following.

$$|c(0)| = \int_R S_{\beta R}(u) H(u)\, du, \qquad (26)$$

where $S_{\beta R}(u)$ is the real part of $S_\beta(u)$ and is given by

$$S_{\beta R}(u) = S_R(u)\cos\beta + S_I(u)\sin\beta, \qquad (27)$$

with $S_R(u)$ and $S_I(u)$ being the real part and the imaginary part, respectively, of $S(u)$.

Since $H(u)$ takes on either -1 or $+1$ for all $u \in R$, an upperbound can be put on the $|c(0)|$ in equation (26) as shown below.

$$|c(0)| \leq \int_R |S_{\beta R}(u)|\, du, \qquad (28)$$

with the equality satisfied if we use

$$H_{BPOF}(u) = \begin{cases} \text{sgn }[S_{\beta R}(u), & \text{for } u \in R, \\ 0, & \text{for } u \notin R. \end{cases} \qquad (29)$$

However, we have not yet proved that the filter in equation (29) does indeed yield a correlation output with the correct phase β. We prove in appendix 2 that the BPOF that yields the highest value of $|c(0)|$ must be of the form in equation (29) with $\beta = \beta^*$, where β^* is the resulting phase of the output correlation peak, that is, we can assume that the optimal BPOF is of the form in equation (29). The resulting PCE$_{BPOF}$ (R) is given by

$$PCE_{BPOF}(R) = \frac{[\int_R |S_{\beta R}(u)|\, du]^2}{\int_R |S(u)|^2\, du}. \qquad (30)$$

Our next objective is to choose R such that PCE$_{BPOF}(R)$ in equation (30) is maximized. Once again, we form the discrete version of the FT magnitudes, set the irrelevant (at least for optimization purposes) constants to unity and obtain the following.

$$PCE_{BPOF}(R) = \frac{[\Sigma_{k \in R} S_{\beta R k}]^2}{\Sigma_{k \in R} S_k^2}, \qquad (31)$$

where S_k was defined earlier in equation (16) and where

$$S_{\beta R k} = S_{\beta R}(k\Delta u). \qquad (32)$$

Once again, the proposition in appendix 1 is instrumental in determining R^*. PCE$_{BPOF}(R)$ in equation (31) is the same as $\eta(R)$ in equation (45) if we assign $x_i = S_{\beta R i}$ and $y_i = S_i^2$ for $i = 1, 2, \ldots, N$. As before, what matters is the ratio $z_i = x_i/y_i = S_{\beta R i}/S_i^2$. The optimal region of support R^*_{BPOF} must be of the following form.

$$R^*_{BPOF} = \{i : z_i \geq T\}, \qquad (33)$$

where T is a threshold to be determined. Since z_i depends on $S_{\beta R i}$ which in turn depends on β, we must try all possible β values before we determine the best R^*_{BPOF}. However, it can be shown [10] that we need to vary β only in $[0, \pi/2]$. Thus, we try β values in increments of $\Delta\beta$. This leads to the following BPOF algorithm.

BPOF algorithm

1. Start with $\beta = 0$.

2. Compute $S_{\beta R k}, S_k$ for $k = 1, 2, \ldots, N$ using equations (16), (25) and (26). Compute $z_k = S_{\beta R k}/S_k^2$.

3. Determine M, the number of distinct magnitudes in the set $\{z_1, z_2, \ldots, z_N\}$. Denote these as $T_1, T_2, \ldots, T_M$. The number M and the levels T_i can change as β is changed.

4. Determine the optimal region of support as below.

$$R^*_{BPOF}(\beta, T_i) = \{k : z_k \geq T_i\} \qquad (34)$$

Compute the associated PCE$^*_{BPOF}$ from equation (31). Do this for $i = 1, 2, \ldots, M$.

5. Determine the maximum PCE$^*_{BPOF}(\beta, Ti)$ as i is varied from 1 to M. This is the maximum possible PCE$^*_{BPOF}(\beta)$.

6. Increment β by $\Delta\beta$. If β exceeds $\pi/2$, go to step 7; otherwise, go to step 2.

7. Determine the maximum among all PCE$^*_{BPOF}(\beta)$ values. Let β^* be the corresponding angle. Output the associated optimal region of support.

When the above algorithm searches over β values in the range $[0, \pi/2]$, we can identify the best BPOF. When $\beta = 0$ and $\beta = \pi/2$, the resulting BPOFs in equation (29) possess the correct even or odd symmetry and the resulting correlation output $c(0)$ will have phase $\beta = 0$ and $\beta = \pi/2$. Thus for $\beta = 0$ and $\beta = \pi/2$, the filters identified by the above algorithm will indeed yield the proper correlation outputs. We have shown in appendix 2 that the filter yielding the highest PCE value also is consistent in that sense. However, there is no such guarantee for other β values. In section 4, we will show numerical results obtained using this algorithm.

3.3. *Complex ternary matched filters*

In this section, we consider CTMFs of the form in equation (8). The resulting PCE_{CTMF} is given by

$$PCE_{CTMF} = \frac{|\int S(u)H(u)\,du|^2}{\int |S(u)|^2 [H_1^2(u) + H_2^2(u)]\,du}$$

$$= \frac{|\int S(u)H(u)\,du|^2}{\int_{R_1} |S(u)|^2\,du + \int_{R_2} |S(u)|^2\,du}, \quad (35)$$

where $H_1(u) = 0$ for $u \notin R_1$ and $H_2(u) = 0$ for $u \notin R_2$; otherwise, $H_1(u)$ and $H_2(u)$ take on -1 or $+1$. To maximize PCE_{CTMF} in equation (35), we need to maximize $|c(0)|$, where $c(0)$ is as in equations (22) and (23). We can write $|c(0)|$ as

$$|c(0)| = \int S_\beta(u)H(u)\,du$$

$$= \int [S_{\beta R}(u) + jS_{\beta I}(u)][H_1(u) - jH_2(u)]\,du$$

$$= \int_{R_1} S_{\beta R}(u)H_1(u)\,du + \int_{R_2} S_{\beta I}(u)H_2(u)\,du,$$

$$\quad (36)$$

where once again, we utilized the fact that $|c(0)|$ must be real. Obviously, $|c(0)|$ attains the highest possible value (for fixed R_1 and R_2) if we can choose $H_1(u)$ and $H_2(u)$ as below.

$$H_1(u) = \begin{cases} \mathrm{sgn}[S_{\beta R}(u)], & \text{for } u \in R_1, \\ 0, & \text{for } u \notin R_1, \end{cases} \quad (37)$$

and

$$H_2(u) = \begin{cases} \mathrm{sgn}[S_{\beta I}(u)], & \text{for } u \in R_2, \\ 0, & \text{for } u \notin R_2, \end{cases} \quad (38)$$

$S_{\beta R}(u)$ is defined in equation (27) and $S_{\beta I}(u)$ is given as follows.

$$S_{\beta I}(u) = S_I(u)\cos\beta - S_R(u)\sin\beta. \quad (39)$$

Once again, there is no guarantee that $H_1(u)$ and $H_2(u)$ in equations (37) and (38) produce a consistent phase β in the correlation output. However, using the associated symmetries, we can show [10] that for $\beta = 0$ and $\beta = \pi/2$, these filters produce correct phases in the correlation output. We can also show [10] that the CTMF that yields the highest $|c(0)|$ value is of the form in equations (37) and (38) with $\beta = \beta^*$ and it produces a correlation output with phase β^*. We are not repeating

that proof here since it is a generalization of the BPOF proof in appendix 2 and can be found elsewhere [10]. When the optimal CTMF, is used, we obtain

$$PCE^*_{CTMF}(\beta) = \frac{[\int_{R_1} |S_{\beta R}(u)|\,du + \int_{R_2} |S_{\beta I}(u)|\,du]^2}{\int_{R_1} |S(u)|^2\,du + \int_{R_2} |S(u)|^2\,du}. \quad (40)$$

If we take the discrete form of the integrals in equation (40) and set some of the constants to unity, we obtain

$$PCE^*_{CTMF}(\beta) = \frac{[\sum_{k \in R_1} S_{\beta R k} + \sum_{k \in R_2} S_{\beta I k}]^2}{\sum_{k \in R_1} S_k^2 + \sum_{k \in R_2} S_k^2}, \quad (41)$$

where S_k is defined in equation (16) and $S_{\beta R k}$ and $S_{\beta I k}$ are samples of $S_{\beta R}(u)$ and $S_{\beta I}(u)$ for $k = 1, 2, \ldots, N$. At first, equation (41) appears to be very different from the $\eta(R)$ in equation (45). However R_1 and R_2 represent regions of support for two different functions and thus are independent. Let us consider the following sequences x_i and y_i, each of length $2N$. For $i = 1, 2, \ldots, N$ define

$$x_i = S_{\beta R i},$$
$$x_{i+N} = S_{\beta I i},$$
$$y_i = S_i^2,$$
$$y_{i+N} = S_i^2. \quad (42)$$

Similarly, construct a composite region of support R such that integers between 1 and N refer to the contents of R_1 and integers between $(N+1)$ and $2N$ refer to the contents in R_2. Then equation (41) can be rewritten as below.

$$PCE^*_{CTMF}(\beta) = \frac{[\sum_{k \in R} x_k]^2}{\sum_{k \in R} y_k}. \quad (43)$$

Now, we can apply the results of the proposition in appendix 1 which indicates that the optimal R^*_{CTMF} must be of the following form.

$$R^*_{CTMF}(\beta) = \{k : x_k/y_k \geq T\}. \quad (44)$$

As before, we need to vary β in increments of $\Delta\beta$ between 0 and $\pi/2$. This leads to the following CTMF algorithm.

CTMF algorithm

1. Start with $\beta = 0$.

2. Compute $S_{\beta R k}$, $S_{\beta I k}$ and S_k for $k = 1, 2, \ldots, N$ from equations (16), (25), (26) and (39). Determine x_i
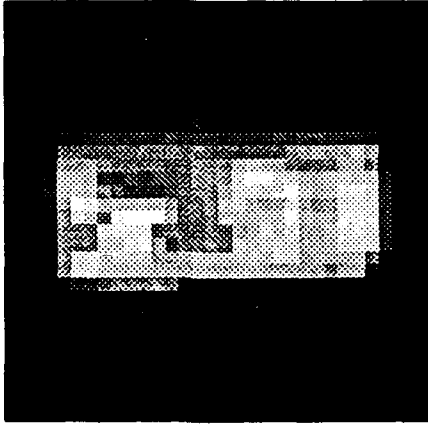
Figure 1. The 32 × 32 image of a missile launcher.

and y_i for $i = 1, 2, \ldots, 2N$ from equation (42). Compute $z_k = x_k / y_k$ for $k = 1, 2, \ldots, 2N$.

3. Determine the number M of distinct magnitudes in the set $\{z_1, z_2, \ldots, z_{2N}\}$. Denote these as $T_1, T_2, \ldots, T_M$.

4. Determine the optimal composite region of support as below.

$$R^*_{\text{CTMF}}(\beta, T_i) = \{k: z_k \geq T_i\}.$$

Identify R^*_1 and R^*_2 associated with the above composite region of support and compute the associated $\text{PCE}^*_{\text{CTMF}}$ from equation (41). Do this for $i = 1, 2, \ldots, M$.

5. Determine the maximum $\text{PCE}^*_{\text{CTMF}}(\beta)$ by varying i from 1 to M.

6. Increment β by $\Delta\beta$. If β exeeds $\pi/2$, go to step 7; otherwise, go to step 2.

7. Determine the maximum $\text{PCE}^*_{\text{CTMF}}(\beta)$. Let β^* be the corresponding angle. Output the associated optimal regions of support.

The same comments as in the BPOF algorithm apply here. We can be sure of the consistency of the filters only for $\beta = 0$, $\beta = \pi/2$ and β^*. We will discuss this issue in some more detail in our next section on numerical results.

4. Numerical results

In this section, we apply the algorithms discussed in section 3 on some sample images to illustrate the improvements possible in correlation filter design. The image used for most of the testing is the 32 × 32 missile launcher image with 256 gray levels shown in figure 1.

Two types of results are shown in this section. The first type, that we refer to as 'numerical evaluations', consists of PCE and SNR values obtained by numerically evaluating various algebraic expressions (such as equations (19), (31) and (41)). The other type of results that we refer to as 'simulation results' are obtained by producing correlation outputs and then estimating SNRs and PCEs from equations (2) and (9). PCE can be obtained from single correlation output. However, SNR estimation requires that we compute expected values and variances and thus we produce many (one hundred in our tests) correlation outputs (using independent input noise realizations) from which we obtain the necessary averages and variances. To avoid getting a circular correlation (wrap-around error), we pad the 32 × 32 input images with zeros to obtain 64 × 64 arrays and perform FFTs of size 64 × 64 on them. When constructing BPOFs and CTMFs, it is important to centre the images properly. We ensure that the centroid of the image coincides with the origin of the array. Finally, when we add noise to the input image, we add it to all 64 × 64 pixels in the array. Otherwise, the resulting noise power spectral density will not be a constant as was assumed in the derivation of OPOF, OBPOF, etc.

4.1. Regions of support

The 64 × 64 FFT of the image in figure 1 is a complex array. In figure 2, we show the magnitude of that array. White regions in figure 2 indicate frequencies with large magnitudes and dark regions indicate low magnitudes. We are including this figure so that the regions of support we show can be properly interpreted. The origin of this array is right at the centre in these figures.

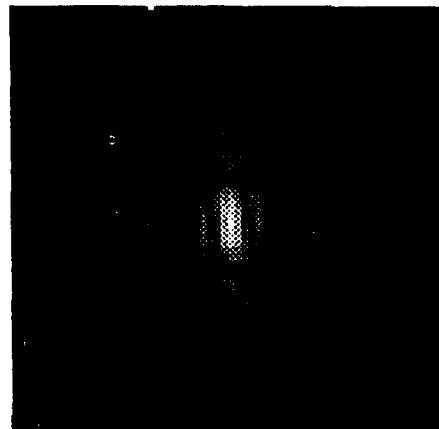We used the algorithms derived earlier [10, 12, 13] to determine regions of support for maximizing the ouput

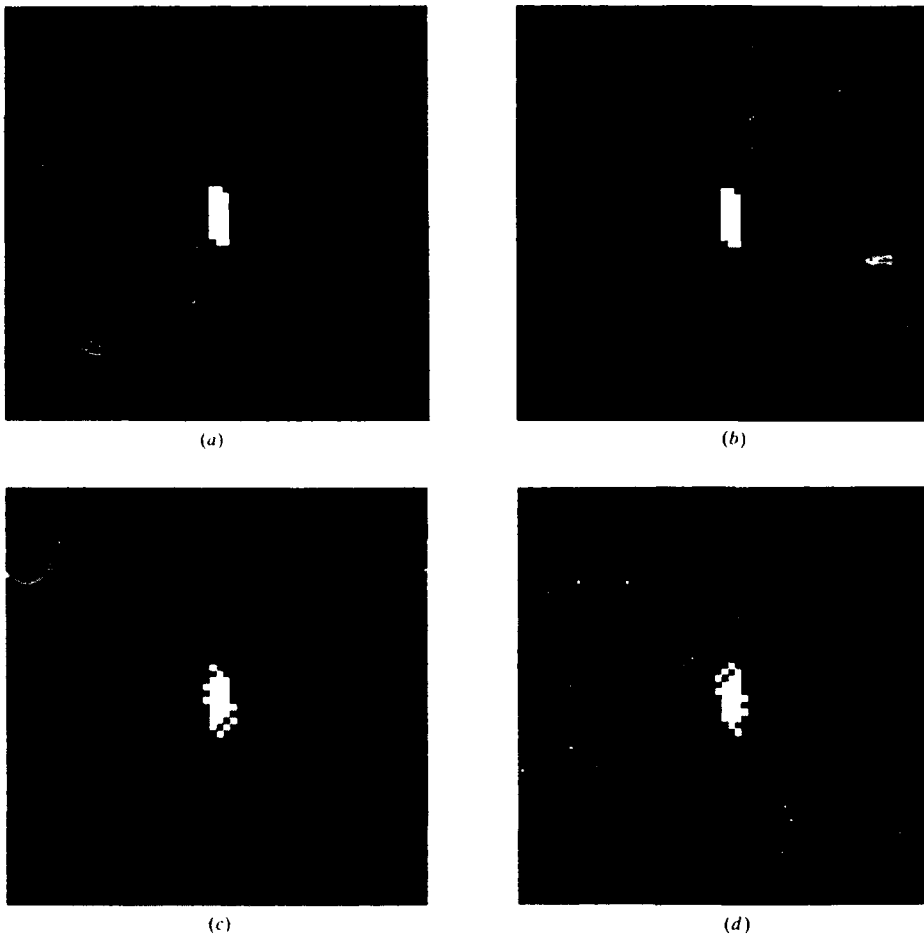Figure 2. The magnitude of the 64 × 64 FFT of the image in figure 1.

Figure 3. Optimal regions of support to maximize output SNR when input noise is additive and white: (a) OPOF, (b) OBPOF (for $\beta = 0°$), (c) region R_1 for CTMF (for $\beta = 45°$), (d) region R_2 for CTMF (for $\beta = 45°$).

SNR when the input noise is additive and white. We show these regions in figure 3. Here white regions indicate where the filter has nonzero magnitude and dark regions indicate where the filter has zero magnitude. Figure 3(a) shows the region of support for OPOF, figure 3(b) for optimal BPOF (for $\beta = 0°$) and figure 3(c) shows R_1 for CTMF (for $\beta = 45°$) and figure 3(d) shows R_2 for CTMF. All these regions allow low frequencies to go through while blocking high frequencies. This is as expected for SNR maximization. However, we could not have predicted the irregular shapes of these optimal regions of support without using the algorithms derived earlier. These filters will not produce sharp correlation peaks because of their low pass nature.

The optimal regions of support obtained by maximizing the PCE values are shown in figure 4. The β value for which we get the highest PCE value when using a BPOF is 90° (not 0°) for this example. The region of support shown in figure 4(b) thus uses $\beta = 90°$. Similarly, the regions of support in figure 4(c, d) for CTMF use a β value of 0°. These regions are very

different from those in figure 3. These emphasize high frequencies. Once again, we could not have predicted the complicated shapes of these regions without the algorithms derived in this paper. Also, note that the OPCE-POF in figure 4(a) is zero for many low frequencies which means that the conventional POF (which is all pass) will result in lower PCE values.

4.2. Output correlations

In this section, let us examine the correlation outputs produced by various filters. Figure 5 shows the correlation outputs obtained using CMF, conventional POF, BPOF (real part binarization), BPOF (imaginary part binarizaton), OPOF, OBPOF, CTMF, OPCE-POF, OPCE-BPOF and OPCE-CTMF. We are including all of these in one place to enable easy comparisons. In all these figures, the input was noise-free. All methods produce peaks at the right location. The filters optimized for SNR produce broad correlation peaks whereas filters designed to optimize PCE produce sharp correlation peaks. Also OPCE-POF produces sharper

(a)

(b)

(c)

(d)
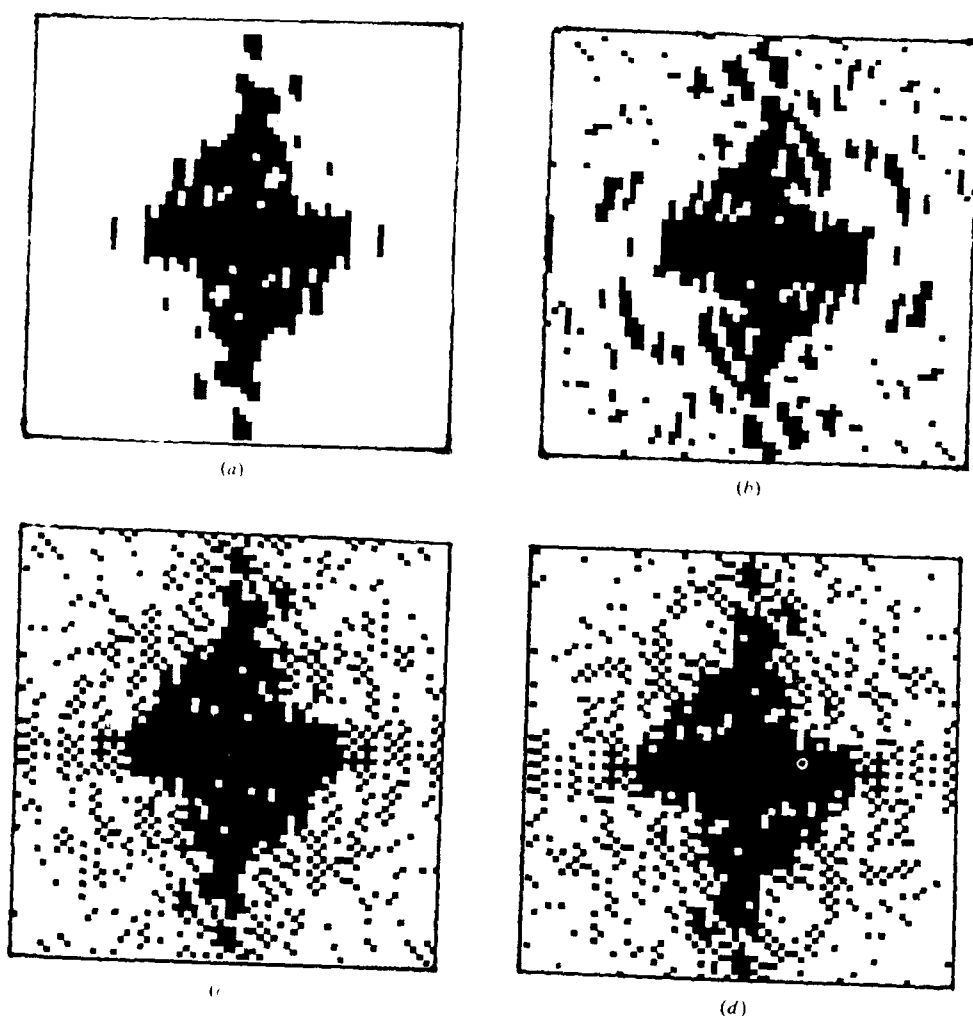
Figure 4. Optimal regions of support to maximize PCE: (a) OPCE-POF, (b) OPCE-BPOF (for $\beta = 90°$) (c) R_1 for OPCE-CTMF (for $\beta = 0°$) (d) R_2 for OPCE-CTMF (for $\beta = 0°$).

peaks than the POF as seen by comparing figure 5(b) and (g). Also, by using OPCE-BPOF and OPCE-CTMF we can use binary or ternary devices in the filter plane and still obtain as sharp output peaks as can be obtained using phase-only filters (see figure 5(g–i).

The above figures describe only part of the story. We show in figure 6 the correlation outputs when the input image was corrupted by zero mean additive white noise. The variance of the noise was selected to be 100 000. This corresponds to an input SNR of −17·73 dB. The input SNR is defined as the ratio of the average energy per pixel in the input image to the variance of the noise. It is obvious that the CMF, OPOF, OBPOF and CTMF are robust to input noise whereas other filters are not. Thus, we have traded away noise tolerance when we maximized PCE. This, we believe, is the unfortunate reality in correlation filter design. We will discuss this issue in more detail in the next section.

4.3. *Comparison of numerical and simulation results*

In this subsection, we present a summary of simulation results we observed for various filters. In table 1, we indicate the SNR and PCE values obtained (using numerical evaluation as well as simulations) for various filters. The numerical results and the simulation results are very close to each other. In fact, for the PCE measure, they are the same since there is no randomness involved in estimating PCE values.

As expected, the CMF provides the highest output SNR; but OPOF and CTMF both produce output SNRs within 2 dB of this. When we realize that CTMF can be implemented with a single MOSLM in the filter plane, we believe that it is now possible to implement highly noise-tolerant correlation filters on commercially available real-time SLMs such as the MOSLM. Partial information filters optimized for peak sharpness yield highly degraded SNRs (about 20 dB lower than that of the CMF). On the other hand, use of OPCE-POF, OPCE-BPOF and OPCE-CTMF improves the PCE by
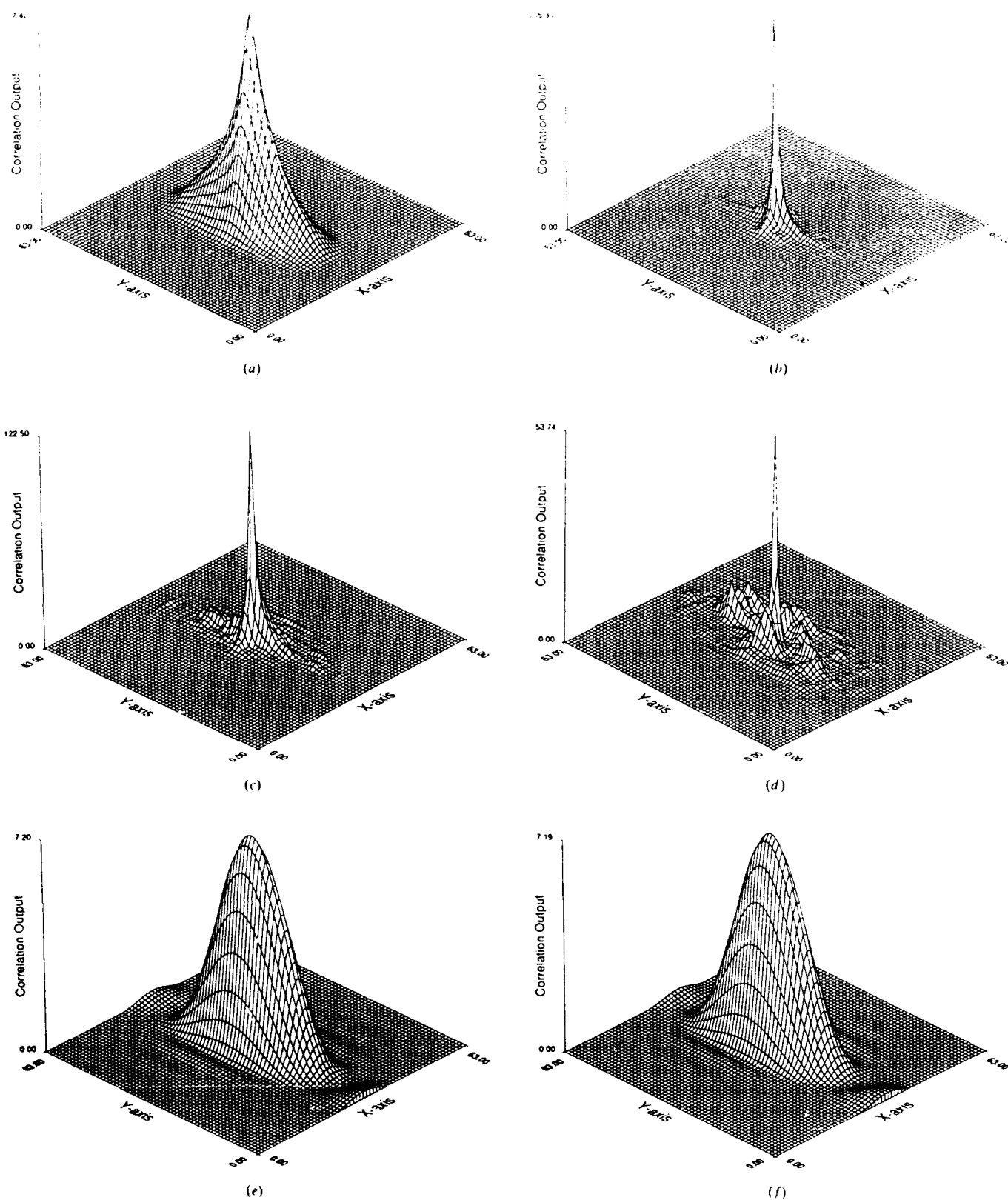
Figure 5. Correlation outputs produced by various filters when input has no noise: (*a*) CMF, (*b*) conventional POF, (*c*) BPOF (real part binarization), (*d*) BPOF (imaginary part binarization), (*e*) OPOF, (*f*) OBPOF, (*g*) CTMF, (*h*) OPCE-POF, (*i*) OPCE-BPOF, (*j*) OPCE-CTMF.
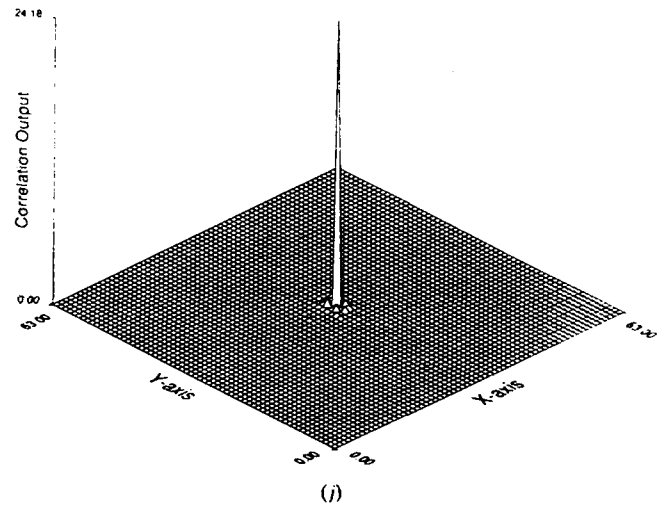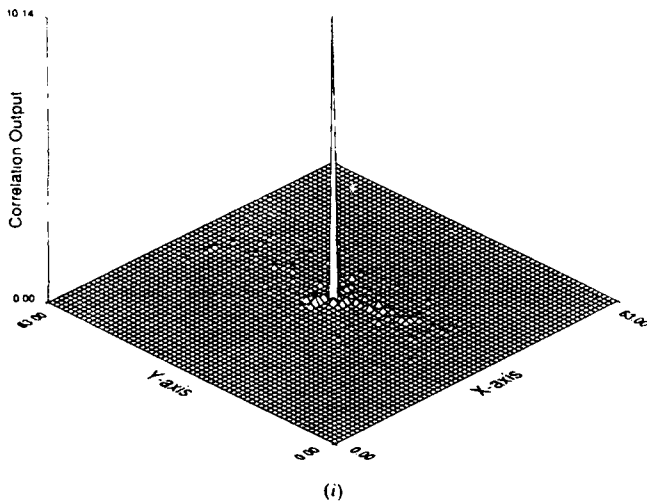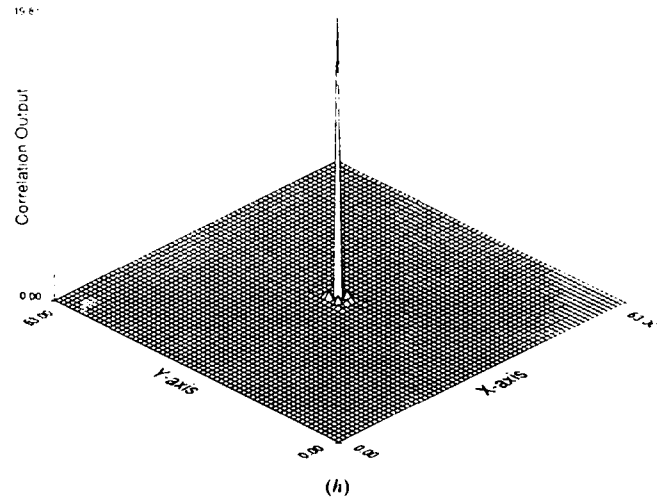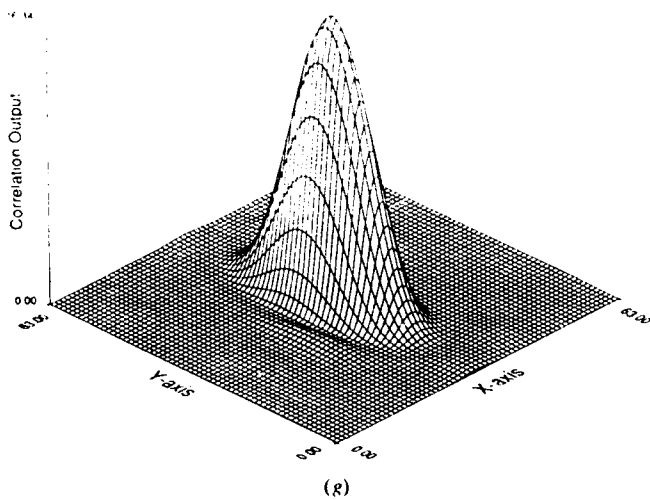
(g)

(h)

(i)

(j)

Figure 5 (continued)

about 15 to 20 dB compared with the CMF. From PCE considerations, OBPOF, CTMF, and OPOF performed the worst. Their PCE values are lower than that of even CMF. Thus, we seem to have conflicting objectives in optimizing SNR and PCE.

4.4. Discrimination performance

To determine whether or not the filters discussed in this paper can discriminate between images, we tested all of the filters (designed for the missile launcher image) on the 32 × 32 image of a tank shown in figure 7. The discussion of these tests is the topic of this subsection.

Table 1. Output SNR and PCE values for various filters for a missile launcher

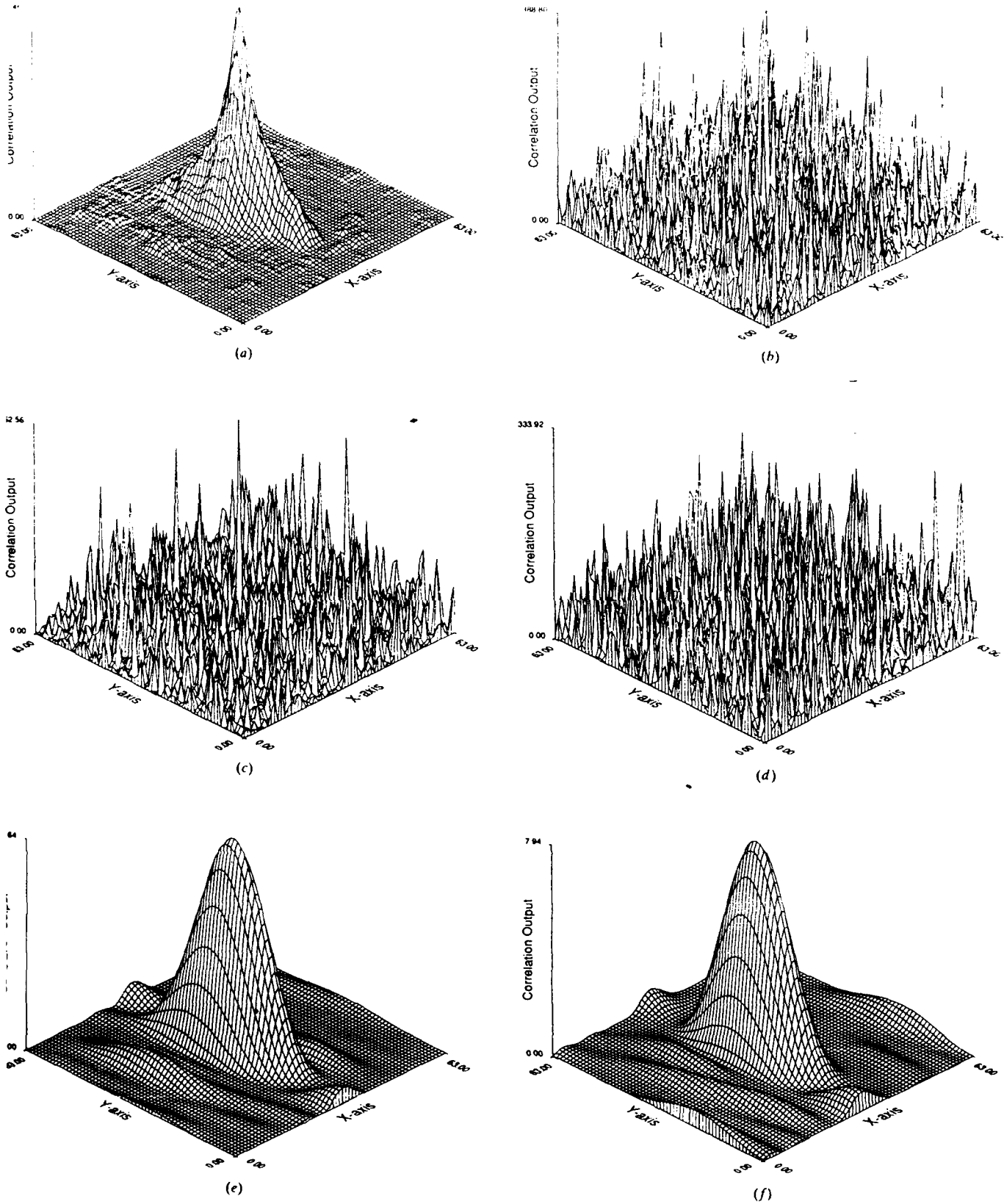| Filter type	SNR (dB)		PCE (dB)	
	Theory	Simulation	Theory	Simulation
CMF	68·63	67·97	16·00	16·00
POF	59·25	59·28	26·74	26·74
BPOF(real)	56·64	56·94	24·13	24·13
BPOF(imag.)	53·42	53·79	20·92	20·92
OPOF	66·84	66·50	13·71	13·71
OBPOF	66·62	66·67	13·49	13·49
CTMF	66·97	67·15	14·47	14·47
OPCE-POF	49·95	49·74	33·54	33·54
OPCE-BPOF	47·56	47·88	30·86	30·86
OPCE-CTMF	48·78	48·29	33·49	33·49

Figure 6. Correlation outputs produced by various filters when the input SNR is −17·73 dB: (*a*) CMF, (*b*) conventional POF, (*c*) BPOF (real part binarization), (*d*) BPOF (imaginary part binarization), (*e*) OPOF, (*f*) OBPOF, (*g*) CTMF, (*h*) OPCE-POF, (*i*) OPCE-BPOF, (*j*) OPCE-CTMF.
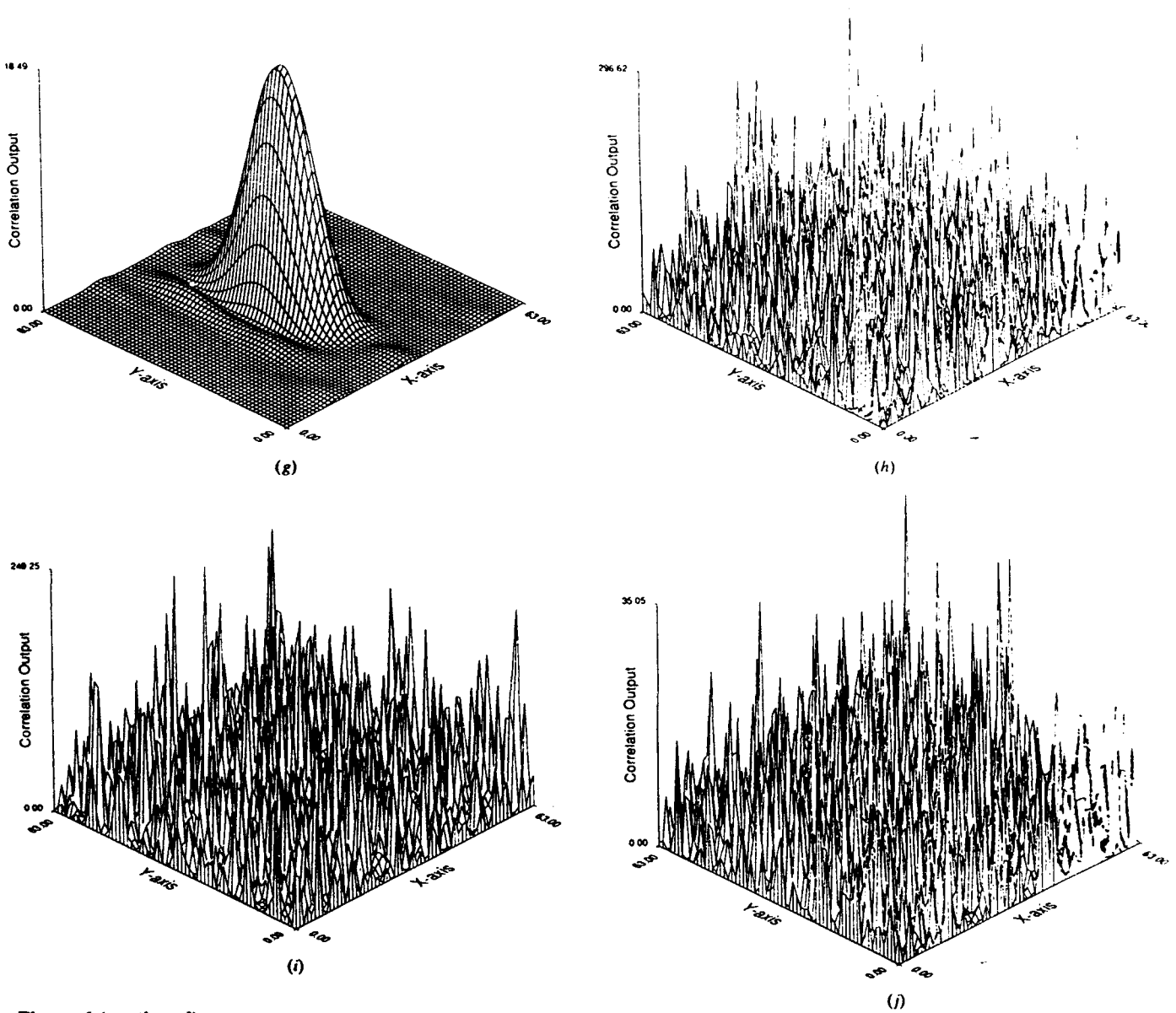
(g)

(h)

(i)

(j)

Figure 6 (continued)

For the simulations, we again zero-padded the image to make it a 64 × 64 array. This image was then corrupted by adding zero-mean gaussian noise with a variance of 1 (the same noise variance as in table 1). For the tank image, the image energy was such that this corresponded to an input SNR of 28·57 dB. In table 2, we show the processing gains provided by various filters (designed with the missile launcher image) when the input image is a missile launcher and when it is a tank. Processing gain (PG) is defined as the difference in dB between the output SNR and the input SNR. Thus PG takes into account the differences in input image energy. We use the difference in the PG for the missile launcher image and the PG for the tank image as a

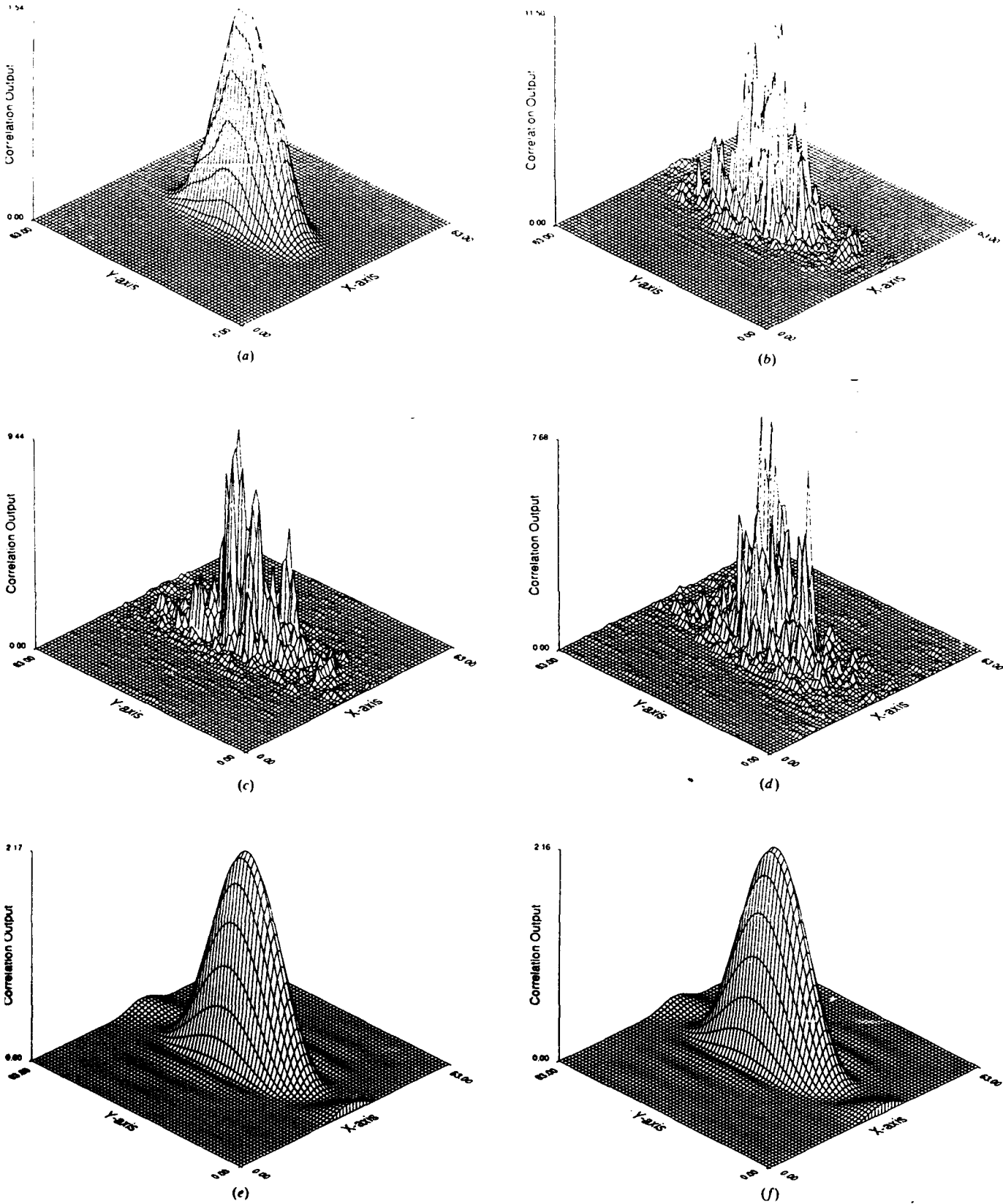**Figure 7.   The 32 × 32 image of a tank.**

42   *B. V. K. Vijaya Kumar* et al.

Figure 8. Correlation outputs produced by various filters for the tank image without noise: (*a*) CMF, (*b*) Conventional POF, (*c*) BPOF (real part binarization), (*d*) BPOF (imaginary part binarization), (*e*) OPOF, (*f*) OBPOF, (*g*) CTMF, (*h*) OPCE-POF, (*i*) OPCE-BPOF, (*j*) OPCE-CTMF.
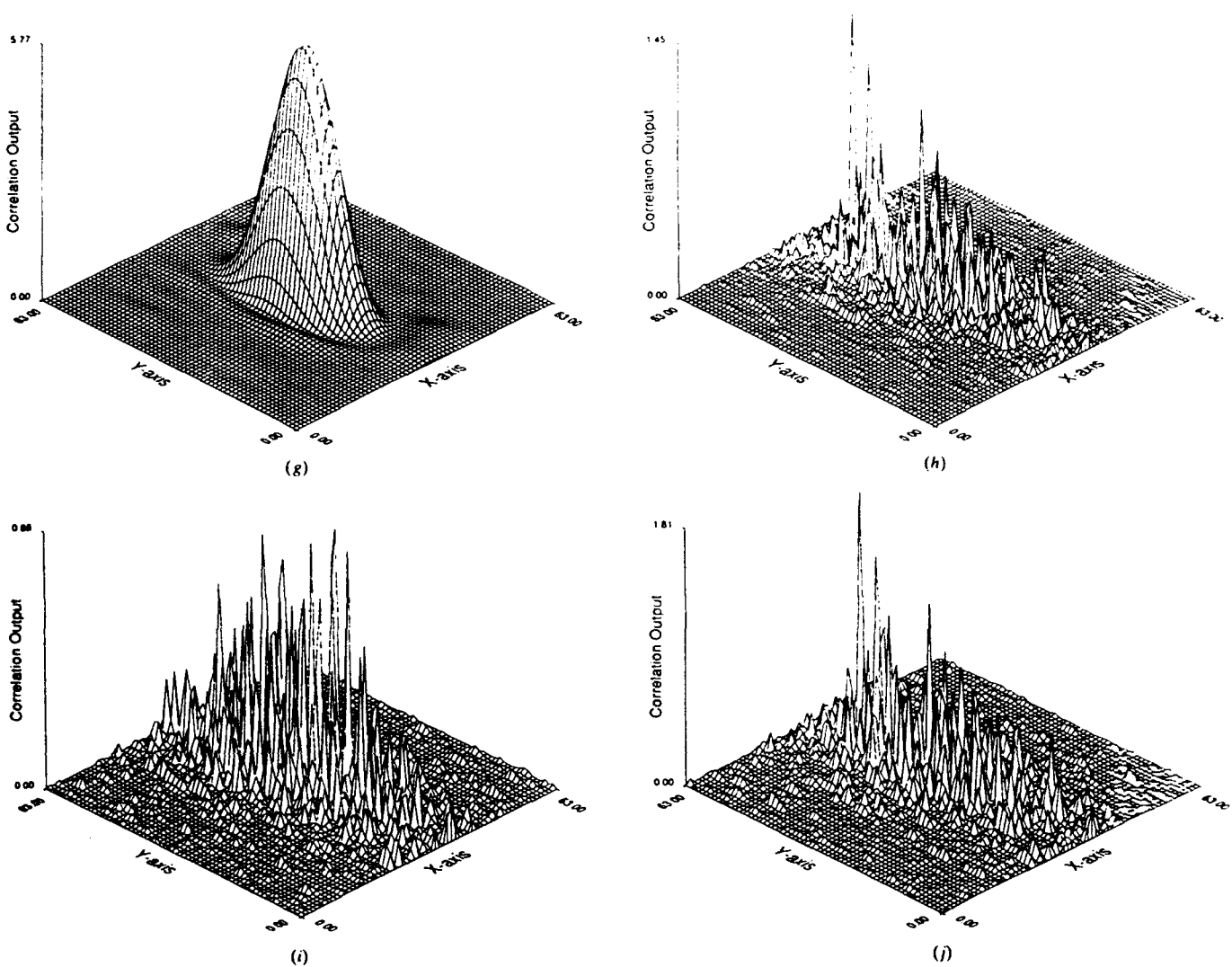
(g)

(h)

(i)

(j)

Figure 8 (continued)

Table 2. Discrimination performance of various filters.

Filter type	PG for missile launcher image (dB)	PG for tank image (dB)	Discrimination gain (dB)
CMF	35·70	32·56	3·14
POF	27·01	16·38	10·63
BPOF(real)	24·67	16·93	7·44
BPOF(imag.)	21·52	3·02	18·50
OPOF	34·23	32·61	1·62
OBPOF	34·40	32·48	1·92
CTMF	34·88	33·83	1·05
OPCE-POF	17·47	−12·65	30·12
OPCE-BPOF	15·61	−5·30	20·91
OPCE-CTMF	16·02	−6·61	22·63

measure of discrimination gain (DG), DG is indicated in table 2.

This table seems to divide the 10 filters into three groups. The filters optimized for SNR (CMF, OPOF, OBPOF and CTMF) provide very poor discrimation. Filters designed to maximize PCE (OPCE-POF, OPCE-RPOF and OPCE-CTMF) provide very good discrimination. Other filters (POF, BPOF-real, BPOF-imaginary) appear to provide intermediate levels of discrimination.

The correlation outputs in figure 8 tell the rest of the story. We correlated the noise-free tank image with the various filters. One can see, by comparing these figures with the corresponding ones in figure 5, that the optimal PCE filters do indeed provide better discrimation

4   *B. V. K. Vijaya Kumar et al.*

rformance. There are no clean peaks in these corre-
tion outputs even though the image is noise-free.
owever, some of the optimal SNR filters (OPOF,
BPOF and CTMF) are virtually unchanged. This is
ue to the fact that the regions of support for these
ters are so small that the largest values for both
ages are passed.

Conclusions

i this paper, we developed and described algorithms
ir designing partial information correlation filters
'OFs, BPOFs and CTMFs) yielding maximum peak-
-correlation energy (PCE) values. Numerical and
mulation results were presented to quantify the resul-
nt improvements. It was observed that significant
CE increases (about 15 dB) can be obtained by use of
suitable region of support. We noted that these
1 rovements in peak sharpness were obtained at the
(pense of degraded noise tolerance. Unfortunately,
oise tolerance and peak sharpness apear to be conflict-
g objectives in correlation filter design. When the
put scene is known *a priori* to be reasonably noise-
ee (as in controlled scenarios such as robotic vision),
makes more sense to optimize PCE. On the other
ind, when the input image can be noisy (as in military
oject recognition), it makes sense to maximize SNR.
here may be other situations where both measures are
iportant and we need to find a compromise solution.
'e are currently working on developing such tech-
ques. Also, filters allowing high spatial frequencies
uch as OPCE-POF, OPCE-BPOF, etc.) provide bet-
r discrimination than the filters allowing only low
atial frequencies.

eferences

1] North, D. O., 1963, An analysis of the factors which determine signal/noise discriminations in pulsed-carrier systems, *Proc. IEEE*, **51**, 1016.

2] Casasent, D., and Furman, A., 1977, Sources of correlation degradation, *Appl. Optics*, **16**, 1652.

3] Vanderlugt, A., 1964, Signal detection by complex spatial filtering, *IEEE Trans. Inf. Theory*, **10**, 139.

1] Horner, J. L., 1982, Light utilization in optical correlation, *Appl. Optics*, **21**, 4511.

5] Caulfield, H. J., and Weinberg, M. H., 1982, Computer recognition of 2-D patterns using generalized matched filters, *Appl. Optics*, **21**, 1699.

o] Hester, C. F., and Casasent, D., 1980, Multivariant technique for multi-class pattern recognition. *Appl. Optics*, **19**, 1758.

[7] Schils, G. F., and Sweeney, D. W., 1986, Iterative techniques for the synthesis of optical correlation filters, *J. opt. Soc. Am.* **3**, 1433.

[8] Horner, J. L., and Gianino, P. D., 1984, Phase-only matched filtering, *Appl. Optics*, **21**, 812.

[9] Psaltis, D., Paek, E. G., Venkatesh, S. S., 1984, Optical image correlation with a binary spatial light modulator, *Opt. Engng.* **23**, 698.

[10] Dickey, F. M., Vijaya Kumar, B. V. K., Romero, L. A., and Connelly, J. M., 1990, Complex ternary matched filters yielding high signal-to-noise ratios, *Opt. Engng*, **29**.

[11] Ross, W. E., Psaltis, D., and Anderson, R. H., 1983, Two-dimensional magneto-optic spatial light modulator, *Opt. Engng*, **22**, 485.

[12] Vijaya Kumar, B. V. K., and Bahri, Z., 1989, Phase-only filters with improved signal-to-noise ratios, *Appl. Optics*, **28**, 250.

[13] Vijaya Kumar, B. V. K., and Bahri, Z., 1989, Efficient algorithm for designing ternary-valued filter yielding maximum signal-to-noise ratio, *Appl. Optics*, **28**, 1919.

[14] Vijay Kumar, B. V. K., Shi W., and Hendrix, C. D., 1990, Phase-only filters with maximally sharp correlation peaks, *Optics Lett.* **15**, 807.

[15] Dickey, F. M., Stalker, K. T., and Mason, J. J., 1988, Bandwidth considerations for binary phase-only filters, *Appl. Optics*, **27**, 3811.

[16] Kast, B. A., Giles, M. K., Lindell, S. D., and Flannery, D. L., 1989, Implementation of ternary phase-amplitude filters using a magneto-optic spatial light modulator, *Appl. Optics*, **28**, 1044.

[17] Horner, J. L., and Leger, J., Pattern recognition with binary phase-only filters, *Appl. Optics*, **24**, 609.

[18] Cottrell, D. M., Lilly, R. A., Davis, J. A., and Day, T., 1987, Optical correlator performance of binary phase-only-filters using fourier and Hartley transforms, *Appl. Optics*, **26**, 3755.

[19] Bracewell, R. N., 1986, *The Hartley Transform* (New York: Oxford University Press).

[20] Dickey, F. M., and Hansche, B. D., 1989, Quad-phase-only filters for pattern recognition, *Appl. Optics*, **28**, 1611.

[21] Hansche, B. D., Mason, J. J., and Dickey, F. M., Quad-phase-only filter implementation, *Appl. Optics*, **28**, 4840.

[22] Mahalanobis, A., Vijaya Kumar, B. V. K., and Casasent, D., 1986, Minimum average correlation energy filters, *Appl. Optics*, **26**, 3633.

[23] Dickey, F. M., and Romero, L. A., 1989, Dual optimality of the phase-only filters, *Optics Lett.*, **14**, 4.

[24] Flannery, D. L., Loomis, J. S., and Milkovich, M. E., 1988, Transform-ratio ternary phase-amplitude formulation for improved correlation discrimination, *Appl. Optics*, **27**, 4079.

Appendix 1

Let $\{x_1, x_2, \cdots, x_N\}$ and $\{y_1, y_2, \cdots, y_N\}$ denote two
sets of positive values satisfying $z_1 \geq z_2 \geq \cdots \geq z_N > 0$
where $z_i = x_i/y_i$, $i = 1, 2, \cdots, N$. We are interested in
choosing a subset of integers R from the set
$\{1, 2, \cdots, N\}$ such that the following ratio is maximized

$$\eta(R) = \frac{(\Sigma_{i \in R} x_i)^2}{\Sigma_{i \in R} y_i}. \tag{45}$$

Let R^* denote such an optimal subset. In this appendix,
we will prove the following important proposition con-
cerning R^*.

Proposition: *If integer n is included in the optimal subset R^*, then all integers $k \leq n$ must also be included in R^*.*

Proof: Suppose $n \in R^*$, $k \notin R^*$, where $k < n$. We will demonstrate a contradiction thus proving that the above cannot happen. To show this, we define the following

$$a = \sum_{i \in R^*} x_i - x_n, \tag{46}$$

$$b = \sum_{i \in R^*} y_i - y_n. \tag{47}$$

Here a and b refer to the summation of all x_is (except x_n) in R^* and the summation of all y_is (except for y_n) in R^*, respectively. Since $\eta(R^*)$ must be larger than $\eta(R)$ where R is obtained from R^* by removing n, we have the following inequality.

$$\frac{(x_n + a)^2}{y_n + b} \geq \frac{a^2}{b}. \tag{48}$$

After some algebraic manipulation, we obtain

$$bx_n^2 + 2abx_n \geq a^2 y_n. \tag{49}$$

Let us now consider the η obtained when we use $R' = R^* \cup \{k\}$, i.e., the subset obtained by adding k to the optimal subset. Let Δ denote the difference between $\eta(R')$ and $\eta(R^*)$

$$\Delta = \eta(R') - \eta(R^*)$$

$$= \frac{(x_n + a + x_k)^2}{y_n + b + y_k} - \frac{(x_n + a)^2}{y_n + b}$$

$$= \frac{(y_n + b)(x_n + a + x_k)^2 - (x_n + a)^2 (y_n + b + y_k)}{(y_n + b)(y_n + b + y_k)}. \tag{50}$$

Since the denominator of equation (50) is positive, we can show that $\Delta > 0$ if we can show that the numerator of equation (50) is positive. That numerator can be simplified as below

$$\text{numerator} = x_k(y_n + b)(x_k + 2x_n + 2a) - y_k(x_n + a)^2. \tag{51}$$

Dividing both sides of equation (51) by the positive term $y_k(y_n + b)$, we get the following result

$$\frac{\text{numerator}}{y_k(y_n + b)}$$

$$= \frac{x_k}{y_k}(x_k + 2x_n + 2a) - \frac{(x_n + a)^2}{y_n + b}$$

$$\geq \frac{x_n}{y_n}(x_k + 2a + 2x_n) - \frac{(x_n + a)^2}{(y_2 + b)}$$

$$= \frac{x_n(y_n + b)(x_k + 2a + 2x_n) - y_n(x_n + a)^2}{y_n(y_n + b)}$$

$$= \frac{(bx_n^2 + 2abx_n - a^2 y_n) + (bx_n^2 + x_n^2 y_n + bx_n x_k + x_n x_k y_n)}{y_n(y_n + b)} \tag{52}$$

where we used the fact that since $k < n$, $z_k = x_k / y_k \geq z_n = x_n / y_n$. The second term in the numerator of equation (52) is obviously positive. The first term is also positive from equation (49). This implies that Numerator > 0, which in turn implies that $\eta(R') > \eta(R^*)$, an absurd result. This is a contradiction thus implying that if integer n is included in R^*, then all $k \leq n$ must also be included in R^*.

This proposition implies that the optimal subset R^* must be of the following form

$$R = \{i : z_i \geq T\}, \tag{53}$$

where the threshold T is not known *a priori*. Thus we can vary T and determine the best R^*.

Appendix 2

Proposition: *For a given region of support R, the BPOF yielding the highest $|c(0)|$ value must be of the form given in equation (29).*

Proof: Let the output $c(0)$ using a BPOF $H(u)$ have magnitude $|c|$ and associated phase β, that is

$$\int_R S(u) H(u) \, du = |c| e^{-j\beta}. \tag{54}$$

Choosing $H(u)$ to maximize $|c|$ is equivalent to maximizing the following

$$|c| = \int_R S(u) e^{-j\beta} H(u) \, du$$

$$= \int_R S_{\beta R}(u) H(u) \, du. \tag{55}$$

Obviously, $|c|$ is maximized (for a given β) if we can use the following BPOF

$$H_\beta(u) = \begin{cases} \text{sgn}\,[S_{\beta R}(u)], & \text{for} \quad u \in R, \\ 0, & \text{for} \quad u \notin R. \end{cases} \quad (56)$$

Let the binary phase-only filter that maximizes $|c|$ result in a output correlation with phase β^*. Then we will prove that this optimal filter must be $H_\beta(u)$ where $H_\beta(u)$ is defined in equation (58). Corresponding maximum value of $|c|$ is denoted by $|c(\beta^*)|$ and is given by

$$|c(\beta^*)| = \int_R S_{\beta^*R}(u) H_\beta\cdot(u)\,du \quad (57)$$

We need to prove that when we substitute $H_\beta\cdot(u)$ in equation (54), we do indeed get a phase of β^*. Let us assume that this filter yields an output phase $\beta_1 \neq \beta^*$. Let $\Delta = \beta_1 - \beta^*$. From equation (57), we can write

$$|c(\beta^*)| = \text{Re}\left[e^{-j\beta^*} \int_R S(u) H_\beta\cdot(u)\,du \right]. \quad (58)$$

Also, since we assumed that when we use $H_\beta\cdot(u)$ in equation (54) we get an output with magnitude $|\hat{c}|$ and phase β_1, we can write the following

$$|\hat{c}|\,e^{j\Delta} = e^{-j\beta^*} \int_R S(u) H_\beta\cdot(u)\,du. \quad (59)$$

The right-hand side of equation (58) is the real part of the right hand side of equation (59) and thus $|c(\beta^*)|$ must be the real part of $|\hat{c}|\,e^{j\Delta}$. Thus, unless $\Delta = 0$, we have

$$|\hat{c}| > |c(\beta^*)|. \quad (60)$$

This is a contradiction since we started with the assumption that $H_\beta\cdot(u)$ is the optimal filter and $|c(\beta^*)|$ is the maximum possible value. Thus β_1 cannot be different from β^*. So, the optimal filter must yield a consistent output phase.

APPENDIX G

TRADEOFFS IN THE DESIGN OF CORRELATION FILTERS

B.V.K. Vijaya Kumar, Charles D. Hendrix and Daniel W. Carlson

Electrical and Computer Engineering Department
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Designing filters for use with optical correlators is really an exercise in trading one performance measure against another. In this critical review, we present several different situations where such a tradeoff is carried out. An informed understanding of this law of nature is important in making sure that our goals in optical pattern recognition are realistic.

1   Introduction

Correlation has been the focal point of much optical pattern recognition research over the past twenty five years. Vanderlugt[1] demonstrated how a coherent optical processor can be used to implement complex frequency response needed in a Classical Matched Filter (CMF). The CMF is attractive in the sense that it maximizes[2] the tolerance to the additive noise in the input. However it suffers from the following drawbacks.

- CMFs are unacceptably sensitive to distortions in the input image (e.g., rotations, changes in scale, etc.).

- CMFs are light-inefficient because their transmittance is less than one at many frequencies.

- The complex-valued frequency response of the CMF makes it inconvenient for implementation on currently available real-time Spatial Light Modulators (SLMs) such as the Magneto-Optic SLM[3] (MOSLM).

Several strategies have been used to alleviate the above problems. To make the optical correlators more light efficient, Phase-Only Filters[4] (POFs) and other variants have been suggested. To make these filters more appropriate for implementation on available devices, filters such as the Binary Phase-Only Filters[5] (BPOFs) have been suggested. Many methods to reduce the distortion sensitivity of the matched filters have been proposed. A good place to start learning more about these attempts is the survey paper by Vijaya Kumar[6]. Most of these methods focus on one aspect (e.g., distortion sensitivity) while ignoring others (e.g., light throughput, noise tolerance, etc.) in the design of correlation filters. The unfortunate reality is that by improving one feature of the filter performance, we are usually degrading others. The

goal of this manuscript is to pull together many different ways that tradeoffs can be carried out. By understanding the implicit tradeoffs, we can become more realistic in setting our goals for filter design.

To help us in judging these tradeoffs quantitatively, we introduce our notation in Section 2 and introduce the relevant performance measures in Section 3. In Section 4, we discuss the Optimal Tradeoff Filters (OTFs) introduced by Refregier[7]. These show how noise tolerance, correlation peak sharpness and light efficiency can be traded off in designing a complex-valued filter. Another method suggested is to record the matched filters nonlinearly. By using fractional power filters, we demonstrate in Section 5 that this nonlinear recording leads to another example of trading off among these performance measures. The POFs and BPOFs as introduced originally are allpass, i.e., these filters have unit magnitude at all frequencies. Recently, it has been shown that the noise tolerance[8,9] and correlation peak sharpness[10] of these filters can be maximized by setting some frequencies to have zero magnitude. The set of frequencies for which the filter magnitude is nonzero is known as the Region of Support (ROS). In Section 6, we describe how the ROS can be selected to provide an optimal tradeoff between noise tolerance and correlation peak sharpness. Another way the ROS can be used is to tradeoff the sensitivity to noise in the input against detector noise. This is discussed in Section 7. In Section 8, we switch gears and consider an important tradeoff in composite filter design. In this section, we show the tradeoff in the number of training images used and the resulting noise tolerance. In Section 9, we discuss the optimal circular harmonic tradeoff filters introduced by Refregier[11] and also summarize the work by Refregier and Figue[12] in relating OTFs to classical Wiener filters. Finally, in Section 10, we provide our concluding remarks.

2 Background

Let us first consider the problem of detecting the presence or the absence of a target image $s(x,y)$ corrupted by additive, zero-mean noise $n(x,y)$. One optical processor designed for this task is the coherent optical correlator shown in Fig. 1.
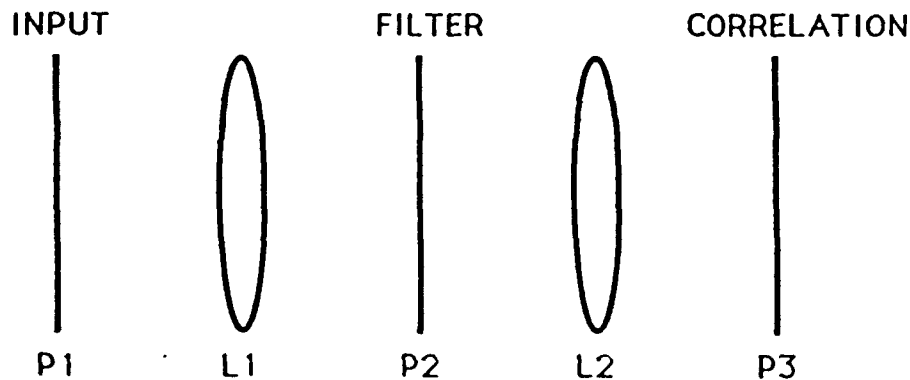
INPUT   FILTER   CORRELATION

P1   L1   P2   L2   P3

Figure 1: A schematic of coherent optical correlator.

Here the input image $r(x,y)$ (which is $s(x,y) + n(x,y)$) is placed in plane P_1 and illuminated by coherent light. Then the light wavefront reaching plane P_2 can be represented by $R(u,v)$ which is the 2-D Fourier Transform (FT) of $r(x,y)$. If we prerecord and place in plane P_2 a complex transmittance $H^*(u,v)$ (with the superscript $*$ denoting the complex conjugation), then the light wavefront $c(\tau_x,\tau_y)$ in

plane P_3 is given as follows.

$$c(\tau_x, \tau_y) = F^{-1}\{R(u, v)H^*(u, v)\}$$
$$= r(x, y) \star h(x, y) \tag{1}$$

where $h(x, y)$ is the inverse Fourier transform of $H(u, v)$ and $\star$ denotes the 2-D crosscorrelation operation. We must emphasize that the complex conjugate of $H(u, v)$ will be placed in plane P_2. The classical matched filter (CMF) uses

$$H_{CMF}(u, v) = S(u, v) \tag{2}$$

where $S(u, v)$ is the 2-D Fourier transform of $s(x, y)$, the reference image. When this $s(x, y)$ is present in the input scene $r(x, y)$, the resultant correlation output $c(\tau_x, \tau_y)$ has a large value (called the "peak") at one location and small values elsewhere. The spatial coordinates of the correlation peak indicate the location of $s(x, y)$ in the input scene.

As described in Section 1, one problem with CMFs is that the filter transmittance $|H(u, v)|$ is less than one at many frequencies and thus causes much of the light incident on plane P_2 to be absorbed by the medium in plane P_2. One solution for this is the POF introduced by Horner and Gianino[4].

$$H_{POF}(u, v) = \frac{S(u, v)}{|S(u, v)|}. \tag{3}$$

Thus $|H_{POF}(u, v)|$ is one for all frequencies and there will be no light attenuation. Sinces POFs are allpass, there is no way to filter out the noise in the input plane. We will discuss this in detail in Section 6.

Some currently available SLMs such as the MOSLM can be made to provide transmittances that are either +1 or -1. For implementation on such devices, we must make $H(u, v)$ to be either +1 or -1. This can be done by binarizing either the real part[5], the imaginary part[13] (or some linear combination[14] of the two) of $S(u, v)$ to get $H(u, v)$. Such filters are known as Binary Phase-Only Filters (BPOFs).

$$H_{BPOF}(u, v) = \text{Sgn}[\alpha S_R(u, v) - \sqrt{1 - \alpha^2} S_I(u, v)] \tag{4}$$

where $0 \leq \alpha \leq 1$, $S_R(u, v)$ and $S_I(u, v)$ are the real part and the imaginary part, respectively, of $S(u, v)$. The Sgn[·] function is defined by

$$\text{Sgn}[x] = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}. \tag{5}$$

Other attempts to design correlation filters suitable for implementation on available SLMs include the work by Juday[15,16] and by Farn and Goodman[17].

Both POFs and BPOFs are sensitive to minor variations in the input image and do not address the problem of distortion-invariance. Several filter design schemes[6] have been proposed to make the correlation outputs more invariant to distortions in the input image. One of these methods is known as the Synthetic Discriminant Function (SDF) approach. The idea underlying the SDF method is to design a synthetic or composite $h(x, y)$ such that when it is correlated against images $\{s_1(x, y), s_2(x, y), \ldots, s_N(x, y)\}$, the resulting correlation outputs have equally strong correlation peaks at the correct locations. Here $s_i(x, y), i = 1, 2, \ldots, N$ represent the original image $s(x, y)$ distorted by various amounts. The "training set" of these N images represents the expected distortions in the image. In the first SDF method[18], the composite image $h(x, y)$ is assumed to be a linear combination of the N training images, i.e.,

$$h(x, y) = a_1 s_1(x, y) + a_2 s_2(x, y) + \cdots + a_N s_N(x, y) \tag{6}$$

where $a_1, a_2, \ldots, a_N$, are weights to be determined. These weights are determined such that $h(x,y)$, when correlated with the training image $s_i(x,y)$ yields a specified cross-correlation output c_i at the origin.

$$h(x,y) \star s_i(x,y)|_{(0,0)} = \iint h(x,y) s_i(x,y)\, dx\, dy$$
$$= c_i, \quad i = 1, 2, \ldots, N \tag{7}$$

Substituting eq. (7) in eq. (6), we can obtain the following set of N linear equations in N unknowns, namely $a_1, a_2, \ldots, a_N$.

$$\sum_{i=1}^{N} a_i R_{ij} = c_j, \quad j = 1, 2, \ldots, N, \tag{8}$$

where R_{ij}, the inner product of $s_i(x,y)$ and $s_j(x,y)$ is given by

$$R_{ij} = \iint s_i(x,y) s_j(x,y)\, dx\, dy. \tag{9}$$

While the basic SDF solves the problem of distortion sensitivity, it introduces other difficulties. There is no accounting of input noise in the basic SDF. Also, since the basic SDF controls what happens to the output at only one point, the resulting correlation outputs usually exhibit sidelobes much larger than the desired value at the origin.

To introduce noise tolerance into SDF design, Vijaya Kumar[19] proposed Minimum Variance SDFs (MVSDFs) that exhibit the smallest output variance while satisfying the SDF constraints in eq. (7). To reduce the false sidelobe problem, Mahalanobis et. al.[20] introduced the Minimum Average Correlation Energy (MACE) filter that minimizes the average correlation plane energy while satisfying the SDF constraints in eq. (7). Recently, Refregier[21] showed how these different design objectives can be combined into a single filter design.

In this section, we provided a brief background about the classical matched filters, phase-only filters, binary phase-only filters and synthetic discriminant functions. In the next section, we review some performance measures useful in assessing the effectiveness of these filters.

3 Performance Measures

As already indicated, it is important to use quantitative measures of performance in evaluating various filter design schemes. In this section, we review some useful measures. This is essentially a condensed version of a longer paper[22] on the same topic.

We have stated in Section 2 that CMFs provide maximum noise tolerance. More precisely, CMFs yield the highest Signal-to-Noise Ratio (SNR) where the SNR (assuming that the correlation peak is at the center) is defined as follows

$$SNR \triangleq \frac{|E\{c(0,0)\}|^2}{Var\{c(0,0)\}}. \tag{10}$$

This SNR is the ratio of the average value of the correlation peak to its standard deviation. Other SNR measures can be found in the literature, but we feel that this is the most appropriate for characterizing

the noise tolerance of a linear processor. This SNR can also be expressed as

$$SNR = \frac{|\iint S(u,v)\,H^*(u,v)\,du\,dv|^2}{\iint P_n(u,v)\,|H(u,v)|^2\,du\,dv}$$ (11)

where $P_n(u,v)$ is the power spectral density of the additive noise $n(x,y)$ in the input. The SNR expression in eq. (11) indicates the explicit dependence on the reference image FT $S(u,v)$, noise spectrum $P_n(u,v)$ and of course, on the filter $H(u,v)$. Using Cauchy-Schwartz inequality, one can easily prove[2] that

$$SNR \leq \iint \frac{|S(u,v)|^2}{P_n(u,v)}\,du\,dv$$
$$= SNR_{max}$$ (12)

and that this equality is achieved by the CMF. All other filters (including POF, BPOF, SDF, etc.) can only yield smaller SNRs. However, we can seek POFs and BPOFs that maximize SNRs. Such SNR-maximal filters are known as Optimal POFs[6] and Optimal BPOFs[23]. In some sense, the MVSDF[19] yields the optimal SNR among all SDFs.

Another desirable attribute for a correlation output is that its peak is sharp and that the sidelobes are low. This is necessary for an accurate localization of the correlation peak (and hence the target image in the scene) and for reducing false peaks. Several measures have been used to characterize the peak sharpness. But the one we find most convenient for analysis is Peak-to-Correlation Energy (PCE) defined below.

$$PCE = \frac{|c(0,0)|^2}{\iint |c(\tau_x,\tau_y)|^2\,d\tau_x\,d\tau_y}.$$ (13)

This measure provides the ratio of the correlation peak to the total energy in the correlation plane. PCE is large for delta-function type correlations and is close to zero for constant correlations. Recently, Horner[24] suggested the use of a modified PCE measure in which the denominator excludes the energy at the peak. The PCE can be rewritten in terms of $S(u,v)$ and $H(u,v)$ as below.

$$PCE = \frac{|\iint S(u,v)\,H^*(u,v)\,du\,dv|^2}{\iint |S(u,v)|^2\,|H(u,v)|^2\,du\,dv}$$ (14)

It is fairly easy to show that PCE is maximized by the inverse filter

$$H_{IF}(u,v) = \frac{1}{S^*(u,v)}.$$ (15)

SNR and PCE measure the noise tolerance and peak sharpness. Another important attribute needed for optical correlation filters is their light throughput efficiency. We would like $|H(u,v)|$ to be as close to one as possible at all frequencies. Since these filters are energy absorbing, $0 \leq |H(u,v)| \leq 1$. Thus CMFs have $|H(u,v)|$ close to zero at many frequencies and much of the light incident on P_2 will not make it to the detector in plane P_3. To quantify this light efficiency, Horner[25] originally introduced the following measure.

$$\eta_H = \frac{\iint |c(\tau_x,\tau_y)|^2\,d\tau_x\,d\tau_y}{\iint |s(x,y)|^2\,dx\,dy}.$$ (16)

The original Horner efficiency considers the ratio of the total energy in the correlation plane to the total energy in the input plane. Since $|H(u,v)| \leq 1$, the Horner efficiency ≤ 1 for all filters. From the

following frequency domain version of eq. (16), it is easy to see that $\eta_H = 1$ whenever $|H(u,v)| = 1$ for all (u,v), i.e., whenever the filter is an all-pass filter (e.g., POF, BPOF, etc.).

$$\eta_H = \frac{\iint |S(u,v)|^2 |H(u,v)|^2 \, du \, dv}{\iint |S(u,v)|^2 \, du \, dv}.$$ (17)

One drawback with η_H is that it includes all the output light in the numerator. However, it is more important to have more light in the correlation peak than anywhere else. Thus it is preferable to use the following modified Horner efficiency along the lines suggested by Caulfield[26].

$$
\eta'_H = \frac{|c(0,0)|^2}{\iint |s(x,y)|^2 \, dx \, dy}
$$
$$
= \frac{|\iint S(u,v) H^*(u,v) \, du \, dv|^2}{\iint |S(u,v)|^2 \, du \, dv}.
$$ (18)

It can be seen from eqs. (17) and (18) that the filter $H(u,v)$ affects only the numerators of both η_H and η'_H. Thus all all-pass filters (e.g., POF, BPOF, etc.) yield the same maximal η_H. However the numerator of eq. (18) is maximized (under the constraint $|H(u,v)| \leq 1$) by the POF in eq. (3). Thus η'_H appears to be the more appropriate light efficiency measure to use.

Several other desirable attributes of correlation filters are more difficult to quantify. One of these is the ability of the filters to locate the correlation peaks accurately. Even when there is no input noise, we cannot be sure that a centered input leads to a correlation peak at the origin unless the phase of the input FT is completely cancelled by the phase of the filter. Such complete phase cancellation occurs with CMF and POF and not with BPOFs. Thus we cannot be sure that the correlation peaks are correctly located when we use BPOFs. When the input is corrupted by noise, the peak locations get affected and the variance in the peak or Peak Location Error (PLE) is a descriptive measure. Recently, Vijaya Kumar et. al.[27] showed that the CMF minimizes the PLE and in that sense is the best filter for locating references in an input.

Another desirable feature for a correlation filter is that it should yield large correlation outputs with a "desired" class of images and small outputs with all others. Filters designed to maximize SNR in eq. (10) are very good at detecting an object in noise, but they may not be good at discriminating one class from another. Thus a measure of the discrimination capability such as Fisher Ratio[28] must be employed.

Filters such as SDFs are designed to provide distortion invariance. Thus a measure for the distortion sensitivity must be used. Since $c(0,0)$ is usually considered in detecting an object, the following measure for distortion sensitivity may prove useful.

$$DS = \frac{|c(0,0)|_{max} - |c(0,0)|_{min}}{|c(0,0)|_{max} + |c(0,0)|_{min}},$$ (19)

where the subscripts "max" and "min" refer to maximum and minimum $|c(0,0)|$ values obtained among all possible distorted inputs. When $|c(0,0)|$ is the same for all distorted images, the distortion sensitivity (DS) is zero. When $|c(0,0)|_{min} = 0$, then DS achieves its highest (worst) value of one. Thus a distortion-invariant filter tries to force DS to zero. Of course, the DS value obtained depends very much on what we mean by the "Distortion Set". When the set of distortions span a small range, it is easier to achieve DS values close to zero.

108

There may be other important considerations such as the complexity of implementation, dynamic range requirements, space bandwidth product requirements, etc. that we need to specify. But for evaluating correlation filter designs, the measures we have introduced should prove to be adequate.

4 Optimal Trade-off Filters

Recently, Refregier[7] introduced Optimal Trade-off Filters (OTFs) which are the subject of this section. In this approach, filters are designed to optimally tradeoff three performance measures, namely: the SNR, the PCE and the modified Horner efficiency η'_H. The OTFs are optimal in the sense that when two of these measures are held constant, the OTFs yield the best value for the third measure among all possible filters.

To understand the theory of OTFs, let us rewrite the three performance measures as below.

$$SNR = \frac{|\text{Peak}|^2}{\text{Variance}},$$ (20)

$$PCE = \frac{|\text{Peak}|^2}{\text{Output Energy}},$$ (21)

and

$$\eta'_H = \frac{|\text{Peak}|^2}{\text{Input Energy}},$$ (22)

where

$$\text{Peak} = \iint S(u,v) H^*(u,v)\, du\, dv,$$ (23)

$$\text{Variance} = \iint P_n(u,v) |H(u,v)|^2\, du\, dv,$$ (24)

$$\text{Output Energy} = \iint |S(u,v)|^2 |H(u,v)|^2\, du\, dv,$$ (25)

and

$$\text{Input Energy} = \iint |S(u,v)|^2\, du\, dv.$$ (26)

Suppose that we want to maintain PCE and η'_H at specified levels and maximize the SNR. Since the input energy cannot be controlled by the filter $H(u,v)$, maintaining η'_H in eq. (22) at a specified value implies keeping |Peak| in eq. (23) at a constant value. Since |Peak| is a constant, maintaining PCE in eq. (21) at a specified value is equivalent to keeping the output energy in eq. (25) a constant. Finally, since |Peak| is a constant, maximizing the SNR in eq. (20) is equivalent to minimizing the variance in eq. (24). Then the OTF $H(u,v)$ can be obtained by minimizing the variance in eq. (24) while holding the |Peak| in eq. (23) and the output energy in eq. (25) at constant values. Refregier[7] goes on to show that this minimization leads to the following filter.

$$H(u,v) = \sigma_\lambda \left\{ \frac{S(u,v)}{\mu P_n(u,v) + (1-\mu)|S(u,v)|^2} \right\},$$ (27)

where the function $\sigma_\lambda\{\cdot\}$ is defined as

$$\sigma_\lambda\{y\} = \begin{cases} \lambda y & \text{if } |y| \le 1/\lambda \\ e^{j\psi} & \text{if } |y| > 1/\lambda \end{cases}$$ (28)

where $y = |y|e^{j\psi}$ and $0 \leq \mu \leq 1$ and $\lambda \geq 0$ are design parameters. As can be seen from eq. (27), ($\mu = 1, \lambda = 0$) leads to the CMF whereas ($\mu = 0, \lambda = 0$) leads to the inverse filter. The CMF maximizes noise tolerance whereas the inverse filter (IF) yields the sharpest correlation peaks. Intermediate values of μ lead to filters that compromise between the two extremes. The parameter λ is used to indicate the importance of light effeciency η'_H. When λ is close to zero, $\sigma_\lambda\{y\}$ of eq. (28) is essentially proportional to y whereas when λ is large, $\sigma_\lambda\{y\}$ is essentially the phase-only version of y. Thus large λ values indicate that the light efficiency is important.

In Fig. 2, we redraw a figure from Refregier[7] to illustrate these tradeoffs graphically. For $\lambda = 0$, we obtain the tradeoff between PCE and SNR with the CMF and IF at the extremes of the curve. As λ approaches infinity, we get a POF with worse (i.e., lower) values for both PCE and SNR. Other filters such as BPOFs will have even lower SNR values for the same PCE value. Of course this tradeoff curve will change as we go to a different image.
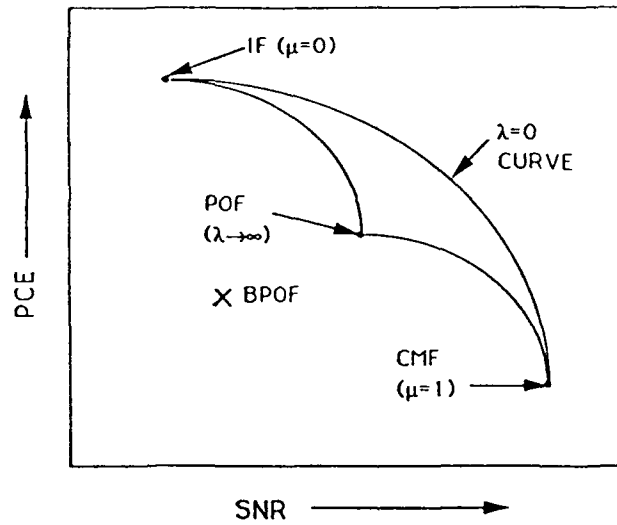
Figure 2: Graphical illustration of the tradeoffs among SNR, PCE and η'_H (Adapted from [7]).

5   Fractional Power Filters

Recently, there has been a growing interest[29] in the use of nonlinearities in the frequency plane of a correlator. With the help of a particular class of nonlinearly recorded matched filters, we illustrate in this section the tradeoffs inherent in such a filter design. We refer to this class of filters as Fractional Power Filters[22] (FPFs).

$$H_{FPF}(u,v) = \begin{cases} [|S(u,v)|]^p exp[j\theta(u,v)] & \text{if} \quad |S(u,v)| \neq 0 \\ 0 & \text{if} \quad |S(u,v)| = 0 \end{cases} \tag{29}$$

where $\theta(u,v)$ is the phase associated with $S(u,v)$ and where p is the power associated with the FPF. It is easy to see that $H_{FPF}(u,v)$ specializes to the IF, POF and CMF for $p = -1, 0$ and $+1$, respectively. Since the phase of $H_{FPF}(u,v)$ is $\theta(u,v)$ for all p, the filter phase completely cancels (when there is no input noise) the phase of the input FT thus ensuring that the resulting correlation peak is at the origin.

Allowing p to take on values other than -1, 0 and +1 leads to the notion of FPFs which employ nonlinear functions of the magnitude of the signal FT.

Substituting $H_{FPF}(u,v)$ in the SNR definition in eq. (11), we get

$$SNR_{FPF} = \frac{[\iint |S(u,v)|^{p+1} du\, dv]^2}{\iint P_n(u,v) |S(u,v)|^{2p} du\, dv}.$$ (30)

For the case of white noise, $P_n(u,v) = N_o$ and SNR_{FPF} in eq. (30) is maximized by $p = 1$ (the CMF). The PCE of FPF can be obtained by substituting eq. (29) in eq. (14).

$$PCE_{FPF} = \frac{[\iint |S(u,v)|^{p+1} du\, dv]^2}{\iint |S(u,v)|^{2(p+1)} du\, dv}.$$ (31)

Once again, PCE_{FPF} is maximized by $p = -1$, the IF. Similarly, the Horner light efficiency η_H of the FPFs can be obtained by substituting eq. (29) in eq. (17).

$$\eta_H = \delta \frac{\iint |S(u,v)|^{2(p+1)} du\, dv}{\iint |S(u,v)|^2 du\, dv}.$$ (32)

Here the constant δ is chosen to ensure that the FPF under consideration (see eq. (29)) has a maximum transmittance of one. The Horner efficiency in eq. (32) is easily related to η'_H and maximizing one is equivalent to maximizing the other.

To illustrate the tradeoff resulting from varying p in FPF design, we show in Fig. 3 how the 3 performance measures (SNR, PCE and η_H) vary as a function of p for the aircraft image shown in Fig. 4. We have allowed p to vary between -2 and +2. We see that SNR increases with p until $p = +1$ (the
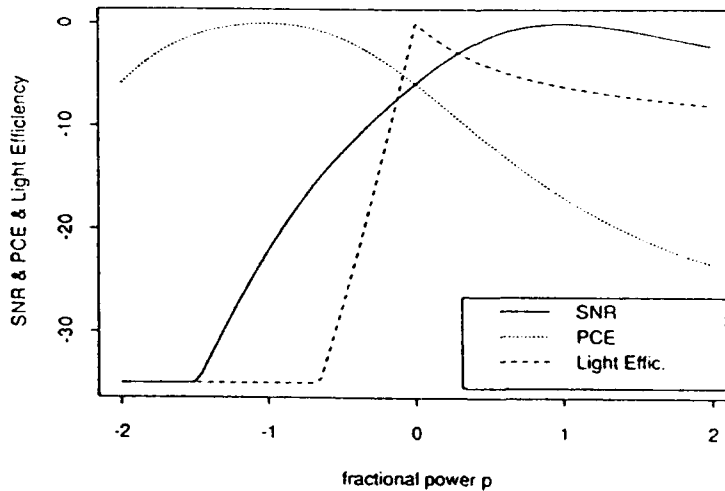

Figure 3: Variation in SNR, PCE and η_H as a function of the fractional power p for the aircraft image.

CMF) and then decreases afterwards. The PCE reaches its maximal value for $p = -1$ (the IF). The light efficiency η_H attains its maximum for $p = 0$ (the POF). It is obvious from the curves in Fig. 3 that while the nonlinearity may help improve one measure, it may hurt the other. For this image, the POF (i.e., $p = 0$) appears to be the best compromise in the sense that it yields only 6 dB less in both SNR and PCE from their optimum values. This type of numerical analysis was carried out[22] with two other images with essentially the same conclusions.
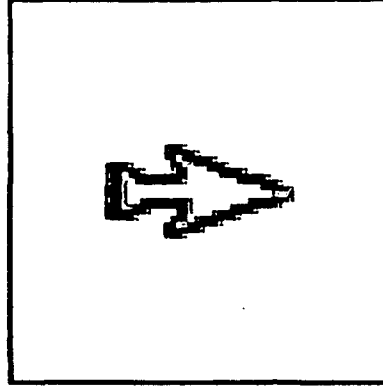
Figure 4: Aircraft (MIG) image used in numerical experiments.

6   Tradeoff by Region of Support Selection

The POF and the BPOF are allpass, i.e., they have a magnitude response of one at all frequencies. This causes the POF and the BPOF to be extremely sensitive to noise in the input[30]. This allpass nature also causes them to become very sensitive to minor distortions in the input. One method to alleviate these problems is to force the filter magnitude response at some spatial frequencies to zero. For POFs, this may be accomplished using a binary (opaque/transmissive) mask next to the conventional POF. When using a MOSLM, Kast et. al.[31] have shown we can achieve 3 levels of transmittance (+1, 0 and -1). Thus the MOSLMs can be used to implement a BPOF with some spatial frequencies set to have zero magnitudes. The set of frequencies for which $|H(u,v)| \neq 0$ is known as the Region of Support (ROS). The selection of a proper ROS is critical and Vijaya Kumar and Bahri showed how the ROS can be selected to maximize the SNR for POFs[8] and BPOFs[9]. More recently, Vijaya Kumar et. al.[32] showed how the ROS can be selected to maximize the PCE for POFs, BPOFs and Complex Ternary Matched Filters (CTMFs). In this section, we will demonstrate how the ROS can be selected to tradeoff among the various measures. We will focus our attention on the design of BPOFs. In particular, we will consider the tradeoff between the SNR and PCE. Note from eqs. (11) and (14) that the two measures have the same numerator but different denominators. This leads to the following empirical compromise performance measure (CPM) that combines the denominators of SNR and PCE using a weighting factor γ, i.e.,

$$CPM = \frac{|\iint S(u,v)H^*(u,v)\,du\,dv|^2}{\iint |H(u,v)|^2 [P_n(u,v) + \gamma|S(u,v)|^2]\,du\,dv}, \tag{33}$$

where γ is a positive constant to be specified by the designer. As $\gamma \to 0$, CPM $\to$ SNR and as $\gamma \to \infty$, CPM $\to$ (PCE/γ) so that at the two extreme values, we get filters that maximize the SNR and the PCE. At intermediate γ values, we get filters that compromise between the two extremes.

Let us now consider BPOFs with a region of support R, i.e.,

$$H(u,v) = \begin{cases} \pm 1 & \text{for}(u,v) \in R \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

Thus the design of these filters consists of two tasks. One task is to choose R and the second task is to decide whether $H(u,v)$ is +1 or -1 at frequencies (u,v) in R. Farn and Goodman[23] have shown that the optimal method for deciding whether $H(u,v)$ is +1 or -1 is to use a Threshold line angle[33] (TLA)
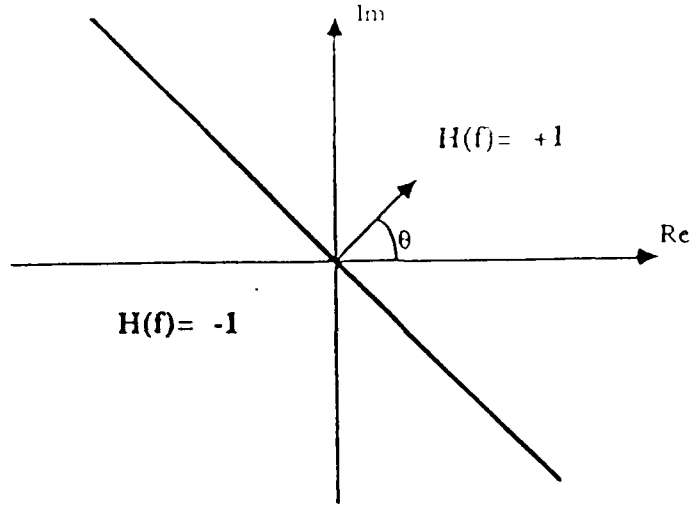
Figure 5: Threshold line angle needed for optical binary phase-only filter.

θ as shown in Fig. 5. If the signal FT $S(u,v)$ is on one side of this line, we choose $H(u,v) = +1$ and otherwise, we choose $H(u,v) = -1$. For $\theta = 0$, $H(u,v)$ is obtained by binarizing the real part of $S(u,v)$ whereas for $\theta = \pi/2$, it is obtained by binarizing the imaginary part of $S(u,v)$.

The best angle θ is not known a priori and must be determined by a search. However, once a particular θ is selected, the ROS leading to the highest CPM can be shown[34] to be of the following form.

$$R = \{(u,v): \frac{S_{\theta R}(u,v)}{P_n(u,v) + \gamma|S(u,v)|^2} \geq T\} \tag{35}$$

where T is a threshold to be determined by search and where

$$S_{\theta R}(u,v) = S_R(u,v)\cos\theta + S_I(u,v)\sin\theta, \tag{36}$$

with $S_R(u,v)$ and $S_I(u,v)$ being the real part and the imaginary part, respectively, of $S(u,v)$. Hendrix et. al.[34] go on to show how ROSs can be designed to tradeoff SNR versus η'_H and PCE versus η'_H. Flannery and Phillips[35] call the BPOFs with a region of support as Ternary Phase-Amplitude Filters (TPAFs) and discuss design tradeoffs in these filters.

To illustrate the tradeoff resulting from the ROS selection, we include here some simulation results obtained using the ROS in eq. (35). We use the boundary image of the truck shown in Fig. 6. We designed Optimal-CPM BPOFs assuming that the background noise is white, i.e., $P_n(u,v)$ is a constant. In Fig. 7 we show a graph depicting how SNR and PCE vary as we vary the γ in our BPOF design. The lower right corner corresponds to $\gamma = 0$ (or maximum SNR) and the upper left corner uses a large value of γ (or maximum PCE). Intermediate values of γ show how SNR can be traded off for PCE and vice-versa.

We have shown elsewhere[34] that as γ is increased, the area of ROS increases and the resulting correlation peak is sharper. However at the same time, the noise sensitivity of the BPOF increases because the ROS starts passing higher frequencies through.
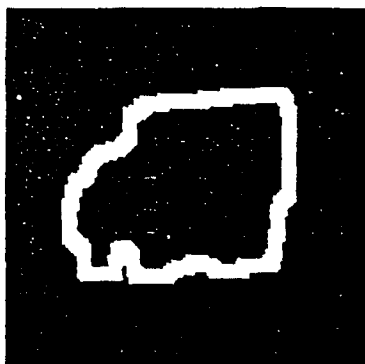
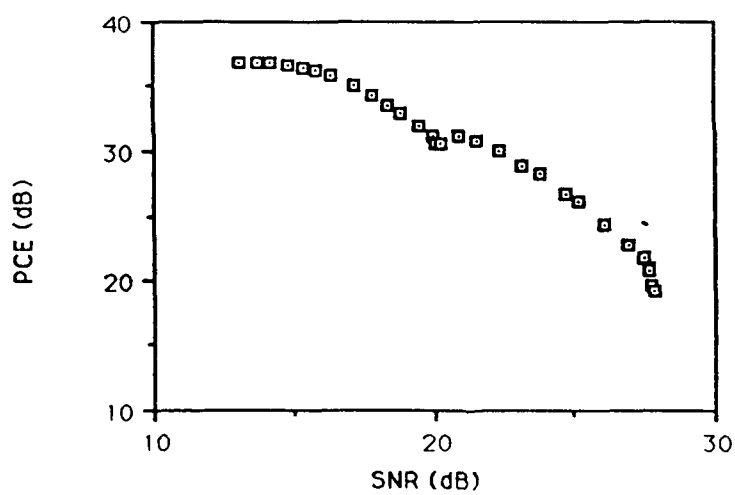Figure 6: Binary, boundary image of a truck [34].

Figure 7: PCE vs. SNR as γ is varied [34].

7 Input Noise vs. Detector Noise

Much research has been carried out in correlation filter design in maximizing the tolerance to noise in the input plane. An often ignored reality is that detectors used in the correlation plane introduce noise in the detected correlation peak value and hence degrade the resulting performance. In fact, if detector noise was absent, we could design correlation filters without worrying about their light efficiency. The main reason to desire a large η_H is to ensure that the correlation plane detector receives enough light to overcome the noise introduced in the detector. The usual procedure of maximizing the SNR (as defined in eq. (10)) ignores this important detector noise. In this section, we demonstrate that the correlation filter design provides a method for trading off tolerance to input noise against tolerance to detector noise. To illustrate this, we start with a very simple model for the detected output y.

$$y = c(0,0) + n_d, \tag{37}$$

where $c(0,0)$ is the correlation output in the absence of noise and n_d is a random variable describing the detector noise. We model n_d as a random variable with mean μ_d and variance σ_d^2. Let H_0 and H_1 denote the two detection hypotheses, i.e., H_0 denotes "signal absent" and H_1 denotes "signal present". Then an appropriate SNR measure is the following.

$$SNR_d \triangleq \frac{|E\{y/H_1\} - E\{y/H_0\}|^2}{\frac{1}{2}[Var\{y/H_1\} + Var\{y/H_0\}]} \tag{38}$$

where the subscript d is used to denote the fact that we are including the detector noise. Bahri[36] showed that this SNR_d can be rewritten in terms of the filter $H(u,v)$ as below.

$$SNR_d = \frac{|\iint S(u,v) H^*(u,v)\, du\, dv|^2}{\sigma_d^2 + \iint P_n(u,v)|H(u,v)|^2\, du\, dv}. \tag{39}$$

This SNR_d measure is very similar to the SNR measure in eq. (11) with the only difference being the extra σ_d^2 term in the denominator. When the detector noise is very low, $SNR_d \cong SNR$ and the usual SNR maximization is carried out. When the detector noise σ_d^2 is large, then SNR_d in eq. (39) is proportional to $|E\{c(0,0)\}|^2$ and we need to maximize the numerator in eq. (39) to maximize SNR_d. The numerator in eq. (39) is maximized by Phase-only filters. Thus, when the detector noise dominates, Phase-only filters (with their 100% light throughput) are optimal. For intermediate situations, SNR_d in eq. (39) must be maximized. Recently, Vijaya Kumar et. al.[37] have shown that the maximization of SNR_d in eq. (39) subject to the constraint that $|H(u,v)| \leq 1$ leads to *Saturated Filters*. Saturated Filters will also be discussed in another paper[38] in this critical review.

Bahri[36] determined the ROSs that maximize the SNR_d in eq. (39). We show in Fig. 8 the 32×32 tank image used in Bahri's simulations. The ROSs obtained for this image are shown in Fig. 9. The input noise spectrum $P_n(u,v)$ is assumed to be a constant at N_o. In Fig. 9a, we show the optimal region of support when $\sigma_d^2/N_o \ll 1$. Dark regions in this figure indicate where $H(u,v) = 0$ and white regions indicate where $|H(u,v)| = 1$. The $(0,0)$ frequency corresponds to the center of the square. As expected, the input noise dominates and the ROS is concentrated around low spatial frequencies so that the strong signal components are allowed through while blocking high frequencies that have mostly noise energy. In Fig. 9b, we show the optimal ROS when $\sigma_d^2/N_o = 1$. Here the ROS has more high frequencies than in Fig. 9a. Finally, in Fig. 9c, we show the Optimal ROS for the case when $\sigma_d^2/N_o \gg 1$. Here the detector noise dominates and the ROS is opened all the way to allow as much of the input light as possible.
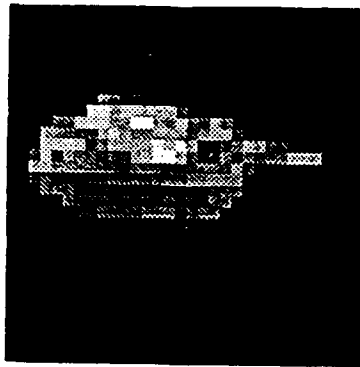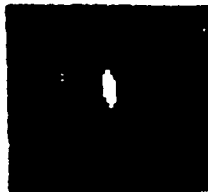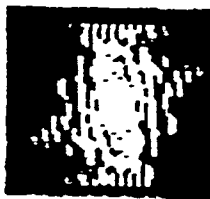
Figure 8: The 32 × 32 tank image used in ROS determination [36].

(a) $\sigma_d^2/N_o \ll 1$ (b) $\sigma_d^2/N_o = 1$ (c) $\sigma_d^2/N_o \gg 1$

Figure 9: Optimal region of support for various combinations of input and detector noise levels [36].

8 Tradeoff in Composite Filter Design

As discussed in Section 2, one of the methods proposed to overcome the distortion sensitivity of the matched filters is the SDF method. In SDF, the filter is made from a composite image $h(x,y)$ that is the weighted sum of N training images $s_1(x,y),\ldots,s_N(x,y)$ as in eq. (6). We know that the CMF yields the highest SNR among all filters. Thus, we must be trading off the noise tolerance in order to obtain distortion tolerance as we go from CMFs to SDFs. In this section, we will illustrate this tradeoff more quantitatively.

To illustrate this tradeoff, we consider only in-plane rotation distortion. Let the N training images be obtained by rotating the original image $s(x,y)$ in increments of $2\pi/N$ radians. Assume that the input image is a rotated version of $s(x,y)$ that is corrupted by zero-mean, additive, white noise. Vijaya Kumar and Pochapsky[39] analyzed how the resulting SNR (defined in eq. (10)) varies as a function of input image space bandwidth product (SBWP), input SNR, the number of training images N and the distortion in the input image. They approached this by modeling $s(x,y)$ as a sample realization from a random process with a specified Autocorrelation function[40]. Vijaya Kumar and Pochapsky[39] used Gaussian-shaped autocorrelation functions (ACFs) as well as exponential-shaped ACFs and obtained similar results.

We show in Fig. 10 the output Signal-to-Noise Ratio (SNR) as a function of input distortion for three different choices of N (1, 12 and 72). The input image $s(x,y)$ has a space bandwidth product of 1000 and the input SNR was such that the output SNR for the CMF is 10 dB when the input is not distorted.
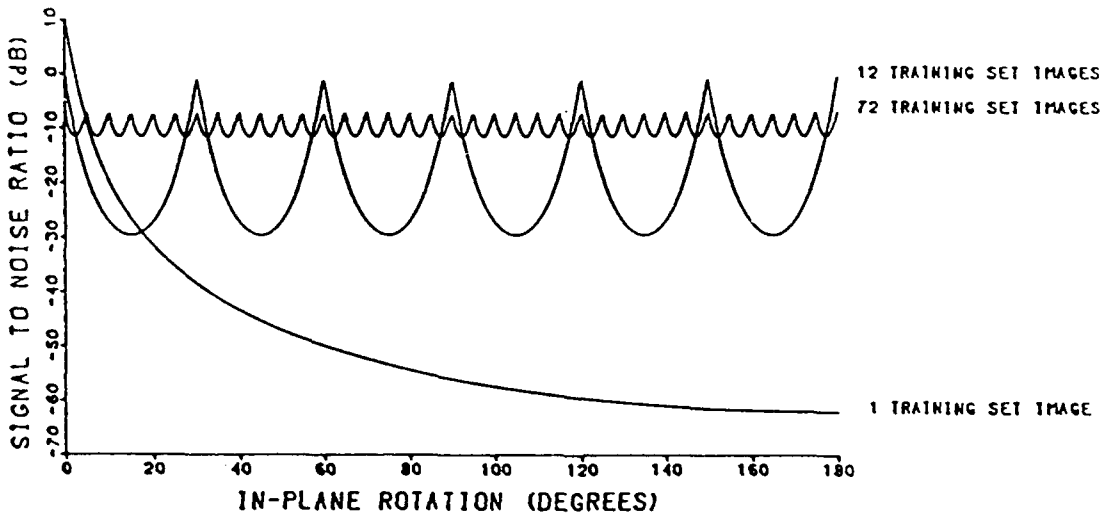

Figure 10: Output SNR as a function of input rotation for different N values. Input image SBWP is 1000 [39].

In this figure, the curve corresponding to one training image shows how the output SNR is affected by in-plane rotation. For no in-plane rotation, the output SNR is 10 dB and it steadily decreases with increasing rotation and becomes as low as -60 dB when the rotation is 180°. The in-plane rotation shown in Fig. 10 is from 0° to 180° only because the remaining rotations can be determined from symmetry.

This curve for "1 training image" clearly demonstrates the distortion sensitivity of the CMF.

In this same figure, we show the output SNR as a function of input rotation for the same image, but using SDFs designed with 12 and 72 training images. When $N = 12$ is used, the output SNR is only 0 dB when the input is undistorted. Similarly, the output SNR is 0 dB when the input is rotated by multiples of 30° (corresponding to one of the 12 training images used). The output SNR reaches its minimum (approximately -30 dB) at $15°, 45°, 75°, \ldots, 345°$. Thus, by using $N = 12$, we improve the worst-case output SNR (from -70 dB for the CMF to -30 dB for the SDF using $N = 12$) while degrading the best-case output SNR (from 10 dB for the CMF to 0 dB for the SDF using $N = 12$). When the number of training images used is increased to $N = 72$, both the best-case and the worst-case output SNRs converge to the value of -10 dB. Any further increase in N is unnecessary since it would not affect the output SNR significantly.

The results in Fig. 10 correspond to an image $s(x, y)$ with SBWP = 1000. Vijaya Kumar and Pochapsky[39] illustrate how output SNR curves are affected by increasing the input SBWP. It is interesting to note that for $N = 1$, the output SNR is higher (for no input distortion) when the input SBWP is higher. But with input distortion, this falls off more rapidly so that the worst-case SNR is usually lower for the image with the higher SBWP. Also, it takes a larger N (for higher SBWP image) to reach the situation where the "best-case" and "worst-case" output SNRs are equal.

Another interesting result that came out of this analysis is the potential danger of using a high SBWP image when N is small. In Fig. 11, we show the worst-case SNR as a function of N for three different input SBWPs. For the smallest SBWP (i.e., 100), the worst-case output SNR levels off for
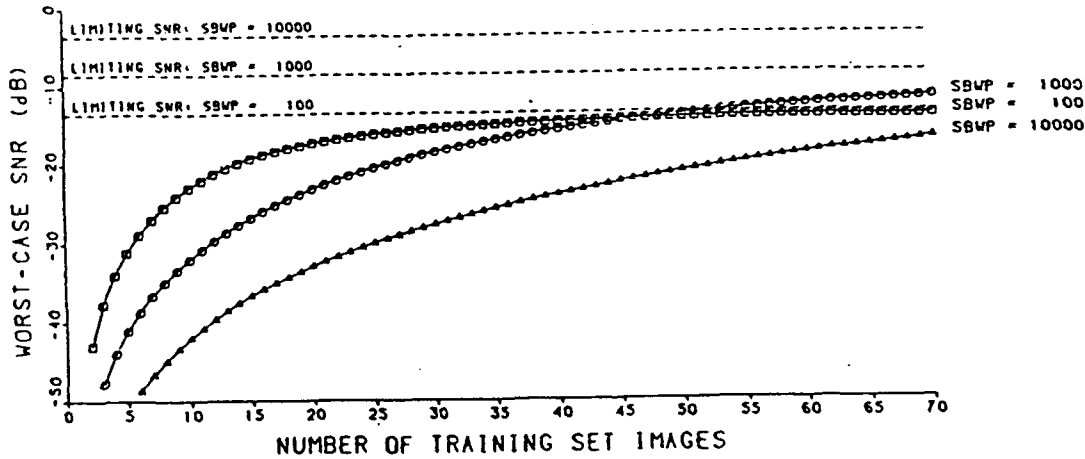

Figure 11: Worst-case output SNR as a function of the number of training images [39].

$N \geq 25$. However, if we use $N = 25$ for an image with SBWP = 10000, the resulting worst-case SNR is significantly lower than that with SBWP = 100 and $N = 25$. Thus, when the number of available training images is limited, it pays to reduce the SBWP of the input images deliberately. In fact, in Fig. 12, we show the optimum SBWP as a function of N, the number of training images. It is gratifying to see that essentially the same curve was obtained for both the Gaussian-shaped and the exponential-shaped ACFs.
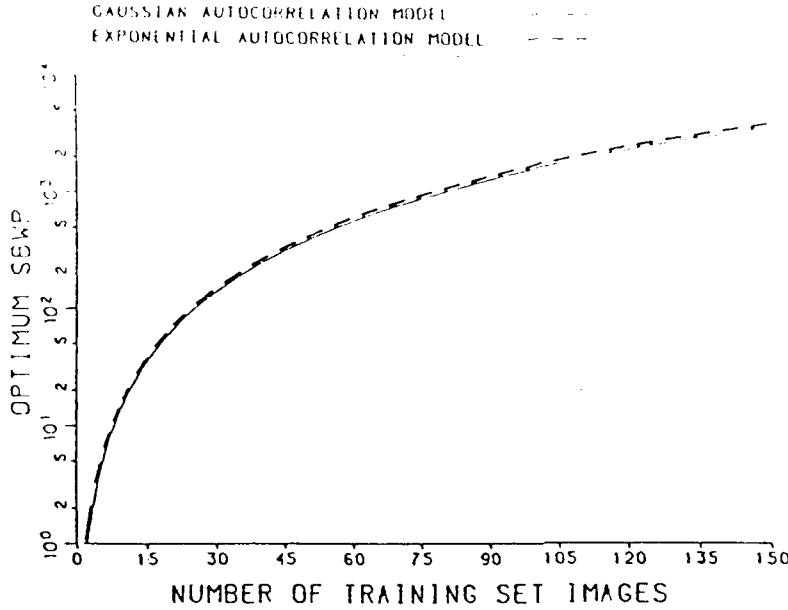
Figure 12: Optimum SBWP as a function of the number of training images [39].

9 Optimal Tradeoff Composite Filters

In Section 4, we discussed OTFs which optimally tradeoff PCE, SNR and light efficiency. In this section, we review some of the recent work by Refregier and Figue[12] in extending the idea of tradeoffs to composite filter design. To introduce this work, we need to introduce vector notation.

Suppose each of the N training images $s_1(x,y), s_2(x,y), \ldots, s_N(x,y)$ can be represented by arrays with d pixels in them. Then we can denote these training images by d-dimensional column vectors $\underline{s}_1, \underline{s}_2, \ldots, \underline{s}_N$. Similarly, the filter $h(x,y)$ can be denoted by the d-dimensional column vector $\underline{h}$. Then the SDF constraints in eq. (7) can be rewritten as

$$\underline{s}_i^T \underline{h} = c_i, \quad i = 1, 2, \ldots, N, \tag{40}$$

where the superscript T denotes the transpose. Eq. (40) represents N linear equations in d variables (i.e., $\underline{h}$). Since $d \gg N$, there are many degrees of freedom available (to optimize other measures) while solving eq. (40).

As discussed in Section 2, the basic SDF approach assumes that $h(x,y)$ is a linear combination of the training images (see eq. (6)) and thus throws away all available degrees of freedom. One problem with the basic SDF is its noise sensitivity. Vijaya Kumar[19] proposed the minimum variance SDF (MVSDF) that uses the additional degrees of freedom to maximize the noise tolerance while satisfying the equations in (40). Another method to take advantage of these additional degrees of freedom is the minimum average correlation energy (MACE) filter developed by Mahalanobis et. al.[20]. Refregier and Figue[12] showed how the two types of filters can be accommodated in a single design.

Suppose the input image is one of the training images corrupted by additive, white noise. Then the

variance σ^2 of the correlation output is given by[19]

$$\sigma^2 = N_o h^T h,$$ (41)

where N_o is the level of the input noise. Minimizing σ^2 in eq. (41) subject to the constraints in eq. (40) leads to

$$h_{MVSDF} = S(S^T S)^{-1} c,$$ (42)

where S contains the N training vectors as its N columns, i.e.,

$$S = [s_1, s_2, \ldots, s_N]$$ (43)

and c is a column vector with the N output constraint values in it, i.e.,

$$c = [c_1, c_2, \ldots, c_N]^T.$$ (44)

With this notation, the N equations in eq. (40) can be written more compactly as

$$S^T h = c.$$ (45)

It is easy to verify that the MVSDF in eq. (42) satisfies the constraint in eq. (45).

While the MVSDF maximizes the noise tolerance, it does nothing to reduce the false peak problem. The false sidelobe/peak problem is attacked by the MACE filter[20] which minimizes the average energy in the correlation plane while satisfying the constraints in eq. (45). The MACE filter is most easily derived in the frequency domain.

Let $\hat{h}$, $\hat{s}_i$ and $\hat{S}$ denote the frequency-domain counterparts of h, s_i and S. Then eq. (45) can be rewritten as

$$\hat{S}^+ \hat{h} = c,$$ (46)

where we have ignored some constants. The superscript $+$ denotes conjugate transpose. The average correlation plane energy E_{avg} is as below.

$$
\begin{aligned}
E_{avg} &= \frac{1}{N} \sum_{i=1}^{N} \iint |H(u,v)|^2 |S_i(u,v)|^2 \, du\, dv \\
&= \iint |H(u,v)|^2 P_s(u,v) \, du\, dv
\end{aligned}
$$ (47)

where the average signal spectrum is given by

$$P_s(u,v) = \frac{1}{N} \sum_{i=1}^{N} |S_i(u,v)|^2.$$ (48)

The average correlation plane energy E_{avg} in eq. (47) can be rewritten in matrix/vector notation as

$$E_{avg} = \hat{h}^+ \hat{D} \hat{h}$$ (49)

where $\hat{D}$ is a $d \times d$ diagonal matrix whose entries are given by $P_s(u,v)$ defined in eq. (48). The MACE filter is obtained by minimizing E_{avg} in eq. (49) while satisfying the constraints in eq. (46). The resulting MACE filter is given[20] by

$$\hat{h}_{MACE} = \hat{D}^{-1} \hat{S} (\hat{S}^+ \hat{D}^{-1} \hat{S})^{-1} c.$$ (50)

The matrix D is diagonal and is easily inverted. It is easy to verify that the filter in eq. (50) satisfies the SDF constraints in eq. (46).

Sudharsanan et. al.[41] suggested a space domain method to design compromise composite filters. In space domain, the matrix equivalent of D is not diagonal and is not convenient. Refregier and Figue[12] showed that the Optimal Tradeoff Composite Filter (OTCF) is given as follows.

$$\hat{h}_{OTCF} = \hat{B}_\mu^{-1} \hat{S}(\hat{S}^+ \hat{B}_\mu^{-1} \hat{S})^{-1} \underline{c}.$$ (51)

where the diagonal matrix $\hat{B}_\mu$ of size $d \times d$ is given by

$$\hat{B}_\mu = \mu \hat{D} + (1 - \mu)\hat{P}_n$$ (52)

where $0 \leq \mu \leq 1$ is a design parameter and $\hat{P}_n$ is the diagonal matrix with the input noise spectral density values.

When $\mu = 0$, the OTCF reverts to MVSDF and when $\mu = 1$, we obtain the MACE filter. For intermediate μ values, we get a compromise filter. Refregier and Figue[12] tested these filters numerically using an image database of 20 trucks and 20 tanks. In Fig. 13, we reproduce the curve from Reference 12 that shows how σ^2 of eq. (41) (termed the Mean Square Error) and E_{avg} of eq. (49) (termed the average correlation energy) vary as μ is varied. The inside box shows the curve in a log scale. Note that both MVSDF and MACE represent rather poor choices since we can improve either of the two performance measures by sacrificing the other by small amounts.
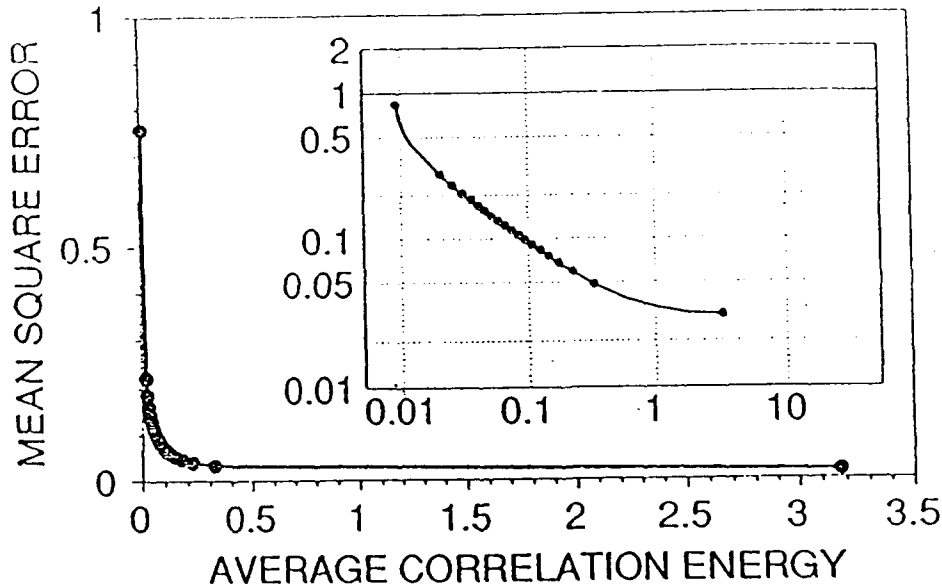

Figure 13: Mean Square Error as a function of the average correlation plane energy for the image database of 20 tanks and 20 trucks [12].

Refregier and Figue[12] make the interesting observation that filters (such as MVSDF) that maximize the noise tolerance appear to be tolerant to distortions, but discriminate poorly. On the other hand, filters (such as MACE) that maximize the peak sharpness tend to discriminate well, but are unacceptably sensitive to even minor distortions. The OTCFs discussed in this section provide a controlled approach to

achieving an acceptable compromise between the two extremes. Refregier[11] suggested a similar optimal tradeoff scheme for designing circular harmonic filters.

10   Conclusions

The objective of this paper is to illustrate the fact that we are always trading off one performance measure against another in correlation filter design. We have discussed the following in order to reinforce this concept.

- Optimal tradeoff filters (OTFs) where we trade off SNR, PCE and η'_H.

- Fractional power filters (FPFs) where we trade off among the three measures by using different powers.

- Region of support (ROS) selection in POFs and BPOFs to trade off SNR and PCE.

- ROS selection to trade off tolerance to input noise for tolerance to detector noise.

- Trading off of the best-case SNR for improved worst-case SNR in designing synthetic discriminant functions (SDFs).

- Optimal tradeoff composite filters (OTCFs) that provide the best compromise between MACE and MVSDF approaches.

It is our hope that the above illustrations will dispel any illusions that filter design schemes exist that are the "best" in all categories. As expected, we must give up in one measure to gain the other. This does not, however, imply that we can randomly choose a filter and use it. Finding a filter that gives the best compromise between two extremes is still a nontrivial task and research into compromise composite filters must be rigorously pursued.

ACKNOWLEDGEMENTS

References

[1] A. Vanderlugt, "Signal detection by complex spatial filtering", *IEEE Trans. Inf. Theory*, Vol. 10, 139-145, 1964.

[2] D. O. North, "An analysis of factors which determine signal/noise discriminations in pulsed-carrier systems", *Proc. of IEEE*, Vol. 51, 1016-1027, 1963.

[3] W. E. Ross, D. Psaltis and R. H. Anderson, "Two-dimensional magneto-optic spatial light modulator for signal processing", *Optical Engineering*, Vol. 22, 485-90, 1983.

[4] J. L. Horner and P. D. Gianino, "Phase-only matched filtering", *Applied Optics*, Vol. 23, 812-816, 1984.

[5] D. Psaltis, E. G. Paek and S. S. Venkatesh, "Optical image correlation with a binary spatial light modulator", *Optical Engineering*, Vol. 23, 698-704, 1984.

[6] B. V. K. Vijaya Kumar, "A tutorial survey of composite filter design techniques", *Applied Optics*, To appear, 1992.

[7] Ph. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak and Horner efficiency", *Optics Letters*, Vol. 16, 829-831, 1991.

[8] B. V. K. Vijaya Kumar and Z. Bahri, "Phase-only filters with improved signal-to-noise ratio", *Applied Optics*, Vol. 28, 250-257, 1989.

[9] B. V. K. Vijaya Kumar and Z. Bahri, "Efficient algorithm for designing ternary-valued filter yielding maximum signal-to-noise ratio", *Applied Optics*, Vol. 28, 1919-1925, 1989.

[10] B. V. K. Vijaya Kumar, W. Shi and C. Hendrix, "Phase-only filters with maximally sharp correlation peaks", *Optics Letters*, Vol. 15, 807-809, 1990.

[11] Ph. Refregier, "Optical pattern recognition: Optimal trade-off circular harmonic filters", *Optics Communications*, To be published, 1991.

[12] Ph. Refregier and J. Figue, "Optimal trade-off filters for pattern recognition and their comparison with Wiener approach", *Optical Processing and Computing*, To be published, 1991.

[13] J. L. Horner and H. O. Bartlet, "Two-bit correlation", *Applied Optics*, Vol. 24, 2889-2893, 1985.

[14] D. M. Cottrell, R. A. Lilly, J. A. Davis and T. Day, "Optical correlator performance of binary phase-only filters using Fourier and Hartley transforms", *Applied Optics*, Vol. 26, 3755-61, 1987.

[15] R. D. Juday, "Optical correlation with a cross-coupled spatial light modulator", *1988 Technical Digest Series*, Vol. 8, 238-241, Optical Society of America, Washington, D.C., 1988.

[16] R. D. Juday, "Correlation with a spatial light modulator having phase and amplitude cross-coupling", *Applied Optics*, Vol. 28, 4865-4869, 1989.

[17] M. W. Farn and J. W. Goodman, "Optimum maximum correlation filter for arbitrarily constrained devices", *Applied Optics*, Vol. 28, 3362-3366, 1989.

[18] C. F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition", *Applied Optics*, Vol. 19, 1758-1761, 1980.

[19] B. V. K. Vijaya Kumar, "Minimum variance synthetic discriminant functions", *JOSA-A*, Vol. 3, 1579-1584, 1986.

[20] A. Mahalanobis, B. V. K. Vijaya Kumar and D. Casasent, "Minimum average correlation energy filters", *Applied Optics*, Vol. 26, 3633-3640, 1987.

[21] Ph. Refregier, "Filter design for optical pattern recognition: Multi-criteria optimization approach", *Optics Letters*, Vol. 15, 854-856, 1990.

[22] B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters", *Applied Optics*, Vol. 29, 2997-3006, 1990.

[23] M. W. Farn and J. W. Goodman, "Optimal binary phase-only matched filters", *Applied Optics*, Vol. 27, 4431-37, 1988.

[24] J. L. Horner, "Metrics for assessing pattern recognition performance", *Applied Optics*, To appear, 1991.

[25] J. L. Horner, "Light utilization in optical correlators", *Applied Optics*, Vol. 21, 4511-4514, 1982.

[26] H. J. Caulfield, "Role of the Horner efficiency in the optimization of spatial filters for optical pattern recognition", *Applied Optics*, Vol. 21, 4391, 1982.

[27] B. V. K. Vijaya Kumar, F. M. Dickey and J. M. DeLaurentis, "Correlation filters minimizing peak location errors", *JOSA-A*, In review, 1991.

[28] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[29] B. Javidi, "Nonlinear joint power spectrum based optical correlation", *Applied Optics*, Vol. 28, 2358-2367, 1989.

[30] F. M. Dickey, K. T. Stalker and J. J. Mason, "Bandwidth considerations for binary phase-only filters", *Applied Optics*, Vol. 27, 3811-3818, 1988.

[31] B. A. Kast, M. K. Giles, S. D. Lindell and D. L. Flannery, "Implementation of ternary phase-amplitude filters using a magneto-optic spatial light modulator", *Applied Optics*, Vol. 28, 1044-1046, 1989.

[32] B. V. K. Vijaya Kumar, W. Shi and C. Hendrix, "Design of partial-information filters yielding maximally sharp correlation peaks", *Optical Computing and Processing*, Vol. 1, 29-46, 1991.

[33] D. Flannery, J. Loomis and M. Milkovich, "Design elements of binary phase-only correlation filters", *Applied Optics*, Vol. 27, 4231-4235, 1988.

[34] C. D. Hendrix, B. V. K. Vijaya Kumar, K. T. Stalker, B. A. Kast and R. Shori, "Design and testing of 3-level optimal correlation filters", *Proc. of SPIE*, Vol. 1564, 1991.

[35] D. L. Flannery and W. E. Phillips, "Case study of design trade-offs for ternary phase-amplitude filters", *Proc. of SPIE*, Vol. 1564, No. 7, 1991.

[36] Z. Bahri, "Phase-only and binary phase-only filters in optical correlators", Ph.D. Dissertation, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, 1989.

[37] B. V. K. Vijaya Kumar, R. D. Juday and P. Karivaratha Rajan, "Saturated Filters", *JOSA-A*, Accepted for publication, October 1991.

[38] R. D. Juday, P. Karivaratha Rajan and B. V. K. Vijaya Kumar, "Some limitations on optical filtering imposed by physical systems", *Proc. of SPIE*, Vol. CR40, No. 14, 1991.

[39] B. V. K. Vijaya Kumar and E. G. Pochapsky, "Signal-to-noise ratio considerations in modified matched spatial filters", *JOSA-A*, Vol. 3, 777-786, 1986.

[40] K. S. Shanmugam and A. M. Breipohl, *Random Signals: Detection, Estimation and Data Analysis*, John Wiley, New York, 1988.

[41] S. I. Sudharsanan, A. Mahalanobis and M. K. Sundareshan, "Unified framework for the synthesis of synthetic discriminant functions with reduced noise variance and sharp correlation structure", *Optical Engineering*, Vol. 29, 1021-1028, 1990.

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.