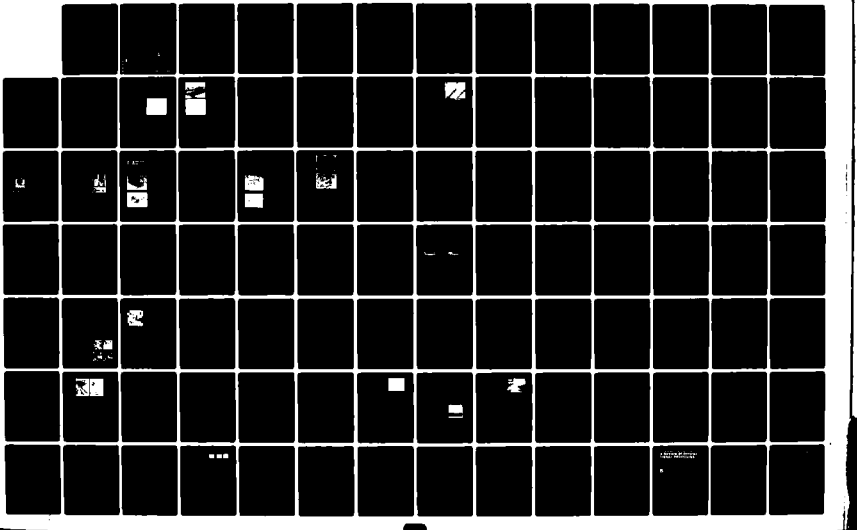
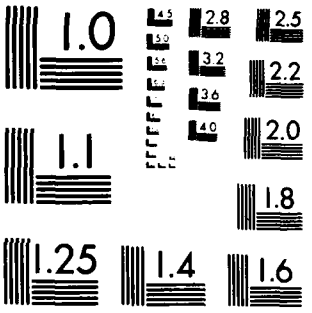
AD-A130 097 OPTICAL PATTERN RECOGNITION FOR MISSILE GUIDANCE(U) 1/2
CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF ELECTRICAL
ENGINEERING D CASASENT 15 NOV 82 AFOSR-TR-83-0556
UNCLASSIFIED AFOSR-79-0091 F/G 17/7 NL





AFOSR-TR- 83-0556

AD A130057

FILE COPY

DTIC
ELECTE
JUL 5 1983
S
A

Approved for public release;
distribution unlimited.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR- 83-0556	2. GOVT ACCESSION NO.: AD-A130097
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Optical Pattern Recognition for Missile Guidance	5. TYPE OF REPORT & PERIOD COVERED: Interim Report September 1981 - October 1982
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): David Casasent	8. CONTRACT OR GRANT NUMBER(s): AFOSR-79-0091
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, PA 15213	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 6-110 2F 2305/B1
11. CONTROLLING OFFICE NAME AND ADDRESS: AFOSR/NE, Building 410, Bolling Air Force Base; Washington, D.C 20332	12. REPORT DATE: 15 November 1982
	13. NUMBER OF PAGES: 102
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

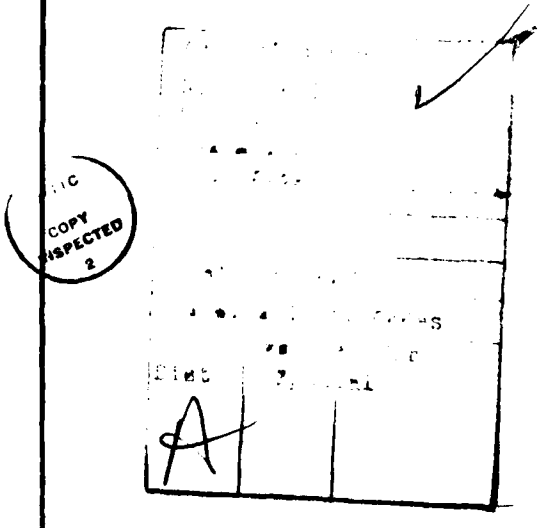
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Adaptive phased array radar processing, discrimination, eigenvalue calculation, holographic pattern recognition, iterative optical processor, pattern recognition, principle component imagery, statistical synthetic discriminant function, PRIZ light modulator, synthetic discriminant functions.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Progress on real-time spatial light modulators, image pattern recognition and optical signal processing for missile guidance is documented. A full description of our test and evaluation of the Soviet PRIZ spatial light modulator is included. In image pattern recognition, a unified formulation of four different and new types of synthetic discriminant functions is advanced. These include synthetic discriminant functions for intra and inter-class pattern recognition and multi-class pattern recognition. In the area of image pattern recognition, we also advance new statistical synthetic discriminant function filter concepts

and a new principal component synthetic discriminant function. These analyses utilize new performance measures and new image models. Conventional holographic pattern recognition research conducted under AFOSR support is also reviewed. Our new AFOSR optical signal processing research concerns optical matrix-vector processors. Initial research in this area includes fabrication of a fiber-optic microprocessor-based iterative optical processor and its use in adaptive phased array radar processing and for the calculation of eigenvalues and eigenvectors of a matrix.



ABSTRACT

Progress on real-time spatial light modulators, image pattern recognition and optical signal processing for missile guidance is documented. A full description of our test and evaluation of the Soviet PRIZ spatial light modulator is included. In image pattern recognition, a unified formulation of four different and new types of synthetic discriminant functions is advanced. These include synthetic discriminant functions for intra and inter-class pattern recognition and multi-class pattern recognition. In the area of image pattern recognition, we also advance new statistical synthetic discriminant function filter concepts and a new principal component synthetic discriminant function. These analyses utilize new performance measures and new image models. Conventional holographic pattern recognition research conducted under AFOSR support is also reviewed. Our new AFOSR optical signal processing research concerns optical matrix-vector processors. Initial research in this area includes fabrication of a fiber-optic, microprocessor-based iterative optical processor and its use in adaptive phased array radar processing and for the calculation of eigenvalues and eigenvectors of a matrix.

TABLE OF CONTENTS

(Continued)

PAGE

1. INTRODUCTION

During the past year, our research in optical data processing for missile guidance has addressed various new algorithms, system architectures, component tests and analyses of various image data bases. As in past years, we have been quite faithful in reporting our AFOSR sponsored research in various journal and conference publications. Copies of the more relevant of these papers are thus included as the chapters of this report to provide concise documentation of our work.

In Section 2, we provide a summary and overview of our research progress made in the past year. Details on the more salient topics of our research are provided in Sections 3-14. In Section 15, we enumerate our AFOSR sponsored research publications, the presentations given on this research in the past year, and the PhD, Master's and research associates that this grant has supported.

2. SUMMARY AND OVERVIEW

2.1 SOVIET PRIZ SPATIAL LIGHT MODULATOR (CHAPTERS 3-5)

A unique opportunity arose for us to conduct a four month test and evaluation of a Soviet light modulator. We immediately accepted the offer. Our results are contained in Chapters 3-5. Five Soviet devices for use as real-time and reuseable spatial light modulators were evaluated in all. These devices were bismuth silicon oxide (BSO) Prom-type devices. The more interesting devices used a new transverse electro-optic effect. In Chapter 3, we describe our initial results. In Chapter 4, we detail our MTF and sensitivity tests on these devices. In general, we found the transverse weighted BSO PRIZ device to have significantly better sensitivity and resolution that the conventional longitudinal electro-optic U.S. Prom device. In Chapter 5, we detail the unique applications of this device in image and signal processing. Its unique features include: automatic edge-enhancement and the ability for dynamic image selection. The latter property is a most unique and attractive feature. Further U.S. research on the Soviet devices is necessary to facilitate fabrication of equivalent U.S. PRIZ devices.

2.2 IMAGE PATTERN RECOGNITION (CHAPTERS 6-10)

As in past years, the major emphasis of our optical data processing for missile guidance research has been directed to novel pattern recognition algorithms (that allow pattern recognition and object classification in the face of various geometrical and textural differences between the input and reference image) as well as the realization of hybrid optical/digital architectures to implement these algorithms. Our image processing research during the past year

has continued to emphasize pattern recognition. Our major pattern recognition technique we are investigating involves the use of synthetic discriminant functions. During the past year, we have devoted increasing attention to discrimination and multi-class pattern recognition with emphasis on maintaining object recognition and correct classification in the presence of various distortions in the imagery.

In Chapter 6, we describe a new statistical two-class SDF. This SDF differs from others: (a) in the performance measure optimized (we use SNR of the output correlation at the peak point, rather than various distance measures as used in other approaches); (b) its application to discrimination (we defined the problem as distinguishing two classes of functions rather than extracting a given function from noise); (c) in our image model (we assume both classes of functions are stochastic processes. Note that the noise is now non-stationary since it is an image with its own spatially varying statistics); (d) in our use of a finite K-L expansion (we approximate both classes of functions by finite K-L expansions of two basis function sets); (e) in our representation of randomly distorted imagery by a basis function expansion with random coefficients; (f) in our basis function set selection (we derive the SDF as a linear combination of the two basis function sets obtained from the finite K-L expansion); and (g) our new iterative digital eigenvalue solution that results (this modified iterative K-L technique is shown to provide results similar to those obtained from the exact but very computationally expensive K-L method). We demonstrate our results for the case of geometrical distortions by considering the discrimination between circles and hexagons of random diameters. Further work is necessary on selection of the parameter γ to increase discrimination but retain a large

correlation peak to ease detection. We hope to be able to continue this research in FY84.

In Chapter 7, we summarize our recent deterministic training set based SDF research. This pattern recognition technique is very attractive since it allows off-line calculation and determination of a pattern recognition filter. The filter is as good as the training set and is appropriate for object recognition as long as the object being searched for (or the other objects possibly present in the image) change. As with all SDFs, this algorithm allows geometrical distortion-invariant pattern recognition. Since it is realized with a correlator, it offers a processing gain and thus functions well in the presence of noise. In our new research, we have unified four different types of SDFs into the same general formulation. This is attractive because of the intense computations necessary and the many matrix operations required in the off-line SDF synthesis. The four new SDFs we have devised include: an equal correlation peak (ECP) SDF (this is appropriate for intra-class pattern recognition; a mutual orthogonal function (MOF) SDF (this is appropriate for inter-class discrimination); a non-redundant filter (NRF) SDF (this is appropriate for multi-class pattern recognition using fewer filters) and a multi-class MOF SDF (this uses more SDFs than does the NRF, but generally functions better). Combinations of inter and intra-class pattern recognition and NPF and multi-class MOFs are subjects of further research, together with the testing and application of these algorithms on various image data bases.

In Chapter 8, we describe a principal component statistical SDF. In this case, we address geometrical distortions and intensity differences as arise in multisensor imagery. We use new image models in which image differences are

modeled as a stochastic process, derive an optimal filter that maximizes correlation SNR and show that it is the principal component or dominant eigenvector in a K-L expansion of the data. As in our other cases, we use a correlator in our final data. Attention is given to the fact that the output has a deterministic and a noisy part. The optimal filter is shown to be equal to a maximum likelihood detector. As the model for our statistical correlator, we use the class of data $\{x(t)\}$ which can only be modeled statistically We derived time-averaged and ensemble-averaged correlation functions for th stochastic process $\{x(t)\}$. These are new statistical averaging techniq .. To derive the optimal SDF, we expand the ensemble-averaged correlation function of the stochastic process in a K-L basis function set. The maximum SNR is shown to result if the dominant eigenvector of the kernel of the ensemble-averaged correlation function is chosen. We demonstrate our results on multisensor images. We compare our results for this new principal component SDF to those obtained using an ECP SDF. We find that requiring optimal SNRs for the correlation yields larger correlation SNR than are obtained from the ECP SDF. Registration accuracy in the location of the correlation peak were also considered. Both filters were found to give perfect results, whereas normal MSFs gave large errors. We also considered the 3dB area of the correlations produced and found small not wide correlation peak widths very close to those obtained with autocorrelations (thus providing excellent registration accuracy and essentially perfect equivalence to the results obtained from an autocorrelation, even though geometrical and intensity distortions were present in the data set).

In Chapter 9, we discuss a modified hyperplane technique for beam forming. This represents an efficient technique (in terms of computer time) for handling

large matrices. We later plan to extend and apply this algorithm to our SDF synthesis work. This is most necessary, because of the large data sets and matrices involved in our off-line SDF synthesis technique.

In Chapter 10, we provide a summary and overview of our AFOSR holographic pattern recognition research. This includes the conventional matched spatial filter, the use of holographic optical elements, laser diodes, a lensless matched spatial filter correlator, weighted matched spatial filters, Mellin transform correlators, hybrid optical/digital pattern recognition systems and synthetic discriminant function matched filters.

2.3 OPTICAL SIGNAL PROCESSING (CHAPTERS 11-14)

A supplemental AFOSR grant enables us to complete earlier research on an iterative optical processor (IOP) for adaptive phased array radar, to extend this system as a general purpose optical processor and to devise new missile guidance applications of it. In Chapter 11, we describe the first laboratory IOP ever fabricated. It uses fiber-optic interconnections, microprocessor control and a 2-D fixed mask. In Chapter 12, we summarize its applications in adaptive phased array radar and in Chapter 13 we note its use as a general purpose processor to perform the very common operation of eigenvalue and eigenvector computations. A summary and overview of the use of optical techniques in signal processing is included as Chapter 14.

3. SOVIET PRIZ SPATIAL LIGHT MODULATOR

Reprinted from **APPLIED OPTICS**, Vol. 20, page 3090, September 15, 1981

Soviet Priz spatial light modulator

D. Casasent, F. Caimi, and A. Khomenko

A. Khomenko is with Ioffe Institute, Shuvalov Laboratories, Leningrad, U.S.S.R.; the other authors are with Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, Pennsylvania 15213.
Received 11 July 1981.
Sponsored by H. J. Caulfield, Aerodyne Research, Inc.
0003-6935/81/183090-03$00.50/0.

Optical data processing systems require 2-D spatial light modulators (SLMs)[1] that are capable of converting an input optical (or electrical) signal in real-time into a transparency suitable for spatial modulation of a collimated laser beam. These devices must operate in real-time and be reusable. A well-documented candidate SLM is the Prom.[2] It uses a bismuth silicon oxide (BSO) crystal that is both photoconductive and electrooptic. In operation, the Prom is illuminated with white light of wavelength λ_W (350–450 nm). The spatial intensity distribution of this λ_W light is converted to a spatial charge pattern in the BSO. When read out with light of wavelength λ_R (633 nm usually), the amplitude of the readout beam is spatially modulated corresponding to the intensity modulation present on the λ_W light pattern.

In the Prom, modulation of the λ_R light occurs by the longitudinal electrooptic effect (the direction of light propagation and the applied electric field are colinear). A [100] cut BSO crystal is used to optimize the corresponding electrooptic tensor coefficients for the longitudinal electrooptic effect. The sensitivity of the Prom is low by comparison with other candidate SLMs.[1] Furthermore its frequency response rapidly decreases with increasing spatial frequency f. The diffraction efficiency is proportional to $1/f^2$ with low f and to $1/f^4$ for high f with a 3-dB point at about 4–5 cycles/mm.[3-6] This performance occurs because the spatial charge layer induced in the Prom by the λ_W light lies within the volume of

the material.[4-6] The electric field has opposite signs on each side of the charge layer, and when integrated across the crystal's thickness a cancellation occurs, low phase modulation results, the magnitude of the electrooptic effect is reduced, and poor sensitivity and spatial frequency result.[4-6] Thus, although the optical quality and potential resolution of the device is quite good, the available light intensity at high spatial frequencies is very low. This limits its use in many practical systems.

The Soviets have fabricated Prom light modulators comparable in quality and performance to the U.S.A. units.[3,4] Recently, they developed the Priz (a Soviet acronym) light modulator.[7,8] It uses a BSO crystal, but its fabrication and insulating layers differ from those in the Prom. For the present discussion, the major issue of concern is that the BSO crystal used in the Priz light modulator has a [110] or [111] cut. This crystal orientation provides a large transverse electrooptic effect and a low longitudinal effect. The Priz is still operated with the applied electric field and the light propagation directions collinear. As in the Prom, the spatially modulated λ_W light beam induces a spatially varying charge pattern in the crystal. However, the charge distribution produced by the longitudinal electric field also has a transverse component. It is this spatially varying transverse charge distribution that produces light modulation in the Priz.

We were fortunate to have recently had five Soviet Prom and Priz light modulators in our laboratories and A. Khomenko, a Soviet scientist who was actively engaged in fabrication and research on these devices. The Priz device differs totally from the French BSO device[9] used for nonlinear optical processing in which the direction of light propagation and the applied electric field are orthogonal. In the Priz, the transverse field (due to the spatial voltage distribution induced by the λ_W light pattern) provides the electrooptic modulation. Because the electric field on both sides of the charge layer is of the same sign in the Priz device, no cancellation of the phase modulation occurs as is the case in the Prom and hence, better device sensitivity results. A detailed resolution analysis with proper modeling of the charge layer within the BSO shows that diffraction efficiency (η) for the Priz decreases at high spatial frequencies $\propto 1/f^2$ rather than $1/f^4$ (as occurs for the Prom).

An interferometric recording system was used to image sine wave fringe patterns on the U.S.A. Prom and the Soviet Priz in λ_W = 476 nm light from an Ar-ion laser. By varying the angle of one of the mirrors in the interferometer, the input spatial frequency recorded on the light modulator was varied from 1 to 60 cycles/mm. The Fourier transform of the fringe pattern was read out in λ_R = 633-nm light from a He–Ne laser, and the intensity in the first and second diffracted orders was measured. Both devices were operated dynamically in real-time at a 2-frame/sec write/read/erase cycle rate with writing and reading being simultaneous and with a 1-msec erase time. Faster cycle times are possible. (In other tests, we operated the device at 30 frames/sec.) The input write light exposure for the data shown in Fig. 1 was E_W = 50 μJ/cm^2 (for both devices). The read light intensity (0.33 mW) was adjusted to insure that the stored pattern on the SLM was not degraded by the λ_R light. From the time history output of the first-order intensity I_1, the diffraction efficiency $\eta = I_1/I_{in}$ vs f was measured. As noted in Ref. 10 unnormalized data are shown to describe more properly the true amount of usable output light. As also noted in Ref. 10 the second-order intensity I_2 in the FT plane was monitored, and the write light intensity and its duration were adjusted to insure that I_2 was over 20 dB below I_1. With these procedures, the spatial frequency response data shown in Fig. 1 correspond to a nonlinearity of $<1\%$.

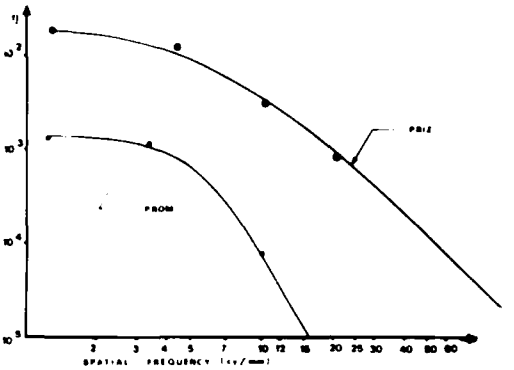


Fig. 1. Diffraction efficiency η vs spatial frequency f response for Prom and Priz spatial light modulators.

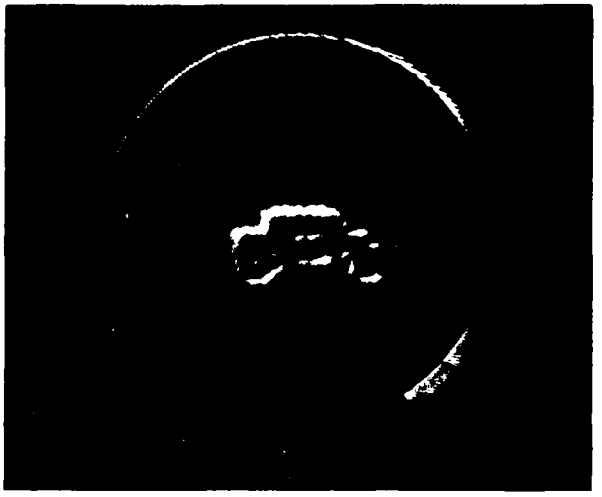
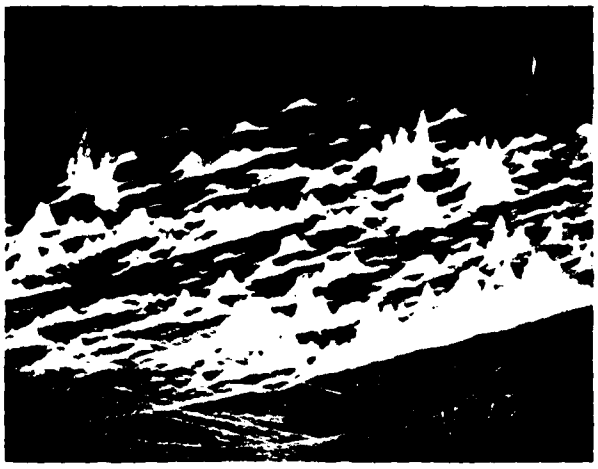


Fig. 2. Real-time image reconstruction showing differentiated version of an IR image recorded on the Priz spatial light modulator.

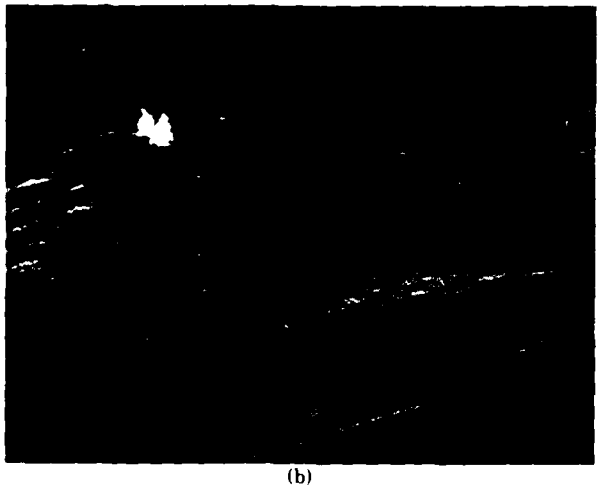
As seen in Fig. 1, the Priz has more useful light at all spatial frequencies (by a factor of 10–100) than the Prom. From Fig. 1, we also see that the rate at which η decreases with frequency is much faster for the Prom ($\eta \propto 1/f^4$) than for the Priz ($\eta \propto 1/f^2$).

The Priz exhibits several other unique features such as automatic dc suppression. This follows directly from the observation that for a uniform charge layer there is no transverse field, and thus the Priz has no dc response. This edge enhancement feature of the device is quite useful in IR image pattern recognition as a preprocessing operation. In Fig. 2 we show the differentiated version of an IR image reconstructed after recording on the Priz. The reconstruction and the image differentiation shown were performed in real-time directly on the device. When linearly polarized read light is used, the Priz is also able to suppress input spatial frequencies oriented in selected directions.[7]

A new and most attractive feature of the Priz is that it can be operated with a fixed applied voltage between electrodes (rather than a voltage switched between positive and negative polarities, as in the Prom). In this operating mode, the re-



(a)



(b)

Fig. 3. Dynamic image selection using the Priz (a) original input (moving spot plus fixed noise background) and (b) reconstruction of only the dynamic portion of the image.

sponse of the Priz is a function of the velocity with which the input data move across the input plane.[7] In our tests on the Priz at CMU, we used this feature of the Priz for change detection. Specifically a moving spot was superimposed on a fixed background noise pattern [Fig. 3(a), the moving spot is in the back left of the figure]. This image was obtained with the high voltage applied to the Priz pulsed between positive and negative polarities in the normal Prom operating mode. When the applied voltage was constant, only the moving spot appeared in the reconstruction. Figure 3(b) shows a snapshot view of such a reconstructed image.

In a later publication, we shall describe in more detail the Priz and our complete test program on these Soviet devices and provide more complete demonstrations and discussions of the unique features of the Priz such as directional filtering and dynamic image subtraction.

We thank Michael Petrov of Ioffe Institute for allowing us to perform this test and evaluation on the Soviet light modulators fabricated in his laboratory and for permitting A. Khomenko to assist us in this test and evaluation program. We also thank the U.S. Air Force Office of Scientific Research (grant AFOSR 79-0091) for supporting publication of this paper.

References

1. D. Casasent, Proc. IEEE, **65**, 143 (1979).
2. B. Horwitz and F. Corbett, Opt. Eng. **17**, 353 (1978).
3. M. Petrov *et al.*, Mikroelektronika **8**, 20 (1979).
4. M. Petrov *et al.*, Ferroelectrics **22**, 651 (1978).
5. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **202**, 110 (1979).
6. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **218**, in press (1980).
7. M. Petrov, at Conference on 4-D Optics Mexico (1980).
8. M. Petrov *et al.*, Sov. Phys. Tech. Phys. **25**, 752 (1980).
9. J. P. Huginard, J. P. Herriau, P. Aubourg, and E. Spitz, Opt. Lett. **4**, 21 (1979).
10. D. Casasent, Appl. Opt. **18**, 2445 (1979).

4. TEST AND EVALUATION OF THE SOVIET PROM AND PRIZ SPATIAL LIGHT MODULATORS

Reprinted from **Applied Optics**, Vol. 20, page 4215, December 15, 1981

Test and evaluation of the Soviet Prom and Priz spatial light modulators

D. Casasent, F. Caimi, and A. Khomenko

Five Soviet bismuth silicon oxide Prom and Priz spatial light modulators were recently tested in the United States. In this program, their performance was quantified and compared with that of the U.S. Prom. The resultant laboratory data show that the Soviet Prom is comparable with the U.S. device and that the Soviet Priz has over ten times the diffraction efficiency and over ten times more usable resolution than the Prom. Theoretical models of the dependence of diffraction efficiency on spatial frequency were also verified by experiments performed on these devices.

I. Introduction

The key device needed for the practical realization of a real-time optical data processing system is the input transducer. This device converts an ambient scene (or an electronic input signal from a mosaic array or other sensor) into a transparency suitable for spatial modulation of a coherent light beam. Such elements are referred to as spatial light modulators (SLM). Many candidate reusable devices exist[1] that can perform the desired operations with varying degrees of resolution and speed. One such candidate SLM consists of a bismuth silicon oxide (BSO) crystal with parylene insulating layers and electrodes on its large faces. This SLM is known as the Prom (Pockels real-time optical modulator).

The Prom is well-documented.[2-4] In operation, an image is focused onto the device in write light (λ_W = 350–450 nm), photocarriers are generated within the photoconductive BSO, and a charge layer is produced with a spatial charge variation proportional to the intensity distribution of the λ_W light. This stored pattern is then read out in read light (λ_R = 633 nm). By the linear longitudinal electrooptic or Pockels effect, the output λ_R light distribution (between crossed polarizers) is a coherent spatially modulated replica of the λ_W light pattern.

In this paper, we report the result of a recent test and evaluation program conducted in our laboratories at Carnegie-Mellon University on five Soviet BSO light modulators. One of these devices was a Soviet Prom,[5,6] and the others were new Priz (Soviet acronym)[7,8] spatial light modulators. Many descriptions, models, and theoretical derivations of the MTF of the Prom exist.[2-6,9,10] Because of the discrepancies in several of these models, a brief unifying summary of the correct model and the predicted spatial frequency response of the Prom are included in Sec. II. This description also provides an excellent vehicle from which to describe the Priz light modulator and the motivation for its development. In Sec. III, we describe the experimental test system used. All tests were performed in real time using a new data acquisition technique to obtain spatial frequency response information on the devices. The results of our experimental program are presented and discussed in Secs. IV and V. We include comparison data on a U.S. Prom we have and on the Soviet Prom in Sec. IV and on the Soviet Priz SLMs in Sec. V. This represents the first time that this many light modulators were tested in the same facility and using the same personnel, measurement techniques, and optical system. We then summarize our results and advance our conclusions in Sec. VI.

II. Prom Theory and Model

The thickness and location of the induced charge layer formed within the BSO by the spatially modulated λ_W light are most important in determining the device's performance. If the charge layer is assumed to be thin and to occur at the crystal (BSO)–dielectric (parylene) interface, the resultant modulation or phase modulation $\Delta\phi$ (between the ordinary and extraordinary wave components) is[11]

A. Khimenko is with Ioffe Institute, Shuvalov Laboratories, Leningrad, U.S.S.R.; the other authors are with Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, Pennsylvania 15213.

Received 24 July 1981.
0003-6935/81/244215-06$00.50/0.

$$\Delta\phi = \frac{d_d\epsilon_d + d_c\epsilon_c}{d_d\epsilon_c(2\pi f)[\epsilon_d \coth(2\pi f d_d) + \epsilon_c \coth(2\pi d_c)]}. \quad (1)$$

where ϵ_d and ϵ_c are the dielectric constants of the parylene in the BSO crystal, respectively, d_d and d_c are the thicknesses of the dielectric layer and the crystal, and f is the spatial frequency of the input data. Since $d_c \simeq 500\ \mu$m and $d_d \simeq 5\ \mu$m, $d_c \gg d_d$, and we approximate $\coth(2\pi f d_c) \simeq 1$, where the approximation will be valid for spatial frequencies $f > 1/2\pi d_c \simeq 0.3$ cycle/mm. From Eq. (1), we see that this model predicts that $\Delta\phi$ should decrease proportionally to $1/f$. Since diffraction efficiency $\eta \propto J_1^2(\Delta\phi)$, where J_1 is a Bessel function of the first order and since $J_1 \propto \Delta\phi$ for low modulation (as occurs in BSO devices), we find $\eta \propto (\Delta\phi)^2$, or from Eq. (1), $\eta \propto 1/f^2$.

In 1978, Petrov *et al.*[5] noted that the charge layer in the Prom was not infinitely thin but had a small but finite thickness d_a. For this model, the resultant dependence of $\Delta\phi$ on spatial frequency f is[5]

$$\Delta\phi = \frac{\sigma_0 n_0^3 r_{41}[1 - \exp(-2\pi f d_a)]}{(2\pi f)^2 d_a \epsilon_0[\epsilon_d \coth(2\pi f d_d) + \epsilon_c \coth(2\pi f d_c)]}. \quad (2)$$

For low spatial frequencies, this model predicts $\Delta\phi \propto 1/f$ and $\eta \propto 1/f^2$. But for higher spatial frequencies ($f > 1/2\pi d_d \simeq 30$ cycles/mm), Eq. (2) predicts $\Delta\phi \propto 1/f^2$ and hence $\eta \propto 1/f^4$. Our experimental data (Sec. IV) confirm these theoretical predictions and are in agreement with earlier Soviet data.[5] Independently, Tanguay and his co-workers in the U.S.A. have subsequently provided a more complete and detailed derivation of the performance of the Prom device.[9,10] This analysis agrees with that predicted by Eq. (2). In this latter work as in the Soviet papers, it was noted that the charge layer does not exist at the crystal–dielectric interface, but rather it is within the volume of the BSO material (but close to the interface). Itek data[4] also note that the charge layer lies within 50 μm of the interface.

The location of the charge layer and its thickness are important because it affects the MTF and spatial frequency response of the device. Moreover, it provides direction for the improved Priz SLM. In Ref. 12, Khomenko *et al.* discuss this subject. The following simplified description will suffice. We recall[13] that the effective extent of the electric field about a sinusoidal charge layer of spatial frequency f is $z_e = 1/2\pi f$. Thus for low spatial frequencies, z_e is large, and at high spatial frequencies it is small. We next recall that $\Delta\phi$ is proportional to the integral of the electric field E in the z direction (optical axis, light propagation direction, and electric field direction). Now, if the charge layer lies within the volume of the BSO, the + and − fields on each side of it will cancel when E is integrated in the z direction. This will decrease the effective $\Delta\phi$ and hence the light modulation and η of the device. Since z_e is larger at low spatial frequencies, we expect the net $\Delta\phi$ to be a maximum at low spatial frequencies and to decrease rapidly as spatial frequency increases. The theoretical and observed $1/f^4$ dependence of η with spatial frequency verifies these remarks.

To overcome these fundamental limitations of the Prom SLM, the Soviets changed the BSO crystal cut from the [100] cut used in the Prom (to maximize the longitudinal electrooptic effect) to [111] and [110] cut crystals (to use the transverse electrooptic effect). These new devices are referred to as Priz (the translation of the associated Soviet acronym is image transformer). In these Priz devices, the optical axis and *the applied electric field* are still collinear, and a spatial charge layer proportional to the λ_W light intensity is still produced. However, the transverse field due to this charge distribution is used (via the transverse electrooptic effect) to produce modulation of the λ_R readout light. The modulation and the resultant η due to the transverse electrooptic effect are larger than for the longitudinal effect because $\Delta\phi$ cancellation does not occur (since the electric field is of the same sign on both sides of the transverse charge) and because of the larger transverse electrooptic coefficients. In the Priz, $\Delta\phi \propto 1/f$ and $\eta \propto 1/f^2$. Thus the spatial resolution as well as the sensitivity of the Priz are expected to be superior to those of the conventional Prom device. These remarks are verified in our experimental data in Secs. IV and V.

III. Experimental Procedure

The five Soviet BSO light modulators we evaluated are referred to as devices 1–5. Device 1 was a Soviet Prom. Device 2 used the conventional Prom fabrication and structure, but the BSO crystal cut used was [111]. Devices 3 and 4 were identical Priz devices with a [111] cut BSO crystal but used a proprietary new fabrication technique. We refer to them (and device 5) as SLMs with a Priz structure. Device 5 was a Priz with a [110] cut BSO crystal. The BSO layers were ~400 μm thick, and each device was 18 mm in diameter. Thin (3–5-μm thick) dielectric layers were used on all devices.

The test system used is shown schematically in Fig. 1. It consists of an interferometer in which the write light beam I is split into two beams by the beam-splitting cube (BSC), reflected from mirrors (M_1 and M_2), and focused at P_0 to form a sine wave fringe pattern.

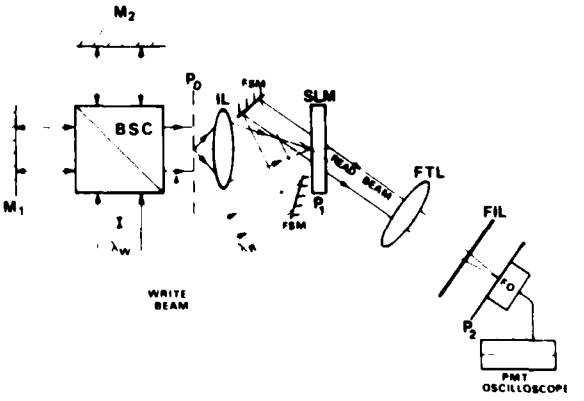


Fig. 1. Schematic diagram of the Prom and Priz spatial frequency response test and evaluation system.

By varying the angle of M_2, different spatial frequencies can be directly produced at P_0. This sine wave fringe pattern at P_0 is then imaged (by lens IL) onto the SLM under the test at P_1. This imaging interferometer system greatly reduces problems that arise with other configurations in which the overlap in the beams from the two interferometer arms varies with the spatial frequency of the signal being recorded. The depth of field of the imaging lens IL used was over 2 mm (over four times the thickness of the SLM used). Thus we were insured that the interference fringe pattern was focused at the optimum location within the SLM.

The write light beam was incident on-axis, and the read light was incident at an angle of 1.7° (This is exaggerated in Fig. 1 to clarify the schematic.) In all experiments, λ_W = 476 nm (Ar-ion laser), λ_R = 633 nm (He–Ne laser), the write and read intensities used were I_W = 0.33 mW, and I_R = 0.52 mW, the write light energy was varied from E_W = 25–200 μJ/cm^2 by varying the write time T_W. The read light intensity I_R was adjusted to insure that it produced no degrading effects on the recorded data. In the way in which the system was used, both the read and write light beams were on simultaneously ($T_R = T_W$, equal write and read light time durations). This facilitated use of a novel data acquisition technique for spatial frequency response information collection. The write light intensity I_W was constantly monitored to insure that no fluctuations occurred in it during the experimental data taking interval.

The sine wave data recorded on the SLM at P_1 are Fourier transformed by lens FTL. [The 633-nm filter FIL insures that only λ_R light reaches the detector or Fourier transform (FT) plane P_2.] At P_2, we measure the dc I_0, first-order I_1, and second-order I_2 light intensities as a function of time. As noted in Ref. 14, the input λ_W light can be adjusted to insure that I_2 is 20 dB below I_1. (This insures that the spatial frequency response data recorded has a nonlinearity below 1%.) In the actual way in which data were acquired in the experiments to be described here, we monitored I_2 and verified that it was at least a factor of 10 below I_1 (in most cases). Since the write and read light beams are on simultaneously in this system, the I_1 vs time output at P_2 rises with time (Fig. 2) as the energy of the write beam accumulates on the SLM. An erase pulse then occurs, and the output rapidly drops to zero as shown in Fig. 2.

Since I_1 is $\sim 10^{-2}$ or 10^{-3} below the intensity of the dc term and since I_2 is over 10^{-2} below I_1, a large output dynamic range is required for measurements. This was achieved by use of a cooled PMT in its photon counting mode (for low light intensity data acquisition) and in its electrometer mode (for higher intensity light data). Calibration between both modes was achieved by use of an attenuator in the readout beam. The PMT system has a 90-dB dynamic range, and with the use of attenuators this range is even greater. In low light level cases, the response time of the PMT ($T_P \simeq 10$ msec) limits the speed at which the data acquisition system can be operated and hence the SLM cycle times that could be used. In all tests, we were also careful to allow sufficient time for charge redistribution to occur in the device after each cycle. In all tests, we insured that the write time T_W was much greater than T_P to reduce detector response time problems. To photograph the output I_1 vs time display, we note that $T_W > 100$ msec is necessary (to reduce detector response time errors), and that to decrease flicker in the output observations on the scope T_W should be <0.5–1.0 sec. As a compromise, we selected $T_W = T_R \simeq 0.5$ sec (or 50 times T_P) for all experiments. This figure was obtained after many experiments with data recorded on the device in the 50–60-cycle/mm spatial frequency range. Thus all our MTF data were acquired dynamically with the SLM operated at about a 2-frame/sec write/read/erase rate. An erase pulse duration of 1 msec was used. This was followed by a 9-msec wait time to insure that full charge redistribution occurred. These parameters provided adequate variations of E_W and adequate user interaction time for photographing the I_1 vs time outputs and monitoring the input light intensity I_{in} and the amount of second-order distortion I_2 present in the data. We note that complete erasure was insured in all cases by proper timing and write and erase exposure selection. After erasure, we measure over 10^3 suppression of the recorded data.

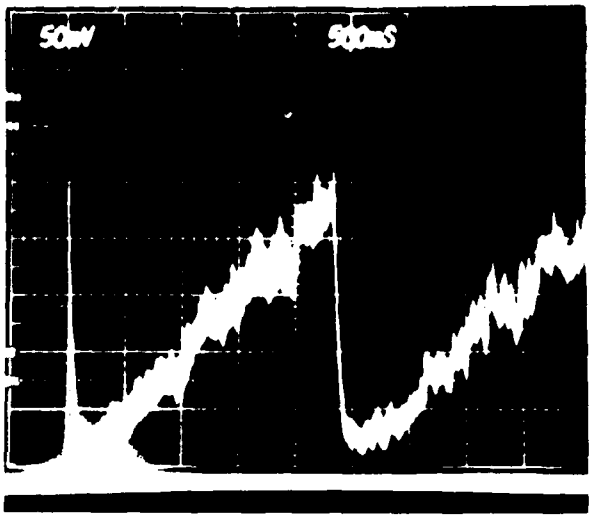


Fig. 2. Typical first-order I_1 vs time output waveform showing the effects of increasing the write light exposure as well as the effectiveness of erasure on the device.

We now consider the data acquisition technique used. Photographs of the I_1 vs time outputs for all five Soviet BSO devices and for the U.S. Prom were obtained for approximately ten different spatial frequencies per device. I_2 vs time was similary photographed at low spatial frequencies where it was measurable. From a given I_1 vs time output (e.g., Fig. 2), we note that the horizontal or time axis also corresponds to the accumulated E_W values since $E_W = I_W t$. Thus each I_1 vs t graph also provides us with I_1 vs E_W for a given spatial frequency f.

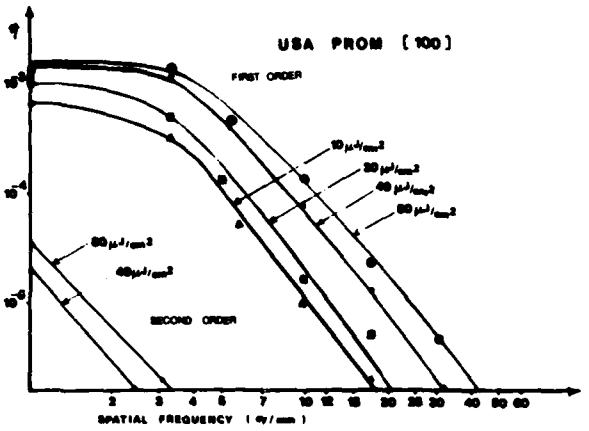


Fig. 3. Diffraction efficiency η vs spatial frequency f for the U.S. Prom with different write light exposure used.

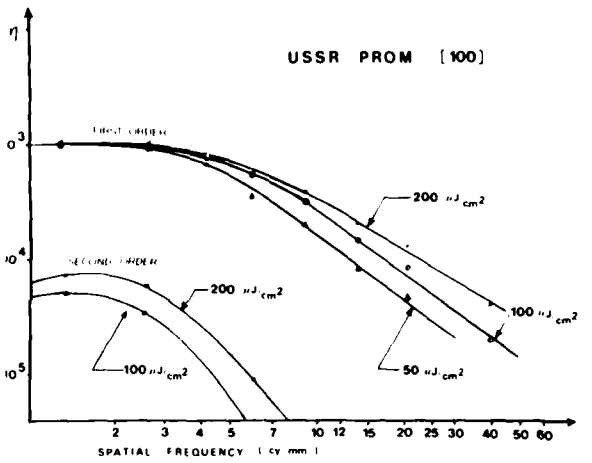


Fig. 4. Diffraction efficiency η vs spatial frequency f for the Soviet Prom with different write light exposures.

From a given I_1 vs time photograph, we thus determine I_1 for a given f and all E_W. Dividing I_1 by $I_R = I_{in}$, we obtain η for one f and all E_W values. By repeating the experiment for ten different input f's, we obtain an η vs f frequency response curve for all E_W. This represents a considerably more efficient and simpler dynamic MTF measurement technique than prior methods. We can also obtain the MTF of the device from a plot of modulation $m = 4\sqrt{\eta}$ vs f. This follows directly since $\eta = m^2/16$.

The FT lens used was of focal length $f_L = 500$ mm. Ony a $d = 15$-mm diam of it was used. Thus the resultant null–null width of the central lobe of each FT plane peak is $2x_1 = 2\lambda f_L/d = 40\ \mu$m. To facilitate detection, we used a fiber optic microscope with a 150-μm diam probe and a 2.5× objective. This produced an equivalent 60-μm diam detector. This probe size (~1.5 times the theoretical diffraction limited spot size) insured adequate detection of the intensities of the first- and second-order terms in the FT plane with little FT plane noise present in the data.

IV. Prom Experimental Data

In Figs. 3 and 4, we show $\eta = I_1/I_{in}$ vs spatial frequency for the U.S. and Soviet Prom light modulators for various write light exposures E_W from 10 to 200 μJ/cm^2 The second-order I_2/I_{in} data vs spatial frequency is also included in both graphs for low spatial frequencies. At high spatial frequencies its intensity is over 10^6 below that at dc, and thus it becomes comparable to the noise of the optical system used. We note that both devices exhibit comparable peak diffraction efficiency values $\eta_0 \simeq 0.1\%$ at low spatial frequencies. We also note that η breaks and starts to decrease at $f \simeq 3$ cycles/mm for both Proms. Higher E_W exposure values yielded slightly larger η values, but second-order distortion becomes quite severe in these cases. Thus data for them were not included.

We notice that the second-order intensity is over 20 dB below the first-order light for $f > 2$ cycles/mm (U.S. Prom) and for $f > 6$ cycles/mm (for the Soviet Prom). We also note that a larger range of E_W values was possible on the Soviet Prom before steady state occurred or before second-order intensity became excessive. Note that below 20-μJ/cm^2 write light exposures, I_2 was not measurable for the U.S. Prom and that similar results occurred for $E_W = 50\ \mu$J/cm^2 or less for the Soviet Prom. Thus all data generally correspond to exposures for which the response of the device has a nonlinearity of ~1% or less.

In all experiments, we will give three parameter measures for use in comparing the performance of the different devices. We first note the device's peak diffraction efficiency η_0 (at low spatial frequencies). We then measure the spatial frequency $f_{0.5}$ at which $\eta = 0.5\eta_0$. (This is the device's 3-dB spatial frequency resolution.) Finally, we will give the spatial frequency $f_{0.01}$ value for which $\eta = 0.01\eta_0$. (This latter spatial frequency response can only be utilized in high dynamic range systems where such low light levels can be detected and used.) For the U.S. Prom unit available for testing, these comparison parameters (for $E_W = 80\ \mu$J/cm^2) are $\eta_0 \simeq 0.1\%$, $f_{0.5} = 6.5$ cycles/mm, and $f_{0.01} = 23$ cycles/mm. For the Soviet Prom device tested, we found $\eta_0 = 0.1\%$, $f_{0.5} = 8$ cycles/mm, and $f_{0.01} > 100$ cycles/mm (this latter value was obtained by linear interpolation of the data) for an $E_W = 100\ \mu$J/cm^2.

We next consider the η vs f dependence of these devices. We note that η decreases faster with increasing f for the U.S. Prom than for the Soviet Prom. However, this is quite misleading, since the thickness of the parylene layer and the crystal are different for both Proms. The Prom's response varies as a function of the thickness of the different layers in the device. Thus, comparing any two Proms must be done with such issues in mind. For the Soviet unit, $d_d = 3\ \mu$m and $d_c = 400\ \mu$m were given parameters. Typical U.S. Proms employ $d_d = 5\ \mu$m and $d_c \simeq 600\ \mu$m. (The exact values were not available for the U.S. Prom tested, but the experimental data obtained support the above values.)

From the data of Fig. 3, we find that the U.S. Prom tested shows an $\eta \propto 1/f^3$ dependence for the $E_W = 40$-$\mu J/cm^2$ curve between the 3–30-cycle/mm frequency range. From Fig. 4, we see that the response of the Soviet Prom decreases less rapidly with $\eta \propto 1/f^2$. This can be directly explained with reference to the theory advanced in Sec. II and the different d_d thicknesses in the two devices. From Sec. II, we recall that an $\eta \propto 1/f^2$ dependence was predicted up to an $f \simeq 1/2\pi d_d$ with an $\eta \propto 1/f^4$ dependence at $f > 1/2\pi d_d$. For the Soviet Prom, $d_d = 3$ μm, and thus an $\eta \propto 1/f^2$ dependence is expected up to $f \simeq 50$ cycles/mm. This is approximately the response observed in Fig. 4. Moreover, with $d_d = 5$ μm for the U.S. Prom, the spatial frequency at which the $\eta \propto 1/f^2$ dependence ends is now $\sim f > 1/2\pi d_d = 30$ cycles/mm.

Thus the more rapid decrease in η with f for the U.S. Prom is the direct result of the larger dielectric layer thickness for this device compared with that of the Soviet Prom. As clearly indicated in Figs. 3 and 4 and explained above, direct comparison of Prom data is difficult unless each device is fabricated with layers of the same thickness. When the different crystal thicknesses ($d_c = 400$ μm for the Soviet Prom and $d_c = 600$ μm for the U.S. Prom) are included, the theoretical models in Refs. 9 and 10 can be used to further refine device comparisons. Since our major intent is only to show (by experiment) the comparable performance of two different Proms and that both support theoretical models advanced in Sec. II, we will not consider more detailed comparisons and experimental verifications of the more advanced Prom device models in Refs. 9 and 10. Rather we direct our attention to the performance data for the Priz SLM in the following section.

V. Priz Experimental Data

In Fig. 5, we show the η vs f experimental data for the Priz (device 3). Similar graphs were obtained for devices 2–4, and thus only the data in Fig. 5 are included here. From Fig. 5, we note only four issues. First, the measured diffraction efficiency of the Priz is over 20 times more than that of the Prom. Thus, it has more useful output light at all spatial frequencies. This verifies the observation noted in Sec. II that the transverse electrooptic effect in this BSO device is larger than the longitudinal effect. Second, these data also show that the spatial frequency response of the Priz is superior to that of the conventional Prom with $\eta \propto 1/f^2$ over the full 3–60-cycle/mm frequency range. From theory (Sec. II), it can be shown that the Priz should exhibit an $\eta \propto 1/f^2$ response over its full frequency range. Third, measured response beyond 60 cycles/mm was observed for this device (due to the larger I_1 light levels available). Fourth, the second-order response is larger for the Priz than the Prom. This appears due to the nonlinearity with which a transverse field is induced from a spatial longitudinal one as in the Priz. However, we note that for $f > 10$ cycles/mm, negligible second-order distortion exists. From this initial Priz device, we find our three comparison parameters to be $\eta_0 = 2 \times 10^{-2}$ (or 20 times larger than for the Prom), $f_{0.5}$ = 8 cycles/mm, and $f_{0.01} > 100$ cycles/mm. We note that these Priz comparison parameters were obtained with a lower $E_W = 50$-$\mu J/cm^2$ maximum exposure than for the Prom device. At higher exposures, the nonlinearity due to the second harmonic term was quite large.

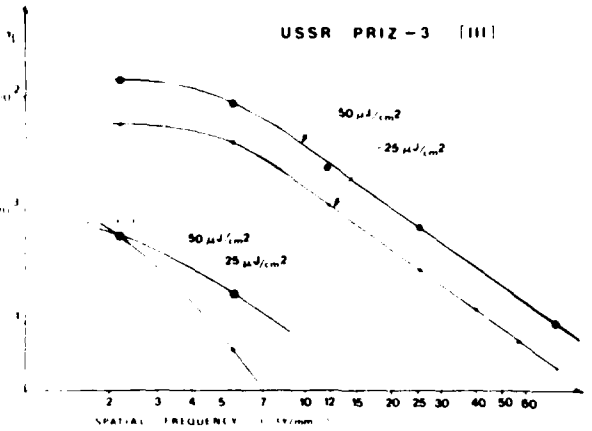


Fig. 5. Diffraction efficiency η vs spatial frequency f for the Soviet Priz for two different write light exposures.

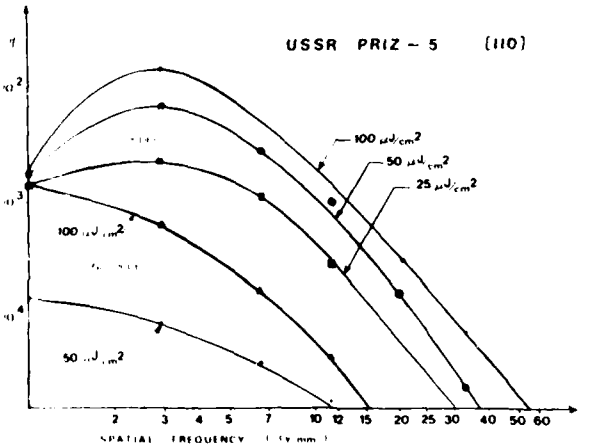


Fig. 6. Diffraction efficiency η vs spatial frequency f for the Soviet Priz for three different write light exposures showing the suppressed dc response possible on the Priz unit.

In Fig. 6, we show the measured response of another Priz (device 5). For this device, the η vs f curve for $E_W = 100$ $\mu J/cm^2$ is shown, and data points at spatial frequencies down to 1 cycle/mm are included. From this figure, we see that the Priz exhibits a peak in η at $f > 0$ ($f \simeq 3$ cycles/mm) and moreover that it suppresses data at dc and low spatial frequencies. This is expected since (Sec. II) at dc the field is uniform across the device and little transverse field exists. Hence little transverse electrooptic effect occurs, and thus dc (and low spatial

Table I. Summarized Comparison of the Prom and Priz Light Modulators

Device	η_0 (%)	$f_{0.5}$ (cycle/mm)	$f_{0.01}$ (cycle/mm)
U.S. Prom	0.1	6.5	23 cycles/mm
U.S.S.R. Prom	0.1	8	>100 cycles/mm
U.S.S.R. Priz	2.5	8	>100 cycles/mm

Device	η at different spatial frequencies		
	10 cycles/mm	20 cycles/mm	40 cycles/mm
U.S. Prom	1×10^{-4}	2×10^{-5}	3×10^{-6}
U.S.S.R. Prom	2×10^{-4}	1×10^{-4}	3×10^{-5}
U.S.S.R. Priz	7.5×10^{-3}	1×10^{-3}	2.5×10^{-4}

frequency) data are expected to be suppressed in this device. In Fig. 6, we also note that, although the second-order distortion is high at low spatial frequencies, it rapidly decreases and becomes negligible for frequencies above 15 cycles/mm. This device also exhibits the same $\eta \propto 1/f^2$ response over its full f = 1–50-cycle/mm measured frequency range. This Priz device exhibits a large η comparable with that of the other units. Its comparison parameter measurements are comparable with those of the other Priz devices (Fig. 5).

VI. Summary and Conclusions

In Sec. II, we highlighted the model and theoretical spatial frequency response to be expected from the Prom and Priz light modulators. The experimental system we used (Sec. III) was shown to be quite stable (in terms of variations in beam overlap area with varying input spatial frequency). The new data acquisition technique we employed was found to be quite useful. The experimental procedure for testing these light modulators followed the ideas advanced in our earlier work[14] and proved to be most appropriate in this example also. Our experimental data on the Prom SLM (Sec. IV) verified the theoretical models advanced in Sec. II and demonstrated how difficult it is to compare two Prom devices with different layer thicknesses. Nonetheless, we clearly showed that the theory advanced in Sec. II was valid, and we also found that the Soviet Prom unit evaluated performed comparable with the U.S. one available. The Priz data (Sec. V) showed that the diffraction efficiency of this device was over ten times that of the Prom (as predicted in Sec. II) and that its η vs f response obeyed an $\eta \propto 1/f^2$ dependence over the full frequency range.

In Table I, we summarized the highlights of our experiments on the Soviet Prom and Priz light modulators as well as U.S. Prom unit. In this table, we note our three comparison parameters for each device. We note that if one considers the MTF resolution of these devices, the $\sqrt{\eta}$ vs f curve is appropriate, and in this case the differences in the performance of the different units will be even more exaggerated (because of their quite different η vs f dependencies). In Table I, we also include a most meaningful set of data (the usable amount of light η available from each device at a given spatial frequency). These data clearly show that the Priz has from 10 to 100 times more usable light than does the Prom.

The Priz has many other features that are quite unique. These include: automatic suppression of dc and low spatial frequencies (see Fig. 6)[15]; directional spatial filtering; and dynamic image selection. Some amplification of the last two items is necessary. From a detailed analysis of the dependence of η on the polarization of the input light (with respect to the crystal's orientation), one can show[7] that the [110] cut Priz has preferential spatial frequency data directions that it will emphasize and that this effect can be controlled by properly adjusting the direction of polarization of the input light with respect to the crystallographic axes. Furthermore, the Priz can be operated with a fixed voltage betwen electrodes. (This greatly simplifies the electronic support system needed for the unit.) In this operating mode,[16] the Priz responds only to changes in the input data and thus exhibits dynamic image selection. In a future paper, we will discuss these unique image processing features of the Priz, other data on the dynamic write and erase performance of the device, as well as its sensitivity, optical quality, and the angular orientation accuracy required to use the unit.

We note in closing that the Priz can exhibit superior performance to that indicated in Figs. 5 and 6 when the write light wavelength λ_W is properly chosen to match the crystal's thickness. In general, when E_W and λ_W are properly selected to match the device's fabrication, optimum performance is possible.

We thank Michael Petrov for allowing us to test and evaluate the Prom and Priz light modulators fabricated in his laboratory at Ioffe Institute (Leningrad) and for allowing A. Khomenko to assist us in this work. We also thank the Air Force Office of Scientific Research (grant 79-0091) for supporting our analysis of this device and publication of this paper.

References

1. D. Casasent, Proc IEEE **65**, 143 (1979).
2. S. Hu and D. Oliver, Appl. Phys. Lett **18**, 325 (1971).
3. R. Sprague J. Appl. Phys. **46**, 1673 (1975).
4. B. Horwitz and F. Corbett, Opt. Eng. **17**, 353 (1978).
5. M. Petrov *et al.*, Ferroelectrics **22**, 651 (1978).
6. M. Petrov *et al.*, Mikroelektronika **8**, 20 (1979).
7. M. Petrov, at Conference on 4-D Optics, Mexico (1980).
8. M. Petrov *et al.*, Sov. Phys. Tech. Phys **25**, 752 (1980).
9. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **202**, 110 (1979).
10. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **218**, 100 (1980).
11. W. Roach, IEEE Trans. Electron. Devices **ED-21**, 453 (1974).
12. A. Khomenko *et al.*, Sov. Tech. Phys. Lett. **5**, 133 (1979).
13. R. P. Feynman *et al.*, *The Feynman Lectures on Physics*, Vol. 5 (Addison-Wesley, Reading Mass., 1967).
14. D. Casasent, Appl. Opt. **18**, 2445 (1979).
15. D. Casasent, F. Caimi, and A. Khomenko, Appl. Opt. **20**, 3090 (1981).
16. M. Petrov *et al.*, Sov. Tech. Phys. Lett. **6**, 165 (1980).

5. APPLICATIONS OF THE PRIZ LIGHT MODULATOR

Applications of the Priz light modulator

David Casasent, Frank Caimi, M. P. Petrov, and A. V. Khomenko

The Priz light modulator suppresses input data at zero spatial frequency, can provide directional spatial filtering, and can perform dynamic image selection or change detection. In this paper, we summarize the Priz's performance and provide experimental confirmation of the above three image processing applications of this device.

I. Introduction

Optical signal processing and image pattern recognition applications require real-time and reusable devices on which the input data to be processed can be recorded for subsequent optical processing. These vital components in an optical processor are known as spatial light modulators (SLM). Although many candidate SLM devices exist,[1] we will restrict our present attention to only one such device, the Priz light modulator.[2-8] Priz is a Soviet acronym that translates as image transformer. This modulator was proposed by a group of researchers of A.F. Ioffe Physico-Technical Institute (FTI) of the Academy of Sciences of the U.S.S.R. It employs the same active element, i.e., the bismuth silicon oxide (BSO) type crystal, as the well-known Prom modulator.[9]

At the Electrical Engineering Department of Carnegie-Mellon University (CMU), five Prom and Priz units fabricated at the FTI laboratory headed by Petrov were tested and evaluated. One worker of FTI participated in the research program (Khomenko).

In Sec. II, we review the structure of the Priz light modulator and the motivation for its fabrication and highlight the spatial frequency response data obtained on the devices we evaluated at CMU. We then include (Sec. III) a summary of the dynamic and optical performance of the Priz. These data were obtained from experiments performed in both the Soviet Union and at CMU. We include several image and signal pattern recognition correlation examples of the use of the device in Sec. IV. Because of the transverse linear electrooptic effect used in the Priz to modulate read light, it exhibits three unique features that are of use in various image processing and pattern recognition applications. These include: suppression of dc and low spatial frequency data as well as directional filtering of input spatial frequencies oriented in selected angular directions (Sec. V) and a quite unique feature referred to as dynamic image selection in which the device responds only to changes in the input image data (Sec. IV). Our summary and concluding remarks follow in Sec. VII.

II. Operation of the Priz

The Prom light modulator[9] consists of a BSO crystal $\sim 20 \times 20 \times 0.4$ mm with Parylene insulating layers and transparent electrodes on the large faces. In operation the spatially modulated data to be processed are imaged or scanned onto the device in λ_W write light (350–450 nm). Photocarriers are generated in the photoconductive BSO, and a spatially modulated charge layer is produced within the BSO. When the device is illuminated with a uniform read light beam at λ_R (usually 633 nm), the λ_R light emerging from the device is polarization modulated spatially with an amplitude of modulation that varies spatially in accordance with the original λ_W input light or data pattern. This λ_R modulation occurs by the linear–longitudinal electrooptic or Pockels effect. The polarization modulation can be converted to amplitude modulation when a crossed analyzer is placed behind the modulator.

In the Priz light modulator,[2-8] a [110] or [111] cut BSO crystal is used rather than the [100] cut crystal used in the Prom. Other proprietary fabrication techniques are employed, but the issue of major importance is that with these different crystal cuts, the device now modulates λ_R light by the transverse rather than the longitudinal electrooptic effect. The spatially varying λ_W light distribution is still incident on the crystal's large faces collinear with the applied electric

M. P. Petrov and A. V. Khomenko are with A. F. Ioffe Physico-Technical Institute of the U.S.S.R. Academy of Sciences, 194021 Leningrad, U.S.S.R.; the other authors are with Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, Pennsylvania 15213.

Received 22 April 1982.
0003-6935/82/213846-09$01.00/0.

field direction, and the spatially varying charge layer parallel to the crystal's large faces is still induced. However, the transverse component of this field is what is used to provide the spatial modulation of the λ_R light.

The Prom exhibits low diffraction efficiency η and a sharp $\eta \simeq 1/f^4$ decrease in usable output light intensity at high spatial frequencies f.[10] If we assume that the sine wave electric charge grating with amplitude σ_0 induced during image writing in the Prom is infinitely thin and that it exists at the crystal–dielectric interface, the phase modulation vs spatial frequency is described by[11]

$$\Delta\phi = \frac{2\pi\sigma_0}{U_{\lambda/2} f(\epsilon_d \coth 2\pi f d_d + \epsilon_c \coth 2\pi f d_c)}, \quad (1)$$

where ϵ_d and ϵ_c are the relative dielectric constants of the dielectric layer and the BSO crystal, respectively, d_d and d_c are the thicknesses of these layers, and $U_{\lambda/2}$ is the halfwave voltage of the crystal. Equation (1) indicates that when f increases at $f \geq \frac{1}{2}d_d$, $\coth 2\pi f d_d \simeq 1$ and $\coth 2\pi f d_c \simeq 1$, and thus $\Delta\phi$ decreases $\propto 1/f$. Since $\eta \propto (\Delta\phi)^2$, Eq. (1) predicts $\eta \propto 1/f^2$. However, in experiments a sharper ($\eta \propto 1/f^4$) dependence was observed. This was attributed to the fact that, in the process of image writing, a volume electric charge distribution is formed within the crystal volume rather than an infinitesimally thin one.[10] Using this new model, it has been shown that in the case when the charge is distributed throughout a layer of thickness d_a near the crystal–dielectric interface

$$\Delta\phi = \frac{\sigma_0[\cosh 2\pi f d_c - \cosh 2\pi f(d_c - d_a)]}{U_{\lambda/2} f^2 d_a(\epsilon_d \coth 2\pi f d_d + \epsilon_c \coth 2\pi f d_c)\cdot \sinh 2\pi f d_c}. \quad (2)$$

Equation (2) predicts an $\eta \propto 1/f^4$ dependence at high f, which agrees with experimental results. Thus the sharp dependence of η on f for the Prom is attributable to the volume character of charge distribution. The most complete and detailed theoretical description of the Prom device can be found in Refs. 12 and 13. These papers discuss a model that includes the location of the charge layer within the BSO, the thickness of the different device layers, the wavelength of the light used. To overcome the disadvantages of the volume charge predicted by Eq. (2), it was suggested in Refs. 14 and 15 to use the transverse electrooptic effect for read light modulation rather than the longitudinal effect used in the Prom. The modulator that uses the transverse electrooptic effect is called the Priz. The corresponding calculation of the phase modulation for the case of the transverse effect shows that[6]

$$\Delta\phi = \frac{4\pi\sigma_0}{U_{\lambda/2}\epsilon_c f}\left\{1 - \frac{\epsilon_d[\cosh 2\pi f d_c + \cosh 2\pi f(d_c - d_a) - 1] + k_1}{2\pi f^2 d_a(\epsilon_d \tanh 2\pi f d_c + \epsilon_c \tanh 2\pi f d_d)\cosh 2\pi f d_c}\right\}, \quad (3)$$

where $k_1 = \epsilon_c \tanh 2\pi f d_d \sinh 2\pi f d_a$.

From this formula, two characteristic features of the Priz device can be seen. First, $\Delta\phi(0) = 0$; i.e., the modulator suppresses the dc component. Second, at high f, $\Delta\phi \propto 1/f$ and $\eta \propto 1/f^2$, i.e., the Priz η vs f characteristic is superior to that of the Prom.

III. Performance of the Priz

In the MTF tests performed on the Priz at CMU, we operated the device at 2 frames/sec. In other experiments, we operated the device at 20 frames/sec. A faster frame rate should be possible, but no effort has yet been made to determine the device's maximum cycle time. However, results of several experiments are useful in providing some indication of the final device performance possible on the Priz.

In the experiments performed in FTI, the device was operated with write times as short at 7 nsec using a pulsed laser source with an intracavity electrooptic modulator. In this case, photocarriers were generated in a negligible time, but the output light pattern was not visible until 1 μsec later, and it peaked after 10 μsec. Carrier mobility and transit times thus appeared to limit the minimum write–read cycle times of the Priz to 10 μsec. In the more conventional operating mode, 1-msec exposure times are used, and the output pattern is then immediately visible.

Erase time is a second limitation on the device's speed. In all tests performed at CMU, a fixed 1-msec erase flash (from the standard erase unit provided with the U.S.A. Prom) was used. However, neither the Prom nor the Priz can be recycled immediately after erasure, and a delay time is necessary to allow redistribution and relaxation of excited carriers within the crystal. One millisecond of relaxation time and hence a total 1-msec erase time appear adequate. High-energy erase pulses cannot decrease this time, since they generate and dislodge other carriers within the bulk of the device, and longer relaxation times between the end of erasure and the start of a new write cycle then become necessary. Thus operation of the Priz at a 10^3-frame/sec (write–read–erase cycle) rate appears possible, but additional theory and experiments are necessary to confirm this. Moreover, an application for which the entire spatial input data changes every millisecond is necessary to merit such an effort together with attention to how one can introduce such a new 2-D λ_W spatial distribution to the device at these 10^3-frame/sec rates.

The storage times for the Priz are adequate for most applications [1-min storage in the dark and 10–20-sec storage under a high λ_R = 633-nm read light intensity (I_R = 2 mW/cm)]. These can be somewhat controlled by varying the thickness of the insulating layers. The lifetime of the Priz, like that of the Prom, appears to be excellent. Selected Priz units have been operated for over one million cycles at a 20-frame/sec rate with no noticeable change in performance. The only concern

with the device's lifetime appears to be its operation with high I_R light levels. In this case, a device with high transmittance and electrodes with good conductivity is necessary. For such cases, InO_2 electrodes are used. The performance of such Priz units appears to be good (as several of them have operated successfully for several years).

The resolution of the Priz is mostly understood and has been experimentally verified by several techniques. Its diffraction efficiency η_0 at low spatial frequencies is ~1%, its 10-dB spatial frequency response $f_{0.1}$ is ~30 cycles/mm, and the spatial frequency $f_{0.01}$ at which $\eta = 0.01\eta_0$ is 100 cycles/mm. At CMU, we operated the device with input data having spatial frequencies as high as 80 cycles/mm. With higher quality optical systems and a different area detection technique, FTI researchers have measured diffraction efficiency on the device beyond 500 cycles/mm. Since the usable output light intensity at these high spatial frequencies is quite low, only in selected applications can such resolution actually be used. The Priz performs well with input write light exposures E_W of 50 μJ/cm^2 or less with corresponding interharmonic distortions and hence response nonlinearities below 1% at these write light energies. The sensitivity of the device defined as the write light exposure E_W necessary to achieve $\eta = 1\%$ at 5 cycles/mm is 50 μJ/cm^2. In our CMU data on these devices,[7,8] an available laser source not optimized to the thickness of the Priz used was employed. With this experimental setup, we obtained only $f_{0.5}$ = 20 cycles/mm and $f_{0.1}$ = 30-mm resolution. If the optimal read light wavelength were used, we expect superior results with $f_{0.5}$ in excess of 30 cycles/mm as obtained at FTI.

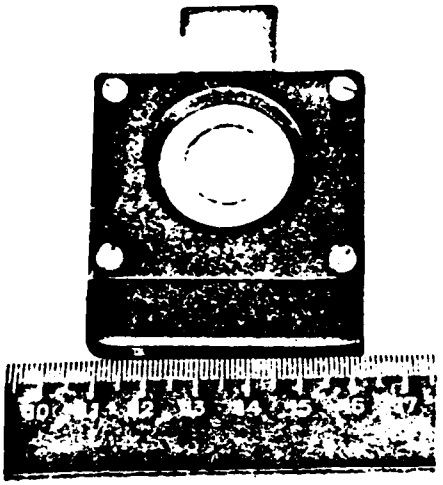


Fig. 1. Priz spatial light modulator.

Both the Priz and Prom modulators have high optical quality and allow use of large crystal sizes. The Priz units evaluated at CMU had a 15-mm diam active area with $\lambda/4$ optical quality. Larger units have been fabricated (up to 30-mm diam) with ~1λ optical flatness. The standard Priz units have also been fabricated, tested, and used with $\lambda/10$ optical quality. The BSO crystal in the Priz is nominally 400 μm thick, and the insulating layers are ~3 μm thick. No substrate is used in the Priz, or the Prom, as clamping effects change the dielectric constants of the materials and induce stresses in the device. One of the units tested at CMU is shown in Fig. 1.

In Table I, we summarize the salient Priz performance parameters. These data were obtained from diverse FTI and CMU tests and experiments. As with any BSO device, a wide range of performance is possible depending upon the thicknesses used for the different layers in the device. The parameters in Table I are all simultaneously obtainable, but they should be interpreted with the above consideration in mind. In all cases, the device should be operated at the intended write light exposures E_W and write light wavelength λ_W if optimum device performance is to be obtained. MTF data are frequently used to describe the spatial resolution of SLMs. Since the Priz has a dc response of zero, the MTF function $\tau(f)$ is not usable for such a device. Rather the diffraction efficiency η of such a device is the

Table I Priz Parameter and Performance Specifications

Parameter	Specification	Parameter	Specification
BSO crystal	400 μm thick	Address time	7 nsec (min)
Insulating layer	3 μm thick	Write/read cycle	1 msec (min)
Electrodes	Pt or InO_2	Erase time	1 μsec (min)
Active area		Erase cycle	1 msec (typical)
Typical	15-mm diam	Frame rate (typical)	20 frames/sec
Maximum	30-mm diam		
Optical quality		Dark Storage	1 min (typical)
Typical	$\lambda/4$	Storage with readout	20–30 sec (typical)
Special request	$\lambda/10$		
Write light		Diffraction effic. (η_0)	1%
Wavelength	350–500 nm	Resolution	
Exposure (typical)	50 μJ/cm^2	at $\eta = 0.1\eta_0$	30 cycles/mm
Sensitivity (for η = 1% at 5 cycles/mm)	50 μJ/cm^2	at $\eta = 0.01\eta_0$	> 100 cycles/mm
Read light		max. measured	> 500 cycles/mm
λ (typical)	633 nm	Spatial freq. response	$\eta \propto 1/f^2$
Intensity (typical)	2 mW/cm^2		

appropriate parameter. This is why all spatial frequency resolution data in Table I are given in terms of η vs f. These data were obtained with the Priz device operated dynamically in real time. In coherent optical processing, the amplitude transmittance $m(f)$ of the SLM vs frequency f is the parameter of interest. If an analyzer is used and the input pattern is the sine wave grating $I = I_0(1 + \sin 2\pi fx)$, the amplitude transmittance for the Priz can be written in the linear approximation as

$$t(x) = m(f)\cos 2\pi fx. \tag{4}$$

Equation (4) implies that the dc component of the Priz response is zero [since from Eq. (3) $\Delta\phi = 0$ at $f = 0$], and the readout pattern has a $\pi/2$ phase shift with respect to the recorded one. In this case, $m(f)$ is directly obtained from the $\eta(f)$ data provided by

$$|m| = 2\sqrt{\eta}. \tag{5}$$

This expression follows from the Fourier transform analysis of Eq. (4).

In Fig. 2, we show the diffraction efficiency η of Priz and Prom units at comparable $E_W = 50\text{-}\mu\text{J/cm}^2$ exposures. From these data, we see the superior diffraction efficiency and resolution of the Priz. We also see that the response of the Priz decreases at higher spatial frequencies at a much lesser rate than that of the Prom. Where $m(f)$ vs f is plotted rather than $\eta(f)$, the differences would be even larger. We also note that the Priz exhibits a suppressed response at low spatial frequencies. This feature follows directly from the fact that, upon uniform illumination of the device with write light, a uniform longitudinal electric field is formed that has little or no transverse component. Since the Priz employs the transverse electrooptic effect, it will not modulate in response to such light. As a result η of the Priz device peaks at a spatial frequency of 4 cycles/mm (for the unit tested) rather than at dc.

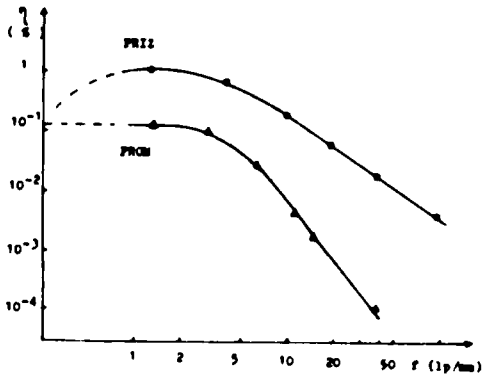


Fig. 2. Diffraction efficiency η vs spatial frequency f for the Prom and Priz spatial light modulators.

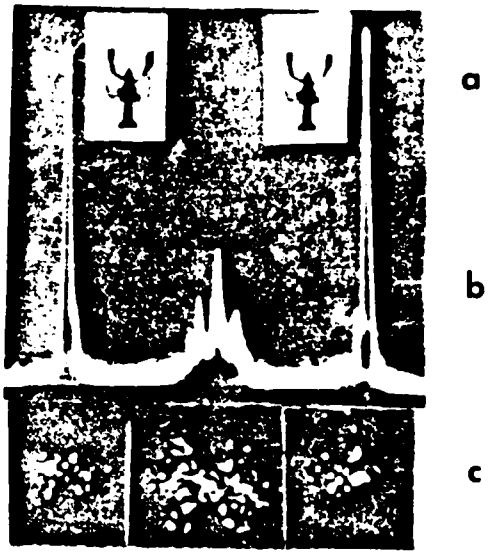


Fig. 3. Real-time image pattern recognition with the Priz used in the Fourier transform plane of a joint Fourier transform correlator: (a) input and reference images; (b) cross-sectional scan; and (c) an image of the output correlation plane pattern. The two peaks to the right and left in (b) and (c) represent the correlations of the two input objects.[3]

IV. Use of the Priz in Pattern Recognition and Signal Processing Correlations

An attractive optical correlator for image pattern recognition is the joint transform correlator.[16] In this system, the reference object being sought is placed beside the real-time input scene in the input plane of a 2-D optical FT system. The objective is to determine if the reference object is present in the input scene and to determine its location. Such pattern recognition applications are appropriate for locating objects on an assembly line and locating areas and landmarks in satellite imagery as well as in missile guidance and many other applications. In the joint transform correlator, the Fourier transform of the input and reference data is formed on an intensity sensitive material (such as film, the Priz, or Prom). The Fourier transform of this joint FT pattern is then formed, and it can be shown[16] that it contains the correlation of the input and reference images. In Fig. 3 we show an example[3] of such a correlation performed on the Priz. The FT of the two input objects [Fig. 3(a)], identical images of lobsters, was recorded on a Priz placed in the FT plane of a lens, which was behind the joint input pattern of Fig. 3(a). The FT of the data recorded on the Priz is shown in Fig. 3(c) and its cross-sectional scan in Fig. 3(b). This full correlation plane pattern contains a central term that is the sum of the autocorrelations of each input object. The large spikes on the left and right in Fig. 3(b) are the correlation of the two input objects. Their presence indicates that the two input objects are similar, and the relative position of the peaks denotes the location of the reference object within the field of view of the input image.

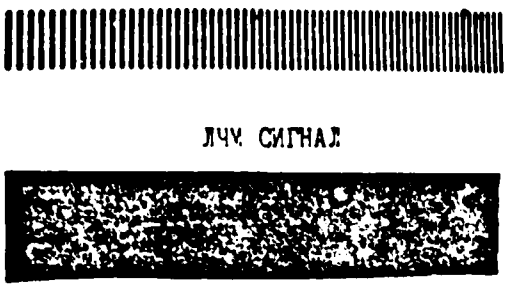


Fig. 4. Real-time correlation or compression of a linear frequency modulated signal using the Priz as the input transducer: (a) linear frequency modulated input signal; (b) compressed output correlation signal.[2]

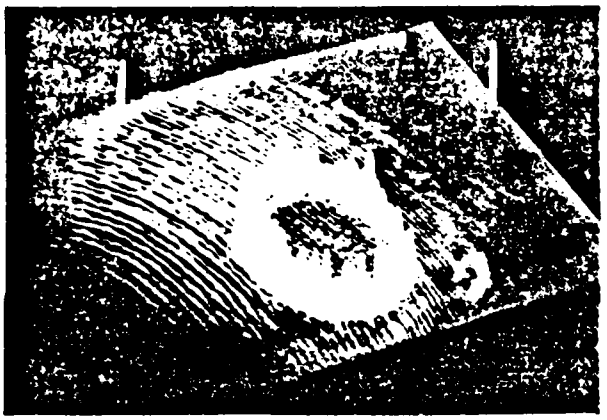


Fig. 5. Real-time image edge enhancement and directional spatial filtering using a [111] Priz: (a) edge-enhanced reconstructed image; (b) directionally filtered reconstructed image.

Signal correlations have also been performed on the Priz. An image of a linear frequency modulated (LFM) signal [Fig. 4(a)] was recorded on the Priz. The FT of this pattern formed with a lens whose focal length was matched to the frequency range of the LFM waveform yields the resultant compressed pulse or output correlation peak[2] shown in Fig. 4(b). The width of the experimentally obtained correlation peak was 1/100th of the width of the original LFM signal. This is in good agreement with the theoretical pulse compression factor of 120 for the LFM signal used. The difference between theory and experiment was due to taper in the input light beam and other effects.

V. Image Spatial Filtering Using the Priz

In Fig. 2 and Sec. III, we noted that because of the transverse electrooptic effect in the Priz, dc and low spatial frequency data are automatically suppressed. In Fig. 5(a), we show the reconstruction of a circular input object recorded on the Priz. As expected, only the edge contour of the object appears (due to the automatic dc spatial frequency suppression performed by the Priz). The amount of dc suppression depends on how closely the read light beam is incident to the normal to the crystal. In our experiments at CMU, we were able to obtain a dc suppression of 10^3 when the normal to Priz was aligned within 4° of the read beam. For a 10^4 suppression factor, 1° alignment is necessary. In the image processing experiments at CMU (Figs. 5 and 7), the read beam was incident at an angle of 1.7° to the normal to the crystal.

The Priz has anisotropic properties arising from those of the linear transverse electrooptic effect.[17] In particular, it exhibits a distinct difference in response to circularly and linearly polarized read light. Figure 6 shows how diffraction efficiency to the Priz depends on orientation of the crystal's axes when read out with linearly and circularly polarized light.[2] In the data of Figs. 5 and 7, the electric vector of the linearly polarized light was along the $[11\bar{2}]$ axis of the crystal. In Fig. 6, the diffraction efficiency η as a function of the angle between the wave vector of the sine wave grating and the [110] crystal axis is plotted in polar coordinates.

The outer circle in Fig. 6 describes the device's response to circularly polarized input light. As seen, it is quite uniform, and thus operation with circularly polarized input light produces no directional preference for input spatial frequency. The reconstructed image in Fig. 5(a) verifies this response and is essentially how the circular outer curve in Fig. 6 was obtained. However, the response of the Priz to linearly polarized read light is quite different. In the two inner figure eight shaped curves in Fig. 6, we show the response for linearly polarized read light. When the device is exposed to linearly polarized input light, it exhibits a preferred response η for input spatial frequencies oriented in one direction, while greatly suppressing input spatial frequencies oriented in the orthogonal direction. The direction in which spatial frequencies are suppressed can be controlled by the polarity of the voltage applied to the modulator if the polarization of the read light is

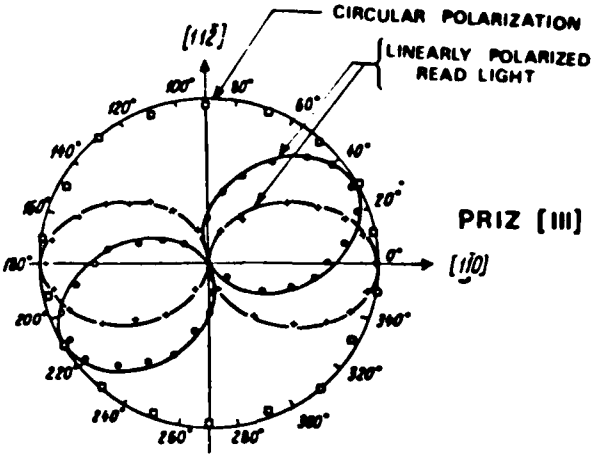


Fig. 6. Diffraction efficiency η for the Priz as a function of the angle between the wave vector of the recorded sine wave grating and the crystal's axes for input read light with circular and linear polarization. Curves 1 and 2 correspond to experimental points for linearly polarized light, and curve 3 corresponds to circularly polarized read light.

fixed. The two figure eight shaped plots in Fig. 6 were obtained with the same polarity of the read light but with different polarities of the applied voltage. Differences arising from changing the voltage polarity can be attributed to the optical activity of the BSO crystal.[17] Figure 5(b) shows the reconstructed image of a circular object recorded on the Priz and read with linearly polarized light. As seen, the spatial frequencies in one direction are suppressed as predicted by Fig. 6.

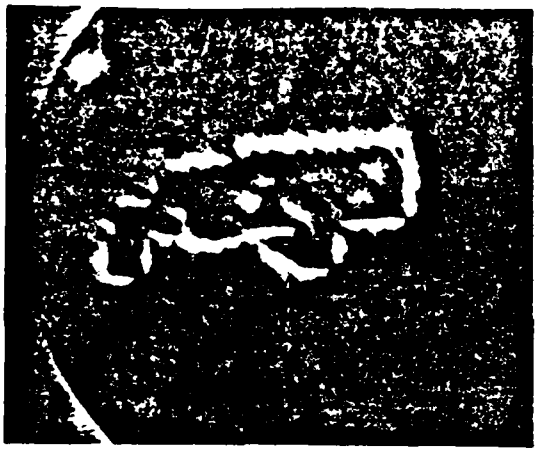
For a [110] cut Priz, similar plots of η vs the read wave vector's direction result. However, for circularly polarized read light, a saddle-shaped response rather than a circular one results. Similarly, a larger η (a factor of 2 larger than for the [111] cut Priz) results when the [110] cut device is operated with linearly polarized read light. Thus the [110] cut device is preferable for multichannel 1-D signal processing applications and others in which directional spatial filtering is desired. Conversely the [111] cut Priz is preferable for image processing where a uniform response is generally desired for all input spatial frequency directions.

To achieve a high degree of suppression of the dc component in the image (both with linearly and circularly polarized read light), the modulator should be placed between a high-quality polarizer and analyzer, which in the case of circularly polarized light can be achieved with a $\lambda/4$ wave plate and a linear polarizer. The dc suppression, directional spatial filtering, and edge enhancement features of the Priz are quite useful preprocessing operations for multisensor and IR pattern recognition. In Fig. 7, we show the original IR image [Fig. 7(a)], the image constructed from a [110] cut Priz with the read light polarized at 45° [Fig. 7(b)] and with vertically polarized read light [Fig. 7(c)]. The reconstructed image in Fig. 7(b) approximates an edge-enhanced version of the original image, whereas the reconstruction in Fig. 7(c) results in enhancement of vertical lines in the original image and suppression of horizontal spatial frequencies in the original input pattern.

VI. Dynamic Image Selection

In investigating the response of the Priz to spatially moving 2-D input patterns, it was found[4] that the device's response was a function of both the spatial frequency of the input data and the velocity with which the input data moved across the input field of view. A modified version of the Priz was used in these experiments. It had no insulating layers, so electrodes were evaporated directly on the crystal's surface.[18] The response of the device to an 0.5-mm wide input line was measured for different velocities (1–40 mm/sec) of the input object across the input plane, and it was found[4] that the response of the device peaked when the velocity was ~7 mm/sec. The response of the Priz is thus a function of both time and space (i.e., the spatial frequency of the input data and the rate at which it

(a)



(b)

(c)

Fig. 7. Real-time image edge enhancement and directional spatial filtering of IR imagery on a [110] Priz: (a) original input image; (b) edge-enhanced reconstructed image; (c) directionally filtered reconstructed image.

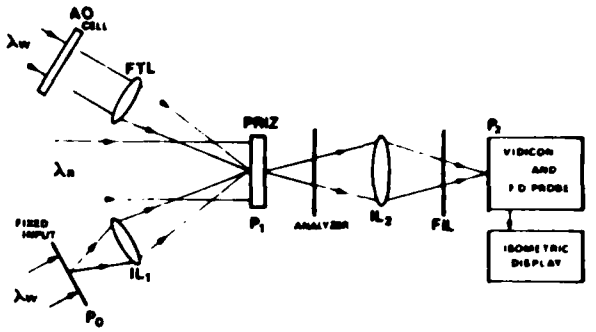
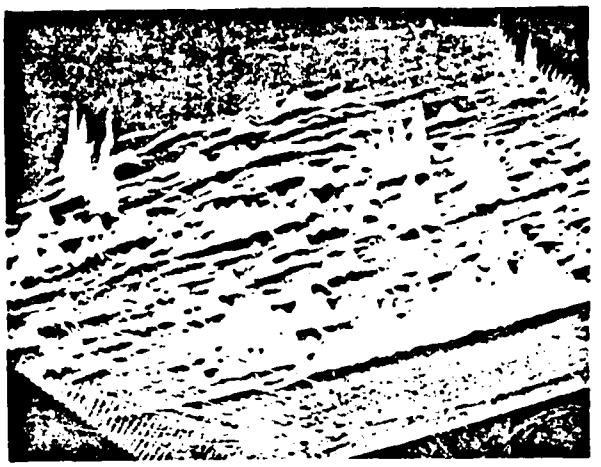


Fig. 8. Schematic diagram of the optical system used to demonstrate dynamic image selection or change detection on the Priz.

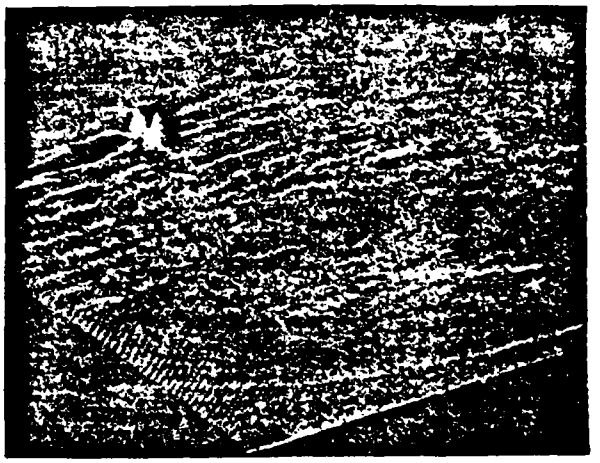
changes or moves with time). An initial attempt to describe the combined time and space response of the device has been reported.[5,18] Initial experiments[4,18] indicate that as the spatial frequency of the input data decreases, the input velocity for which the response η of the device peaks shifts to higher velocities.

To obtain this effect, the Priz is operated with a fixed voltage across its electrodes rather than with the applied voltage switched between positive and negative polarities. In the normal mode (polarization of the applied voltage switched), the device performs like the Prom. Only with a fixed voltage will it perform dynamic image selection. In this mode, the device responds only to changes in the input image. This operating mode is attractive for many applications such as change detection, and it also greatly simplifies the electronic support system necessary (since a fixed rather than a switching high-voltage supply can be used). Over a selected range of input temporal frequencies f_t (where this range varies as a function of the input spatial frequency f_x), the η vs f_t response is linear, and the device performs a time differentiation of the input data. This range of f_t is quite small, and moreover it varies with the intensity of the input write light. For these reasons, this Priz device features is best termed dynamic image selection (i.e., the device's output represents only the changing part of the input data) rather than temporal differentiation.[4] This effect can also be observed with a fixed input and with the write light beam pulsed on and off. In this case, whenever the write light changes (goes on or off), an output image of the input data appears and then decays with a time constant that is a function of the intensity of the write light. If the differential phase $\Delta\phi$ of the output light is measured, it is seen to be of opposite sign when the write light is switched from off to on compared to when it is switched from on to off.

When the Priz units were being tested at CMU, we found this Priz feature to be most attractive and thus assembled the system of Fig. 8 to demonstrate the use of the Priz in change detection. The system of Fig. 8 contains two input planes. Plane P_0 contained a fixed image, in our case a random pattern of uncorrelated noise and correlated noise of different correlation lengths and with different mean values. This fixed P_0



(a)



(b)

Fig. 9. Demonstration of dynamic image selection or change detection in real time on the Priz: (a) composite image of a moving target on a fixed noise background; (b) dynamic real-time selection of the moving target portion of the image in (a).

pattern was imaged onto the Priz at P_1 (by imaging lens IL_1) together with the FT of an acoustooptic) (AO) cell (using the FT lens *FTL*). The AO cell was operated in the scanning mode with a repeated LFM input signal. This caused a scanning spot to traverse the Priz (superimposed on the fixed noise pattern from P_0), thus simulating a moving object on a noise background. The velocity of the scanning spot was adjusted to be 2.8 mm/sec, and its size was ~1 pixel (40 μm).

Both the fixed and moving input patterns were imaged onto the Priz in λ_W = 476-nm light from an argon-ion laser. Readout was performed in λ_R = 633-nm light incident normal to the Priz as shown in the left side of Fig. 8. The pattern on the Priz was then imaged onto P_2 using imaging lens IL_2 through a crossed analyzer and a 633-nm filter (*FIL*). The output P_2 pattern was detected on a vidicon, and the dynamic moving output was visible on an isometric display where it could be photographed. A fiber-optic (FO) probe with a microscope and PMT was also placed at P_2 to allow quantitative measurements of the output plane to be made.

In Fig. 9(a), we show the full output image at P_2 (with the high-voltage Priz power supply operated in the normal pulsed mode). This output shows the fixed background noise pattern and the moving spot. (The moving object or spot is present in the back left of the figure.) When the high-voltage Priz power supply polarity was fixed, only the time varying portion of the input pattern appeared at the output. In this case, only the moving spot produced by the scanning AO cell was visible. In Fig. 9(b), we show the P_2 output for one location of the scanning spot (corresponding to a simulated moving target in the constant noise background image). As can be seen, the Priz suppresses the fixed background noise quite well.

Table II Quantitative Dynamic Image Selection Data on the Priz

R_{in}	1	2.5	10	25
R_{out}	5	50	48	49

To quantify the amount of background suppression obtained, the intensity of the moving spot was varied and the ratio

$$R = \frac{\text{intensity of the moving spot}}{\text{average intensity of the fixed background}} \quad (6)$$

was measured at the input and output of the system (i.e., with and without dynamic image selection or change detection). The results are summarized in Table II. From these data, we note that once the intensity of the scanning spot has been increased so that $R_{in} = 2.5$ or greater, a constant $R_{out} \simeq 50$ ratio results with the intensity of the dynamic part of the output image being 50 times the average background level in the P_2 output. This occurred because when R_{in} was increased above 2.5, saturation of the Priz occurred, and thus no further changes in the effective R_{in} resulted. This is expected since, with the Priz in the FT plane of the AO cell, all the light from the cell was concentrated onto a single pixel on the Priz.

VII. Summary and Conclusion

A complete and unifying summary (at this present time) of a new light modulator, the Priz, has been presented. Many new experimental demonstrations recently obtained at CMU were included. The theory of operation of the Priz was reviewed and experimentally verified, and its similarity and differences from the Prom were noted (Sec. II). The first unifying summary of the performance parameters of the Priz was also advanced (Sec. III). A summary of many of the possible applications of the Priz was then presented. This included conventional optical pattern recognition and optical signal processing correlators (Sec. IV) plus three new image processing operations: dc suppression and directional filtering (Sec. V) plus dynamic image selection (Sec. VI). The first image processing operation had been experimentally demonstrated previously. Directional spatial filtering demonstrations had not been previously described. Our dynamic image selection experiments together with Ref. 18 represented the first examples of the use of the device for the selection of the dynamic part of an image from a constant fixed noise background (i.e., change detection).

Many aspects of the Priz device are well understood. However, further theoretical analyses and modeling together with further device fabrication and experimental testing are necessary to understand fully and describe many of the observed features of the device. Issues meriting further analysis include the nonlinearity associated with the transverse electrooptic effect, the combined time and space dependence of the resolution of the device, and a theoretical formulation of the dynamic image selection feature of the device with attention to the selection of device parameters to optimize and control this effect.

The promising performance parameters tabulated for the Priz light modulator and the experimental verification included of several of the novel features of this device indicate that a wealth of new research efforts and applications are still possible in the field of real time and reusable spatial light modulators.

References

1. D. Casasent, Proc. IEEE **65**, 143 (1979).
2. M. Petrov, at Conference on 4-D Optics, Mexico (1980).
3. M. P. Petrov *et al.*, Sov. Phys. Tech. Phys. **25**, 752 (1980).
4. M. P. Petrov *et al.*, Sov. Tech. Phys. Lett. **6**, 165 (1980).
5. M. P. Petrov and A. Khomenko, Opt. Commun. **37**, 253 (1981).
6. M. P. Petrov *et al.*, Zh. Tekh. Fiz. **51**, 1422 (1981) [Sov. Phys. Tech. Phys. **26**, 816 (1981)].
7. D. Casasent, F. Caimi, and A. Khomenko, Appl. Opt. **20**, 3090 (1981).
8. D. Casasent, F. Cimi, and A. Khomenko, Appl. Opt. **20**, 4215 (1981).
9. B. H. Horwitz and F. C. Corbett, Opt. Eng. **17**, 353 (1978).
10. M. Petrov *et al.*, Ferroelectrics **22**, 651 (1978).
11. W. Roach, IEEE Trans. Electron. Devices **ED-21**, 453 (1974).
12. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **202**, 110 (1979).
13. Y. Owechko and A. Tanguay, Proc. Soc. Photo-Opt. Instrum. Eng. **218**, 67 (19809).
14. M. Petrov *et al.*, Mikroelektronika Akad. Nauk SSSR **8**, 20 (1979).
15. A. Khomenko *et al.*, Sov. Tech. Phys. Lett. **5**, 133 (1979).
16. J. Rao, J. Opt. Soc. Am. **57**, 798 (1967).
17. M. Petrov and A. Khomenko, Fiz. Tverd. Tela (Leningrad) **23**, 1350 (1981) [Sov. Phys. Solid State **23**, 789 (1981)].
18. M. Petrov, in *Current Trends in Optics, Proceedings, ICO-12* (Taylor and Francis, London, 1981), pp. 161–172.

Two of the authors (D.C. and F.C.) thank the Air Force Office of Scientific Research (grant 79-0091) for supporting our time during this study and for supporting publication of this paper. The other authors (M.P. and A.K.) thank the U.S.S.R. Academy of Sciences for research support of their ongoing Prom and Priz light modulator studies.

6. A CORRELATOR FOR OPTIMUM TWO-CLASS DISCRIMINATION

A CORRELATOR FOR OPTIMUM TWO-CLASS DISCRIMINATION

by

D. Casasent, B.V.K. Vijaya Kumar and H. Murakami*
Carnegie-Mellon University, Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

*Present Address: Toshiba Corporation, 70, Yanagi-Cho, Saiwai-Ku, Kawasaki, 210 Japan

ABSTRACT

The optimal correlation filter for the two-class discrimination problem is considered. A simple iterative procedure is suggested for the design of the filter. Experimental verification of the performance of this optimal correlator in discriminating the class of circles from the class of hexagons is included.

1. INTRODUCTION

One of the basic problems in pattern recognition is the discrimination between two classes of data objects. This problem is usually solved by first extracting the relevant features and then classifying them. When the class of objects of interest consists of randomly distorted versions of one object, statistical techniques are needed to obtain useful features. Several general approaches to statistical feature extraction are available in the literature [1,2]. Most of these methods optimize some distance criteria such as divergence [3] or Fisher discriminant [4] between the two classes. These distance measures are not easily available for comparison purposes in optical correlation experiments. A more useful quantity to optimize is the signal to noise ratio since it can be easily measured. In this paper, we will develop a new two-class discrimination correlator that optimizes the output correlation SNR.

The Karhunen-Loeve (K-L) expansion can be shown [5,6] to represent a random process in an optimal sense. Fukunaga and Koontz [3] suggest a modification of this technique to improve its discriminatory power. According to their method, the auto-correlation matrix of each class is transformed such that the best fit eigenvector of one class represents the other class very poorly. Foley and Sammon [4] determine the optimal discriminant vector that maximizes the Fisher ratio, the ratio of the between-class variance to the sum of the within-class variances [1]. Caulfield, et al. [7] have used the same Fisher criteria to design optically implementable discriminators. In this paper, we derive an optimal correlator to discriminate one class of signals $\{x(t)\}$ from another class of signals $\{y(t)\}$, where both $\{x(t)\}$ and $\{y(t)\}$ are considered as stochastic processes. The data sets $\{x(t)\}$ and $\{y(t)\}$ are in general images and, for the problem we consider, each set consists of different geometrically distorted versions of a different object. Our new approach is based on optimizing the cross-correlation output SNR. This parameter is of more practical significance in optical processors than the other measures. This optimal correlator will be derived by extending the conventional matched filter model. The matched filter is a discriminator for the signal $\{x(t)\}$ in the stationary white noise process $\{n(t)\}$, whereas the optimal correlator considered in this paper is a discriminator of a class of signals $\{x(t)\}$ from the nonstationary noise process $\{y(t)\}$.

In Section 2, the nonstationary noise model is derived from the two-class problem. By assuming that both stochastic processes $\{x(t)\}$ and $\{y(t)\}$ can be approximated by a finite K-L expansion with basis function sets $\{\Phi_i\}$ and $\{\xi_\ell\}$ respectively, we will prove that the optimal discriminator can be represented by a linear combination of the $\{\Phi_i\}$ and $\{\xi_\ell\}$. In Section 3, the optimal discriminator for the two-class problem is determined. We present experimental results in Section 4 to support our conclusions. As our data base for simulations, we use circles with randomly-varying diameters as one image class and hexagons with randomly-varying diameters as another class.

2. NONSTATIONARY NOISE MODEL

In the two-class discrimination problem, one class of signals $\{x(t)\}$ is to be discriminated from another class of signals $\{y(t)\}$. In order to find the optimal filter which can discriminate $\{x(t)\}$

from {y(t)} in the presence of noise n(t), we use the model in Figure 1. The noise process {y(t)} is nonstationary since it represents an image class with its own statistics that may be spatially varying. We thus call this model a nonstationary noise model.

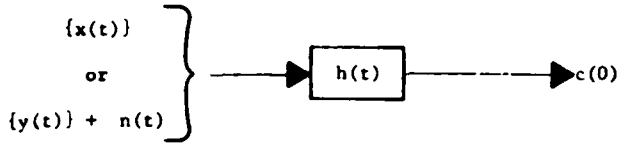


FIGURE 1 Schematic of the nonstationary noise model.

The optimal filter is derived by maximizing the output SNR defined as

$$SNR = \left. \frac{E[\int x(t)h(\tau-t)dt]^2}{E[\int\{y(t) + n(t)\}h(\tau-t)dt]^2} \right|_{\tau=0} , \tag{1}$$

where we assume 1-D functions for notational simplicity and where (with no loss of generality) we assume that the correlation peak occurs at $\tau = 0$. Letting $\tau = 0$ and $h(-t) = f(t)$ in (1), we obtain

$$SNR = \frac{E[\int x(t)f(t)dt]^2}{E[\int\{y(t) + n(t)\}f(t)dt]^2} . \tag{2}$$

It can be reasonably assumed that each class of signals {x(t)} and {y(t)} can be represented with sufficient accuracy by a finite dimensional subspace [8]. Thus when {x(t)} or {y(t)} is represented using the K-L expansion, they can be approximated by the subspace spanned by a small number of K-L basis functions. Let us assume that {x(t)} can be represented with sufficient accuracy by the basis function set $\{\Phi_i\}$ and that {y(t)} can be similarly represented by the basis set $\{\xi_\ell\}$. Then, we can compose a basis function set $\{\Psi_n\}$ (with N elements) which contains both $\{\Phi_i\}$ and $\{\xi_\ell\}$ and we can prove the following theorem. The optimal filter f(t) can be represented by a linear combination of the basis functions $\{\Psi_n\}$.

The proof proceeds as follows. Let the optimal filter f(t) be decomposed into $f_1(t)$ and $f_2(t)$, where $f_1(t)$ belongs to the subspace spanned by $\{\Psi_n\}$ and where $f_2(t)$ is orthogonal to that subspace, and hence to the spaces spanned by $\{\Phi_i\}$ and $\{\xi_\ell\}$. Then, the SNR in (2) becomes

$$SNR = \frac{E[\int x(t)f_1(t)dt]^2}{E[\int\{y(t) + n(t)\}f_1(t)dt]^2 + E[\int\{y(t) + n(t)\}f_2(t)dt]^2} . \tag{3}$$

From (3), the following inequality can easily be proved,

$$SNR \leq \frac{E[\int x(t)f_1(t)dt]^2}{E[\int\{y(t) + n(t)\}f_1(t)dt]^2} \quad . \tag{4}$$

The right hand side of (4) is the SNR obtained using the filter $f_1(t)$. Thus, (4) shows that the filter $f_1(t)$ gives a larger correlation output SNR than the filter $f(t) = f_1(t) + f_2(t)$. Thus the optimal filter f(t) can be completely specified by $f_1(t)$ and hence by $\{\Psi_n\}$. We will use this theorem in determining the optimal filter in the next section.

3. OPTIMAL FILTER

To derive the optimal filter, we first describe the class of signals $\{x(t)\}$ in terms of the orthonormal basis $\{\Psi_n\}$ as

$$x(t) = a_1\psi_1(t) + a_2\psi_2(t) + \ldots + a_N\psi_N(t) \quad , \tag{5}$$

where the coefficients $\{a_n\}$ are random variables and where N is the number of basis images $\psi(t)$. To our knowledge, such a representation of randomly distorted imagery by a set of basis functions with random coefficients and the use of SNR in (1) has not been employed before. Similarly, the noise process $\{y(t)\}$ can be represented by

$$y(t) = w_1\psi_1(t) + w_2\psi_2(t) + \ldots + w_N\psi_N(t) \quad , \tag{6}$$

where the coefficients $\{w_n\}$ are random variables. From our theorem in Section 2, the optimal filter h(t) is a linear combination of the $\{\Psi_n\}$, i.e.

$$h(t) = p_1\psi_1(t) + p_2\psi_2(t) + \ldots + p_N\psi_N(t) \quad . \tag{7}$$

The optimal filter f(t) can be determined by solving for the optimal coefficients $\{p_n\}$ in (7) which maximize the SNR in (2). Since $\{\Psi_n\}$ is a set of orthonormal functions, the SNR in (2) becomes, after substitution of (5), (6) and (7),

$$SNR = \frac{E[\sum_{i,j=1}^{N} a_i a_j p_i p_j]}{E[\sum_{i,j=1}^{N} w_i w_j p_i p_j] + I\sum_{i,j=1}^{N} p_i p_j} \quad . \tag{8}$$

We will rewrite (8) more succinctly as

$$SNR = \frac{\underline{p}^T \underline{A}\,\underline{p}}{\underline{p}^T \underline{W}\,\underline{p} + I\underline{p}^T\underline{p}} \quad , \tag{9}$$

where $\underline{p}$ is the N × 1 column vector with p_n as its n-th element, I is the white noise power, $\underline{A}$ is the N × N correlation matrix with its i,j-th element given by $E[a_i a_j]$ and $\underline{W}$ is a similar N × N correlation matrix with its i,j-th element given by $E[w_i w_j]$. Finding the optimal filter h(t) which maximizes the SNR in (9) is equivalent to finding the optimal vector $\underline{p}$ which maximizes

$$SNR = \frac{\underline{p}^T \underline{A} \underline{p}}{I + \underline{p}^T \underline{W} \underline{p}} \tag{10}$$

under the condition $\underline{p}^T \underline{p} = 1$. Note that a simple scaling of the filter vector $\underline{p}$ does not change the SNR.

The optimization problem in (10) is solved in the Appendix. The solution can be written as

$$(\underline{A} - \alpha \underline{W})\underline{p} = I\alpha \underline{p} \ . \tag{11}$$

where α is a scalar. Because of the inter-dependence of $\underline{p}$ on α, an iterative method is used to solve (11) for $\underline{p}$. The four steps in this new iterative algorithm follow:

STEP 1: Set $\alpha_0 = 1$.

STEP 2: Determine the largest eigenvalue λ_m and the corresponding eigenvector $\underline{p}_m$ for the matrix equation

$$(\underline{A} - \alpha_m \underline{W})\underline{p}_m = \lambda_m \underline{p}_m \ . \tag{12}$$

STEP 3: Modify α_m according to the λ_m value obtained as

If $\lambda_m > \alpha_m I$, set $\alpha_{m+1} = \alpha_m + \Delta\alpha$ and

if $\lambda_m < \alpha_m I$, set $\alpha_{m+1} = \alpha_m - \Delta\alpha$,

where $\Delta\alpha$ is a positive increment in α.

STEP 4: Repeat Step 2 and Step 3 until

$\lambda_m > \alpha_m I$ and $\lambda_{m+1} < \alpha_{m+1} I$

or

$\lambda_m < \alpha_m I$ and $\lambda_{m+1} > \alpha_{m+1} I$.

The eigenvector $\underline{p}_m$ obtained is the solution of (11) and by (7) defines the optimal filter f(t). For improved accuracy, we select $\Delta\alpha$ small; whereas for faster computation, a larger $\Delta\alpha$ is chosen. In the initial iteration steps, $\Delta\alpha = 0.1$ is chosen. As λ approaches its true value, $\Delta\alpha$ is dynamically reduced to 0.01 to yield accurate final estimates of λ_m. Use of this iterative procedure can be justified because the maximum eigenvalue λ_m of (12) decreases for increasing α and vice versa. Let us evaluate the maximum SNR obtained by this method. Substituting the solution of (11) into (10), we find

$$(SNR)_{opt} = \frac{\underline{p}^T\underline{A}\underline{p}}{I + \underline{p}^T\underline{W}\underline{p}} = \frac{\underline{p}^T[\alpha I\underline{p} + \alpha\underline{W}\underline{p}]}{I + \underline{p}^T\underline{W}\underline{p}} = \frac{\alpha[I + \underline{p}^T\underline{W}\underline{p}]}{I + \underline{p}^T\underline{W}\underline{p}} = \alpha . \tag{13}$$

Since α must satisfy (11), and since λ_m is the largest eigenvalue of $(\underline{A} - \alpha\underline{W})$, then $I\alpha = \lambda_m$ and

$$(SNR)_{opt} = \alpha = \lambda_m/I. \tag{14}$$

The second class of images in Figure 1 can be treated as a stationary white noise part with power I and a nonstationary noise part with the correlation matrix $\underline{W}$ defined in (6). The values of the elements of the matrix $\underline{W}$ are indicative of the amount of nonstationary noise present in image class $\{y(t)\}$. To determine the effects of nonstationary noise on the optimal filter, we rewrite (9) as

$$SNR = \frac{\underline{p}^T\underline{A}\underline{p}}{I\underline{p}^T\underline{p} + \underline{p}^T\underline{W}\underline{p}}$$

$$= \gamma\frac{\underline{p}^T\underline{A}\underline{p}}{\underline{p}^T\underline{p} + \gamma\underline{p}^T\underline{W}\underline{p}}$$

$$= \gamma SNR^1 , \tag{15}$$

where

$$\gamma = \frac{1}{I} ,$$

and

$$SNR^1 = \frac{\underline{p}^T\underline{A}\underline{p}}{\underline{p}^T\underline{p} + \gamma\underline{p}^T\underline{W}\underline{p}} . \tag{16}$$

We see from (15) that optimizing SNR^1 is equivalent to optimizing SNR for a given γ. For a fixed amount of nonstationary noise, γ will decrease as the amount of stationary noise I increases. Thus, γ can be viewed as a measure of the ratio of the nonstationary noise power to the white noise power in the image. By analogy with the optimal weight solution of (11) for (9), the optimal weight $\underline{p}$ to maximize SNR^1 in (16) is the solution to

$$(\underline{A} - \alpha\gamma\underline{W})\underline{p} = \alpha\underline{p} . \tag{17}$$

The iterative procedure outlined previously is also applicable to the solution to (17).

From (17), we see that the optimal weight vector $\underline{p}$ will differ for different choices of γ. Since γ is not easily measurable for a general problem, we must estimate γ using a priori information about whether the data is dominated by stationary noise (small γ) or non-stationary noise (large γ) and then use the appropriate filter h derived from $\underline{p}$ in (17). For $\gamma = 0$, we obtain the stationary noise case and (11) and (17) reduce to $\underline{A}\underline{p} = \alpha\underline{p}$, which states that the eigenvalues and eigenvectors of $\underline{A}$ alone determine $\underline{p}$ in this case. We will refer to the solutions $\underline{p}$ to (17) for $\gamma \neq 0$ as a modified K-L solution or MKL(γ) solutions.

4. EXPERIMENTAL VERIFICATION

We have shown that the iterative solution to (11) or (17) can be used to obtain the optimal weight vector $\underline{p}$ and to have the coefficients for the discrimination function h in (7), where the $\{\psi_i\}$ are the dominant eigen-images (with the largest eigenvalues) of the data set $\{x_i\}$. Use of the iterative solution in Section 3 requires knowledge of the correlation matrix $\underline{A}$ for the $\{x_i\}$ and the correlation matrix $\underline{W}$ for the nonstationary noise or the second class of data $\{y_i\}$. Once the ψ_i and λ_i eigenvectors and eigenvalues of $\{x_i\}$ have been obtained, we can use them to compute $\underline{A}$ and $\underline{W}$ using

$$A_{i,j} = \lambda_i \delta_{i,j}, \tag{18}$$

and

$$W_{i,j} = \frac{1}{M} \sum_{k=1}^{M} (y_k \cdot \psi_i)(y_k \cdot \psi_j) \tag{19}$$

where in (19), M is the number of y_k patterns, and the $y_k \cdot \psi_j$ are the projections of the elements y_k onto the ψ_j, i.e. each y_i can be written as

$$y_i = \sum_{k=1}^{N} (y_i \cdot \psi_k)\psi_k \ . \tag{20}$$

In this section, we present digital simulation results to verify the performance of our optimal discriminator for a specific two-class problem. As our data base, we consider two classes of geometrical shapes: circles and hexagons. Randomness is introduced in each class by varying the radius of the circles and hexagons from 10.0 units to 10.95 units in steps of 0.05 units. Each picture is recorded in a 32 × 32 array and each class contains M = 20 such pictures.

We denote the M = 20 circles by $\{x_i\}$ and the M = 20 hexagons by $\{y_i\}$. We computed the N eigenvectors $\{\psi_i\}$ and associated eigenvalues from the $\{x_i\}$. For simplicity, we retained the first five eigenvectors denoted by $\{\phi_i\}$, where i = 1,...,5 (those with the largest eigenvalues). For simplicity, we used these $\{\phi_i\}$ as the basis set for the optimal filter f rather than computing the composite basis set $\{\psi_n\}$ that includes the primary components of both $\{x_i\}$ and $\{y_i\}$. The $\underline{A}$ and $\underline{W}$ matrices needed in (11) and (17) were then formed as in (18) and (19) with N = 5 basis functions and M = 20 images. The iterative procedure described in Section 3 was then used to compute the solution $\underline{p}$ to (17) for different values of γ. The principal eigenvector h found for γ = 0 (data dominated by stationary noise) and for γ = 50 (data dominated by nonstationary noise) are shown in Figures 2a and 2b respectively. In Figure 2a (γ = 0 case), the exact K-L technique was used, whereas for Figure 2b (γ = 50 case), our iterative MKL method was employed. Both figures appear quite similar, except for small contrast differences. Our iterative solution thus appears to produce results quite similar to those obtained using the exact but computationally expensive K-L analysis.

The optimal discrimination filters h were calculated for eighteen different values of the parameter γ and each of these h's was cross-correlated with all 20 circles and all 20 hexagons. The average of the correlation intensities at the origin (correct peak location) was then calculated (for the cases of circle and hexagon inputs) for each different γ. These results are presented in Table 1, where C-PEAK and H-PEAK denote the average correlation peak intensities for the case of the circles and the hexagons respectively. In general, γ is not easily measurable for a given two-class problem and must thus be appropriately chosen for a proper filter design.

(a) KL (b) MKL(60)

FIGURE 2 K-L and MKL filters, where intensity is quantized to 16 levels (with 0 to 9 and + denoting positive values and the letters A to E denoting negative values).

(a) The primary K-L component of the 20 circles.
(b) The optimal MKL(γ) discriminator for a power ratio γ = 50.

TABLE 1: *Optimal discriminator performance* by changing the ratio of the nonstationary noise power to the white noise power.

γ	C-PEAK	H-PEAK	C/H
0	0.8255	0.1755	4.70
1	0.8056	0.1159	6.95
2	0.7751	0.0852	9.10
3	0.7483	0.0681	10.99
4	0.7265	0.0577	12.59
5	0.7090	0.0509	13.93
10	0.6580	0.0368	17.88
15	0.6328	0.0322	19.64
20	0.6166	0.0299	20.62
25	0.6044	0.0284	21.26
30	0.5941	0.0273	21.75
35	0.5848	0.0264	22.15
40	0.5760	0.0256	22.49
45	0.5675	0.0249	22.81
50	0.5594	0.0242	23.09
125	0.4463	0.0169	26.43
250	0.3201	0.0105	30.42
∞	0.0314	0.0002	128.25

γ: The ratio of the nonstationary noise power to the white noise power.
C-PEAK: Average peak value of the cross-correlation for 20 circles.
H-PEAK: Average peak value of the cross-correlation for 20 hexagons.
C/H: Ratio of C-PEAK to H-PEAK.

We see from Table 1, that the filters h give larger relative values for C-PEAK than for H-PEAK for all γ values as expected. We also include the ratio C/H = C-PEAK/H-PEAK in Table 1 and note that it is always greater than 4.70. This indicates that the optimal discriminator results in high average correlation peaks for circles and low average correlation peaks for hexagons. The variance of these peak correlation values from the average levels quoted was moderate. From this example, we see that our optimal discriminator fulfills the intended purpose of discriminating two classes of data (here circles and hexagons were used). The ratio (C/H) is a measure of discrimination performance of the filter. For this simulation example, C/H is seen to improve as the value chosen for the filter parameter γ used in (17) is increased. This occured because no white noise was included in our simulations and thus the input data set had $\gamma_D = \infty$. Thus, for this data test, we expect to find that the optimal filter would have a parameter γ that matched the γ_D of the data set. In general, a filter designed for a large γ will better discriminate against the class of hexagons. However, higher values of γ yield smaller values of C-PEAK (as can be seen in Table 1) and this will make detection of the correlation outputs more difficult and the measured SNR values susceptible to system noise. If the assumed white noise for which the filter is to be designed is low (i.e., high γ), then a lower C-PEAK value will be adequate to provide sufficient detection. The smaller C-PEAK values that resulted for larger choices of the filter parameter γ in Table 1 show this trend. Thus γ must be chosen as a compromise between the two conflicting objectives of large C/H and large C-PEAK. This choice can best be made only in specific applications and pattern recognition scenarios.

Numerical values associated with the correlations of the K-L and MKL filters with the circles and hexagons in the input data sets $\{x(t)\}$ and $\{y(t)\}$ are presented in detail in Table 2. Table 2a lists the correlation peak values obtained when the circles were correlated with the K-L and MKL filters. We see from this table that the K-L filter yields cross correlation peak values between 0.75 and 0.93 whereas the MKL (50) filter provides correlation peak values between 0.22 and 0.89. We note that in general, smaller peak values were obtained as γ was increased. This is as expected because the K-L ($\gamma = 0$) filter should be the optimal filter in the presence of white noise.

The cross correlation peak values obtained with hexagons as inputs are shown in Table 2b. The K-L filter is seen to correlate much better with the hexagons than the MKL filters do. As a result, the simple K-L filter cannot discriminate between circles and hexagons as well as the MKL filters can. This aspect is more vividly demonstrated in Table 2c, where the ratios of the correlation peak values in Table 2a to those in Table 2b are tabulated. We see that MKL filters provide much better discrimination (higher ratios) than simple K-L filters.

TABLE 2: Performance of discriminator.

(a) Peak value of correlation with circle

R	KL	MKL(1)	MKL(25)	MKL(50)
10.1	0.7593	0.4271	0.2707	0.2204
10.3	0.9056	0.7101	0.5735	0.5266
10.5	0.9284	0.8089	0.6759	0.6133
10.7	0.8860	0.8869	0.7787	0.7042
10.9	0.7347	0.8699	0.8833	0.8899

(b) Peak value of correlation with hexagons

R	KL	MKL(1)	MKL(25)	MKL(50)
10.1	0.0846	0.0536	0.0422	0.0418
10.3	0.1020	0.0648	0.0511	0.0507
10.5	0.1397	0.0620	0.0543	0.0639
10.7	0.1935	0.0770	0.0614	0.0631
10.9	0.4527	0.1961	0.1086	0.0934

(c) Ratio of cross correlation peaks with circle to cross correlation peaks with hexagons

R	KL	MKL(1)	MKL(25)	MKL(50)
10.1	8.98	7.97	6.41	5.27
10.3	8.88	10.96	11.22	10.39
10.5	6.65	13.05	12.45	9.60
10.7	4.58	11.52	12.68	11.16
10.9	1.62	4.44	8.13	9.53

5. CONCLUSIONS

The optimal correlator which discriminates between two classes was obtained by applying a new image model to the conventional matched filter derivation. In our new model, one class was considered as stochastic signals and the other class as a nonstationary noise process. A primary application for this two-class discriminator is to recognize and distinguish two objects, when distorted versions of each can be present. In our model, each image class was described by a linear combination of basis functions whose weighting coefficients were random. This new description for a set of geometrically distorted versions of an object allows us to formulate a new two-class discriminator.

Using the output SNR as our optimality criteria, we derived the optimal filter. We also proposed a new iterative procedure that makes the discriminator's design practical.

This new discriminator synthesis concept was tested on a data base consisting of 20 circles and 20 hexagons, each with a different radius. From our results, we found that there were tradeoffs between the discrimination power of the filter and the correlation output peak values for a given discriminant filter and that these could be adjusted by changing one filter design parameter γ, the ratio of the nonstationary noise power to the white noise power in the image set. We showed that by changing γ, we could make the filter perform more efficiently in the stationary or nonstationary noise cases.

REFERENCES

1. R.O. Duda and P.E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York (1973).

2. K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, New York (1972).

3. K. Fukunaga and W.L.G. Koontz, "Application of the Karhunen-Loeve Expansion to Feature Selection and Ordering", IEEE Trans.Comput., Vol. C-19, 311-318 (1970).

4. D.H. Foley and J.W. Sammon, Jr., "An Optimal Set of Discriminant Vectors", IEEE Trans.Comput., Vol C-24, 281-88 (1975).

5. Y.T. Chien and K.S. Fu, "On the Generalized Karhunen-Loeve Expansion", IEEE Trans.Inform.Theory, Vol. IT-13, 518-20 (1967).

6. K. Fukunaga and W.L.G. Koontz, "Representation of Random Processes Using the Finite Karhunen-Loeve Expansion", Information and Control, Vol. 16, 85-101 (1970).

7. H.J. Caulfield and R. Haimes, "Composite Matched Filters", Applied Optics, 19, 181-83 (1980).

8. C.W. Helstrom, Statistical Theory of Signal Detection, Pergamon Press, New York (1968).

ACKNOWLEDGEMENT

The support of the Air Force Office of Scientific Research (Grant AFOSR 79-0091) for this research is gratefully acknowledged, as is the support of H. Murakami by Toshiba Corporation during his stay at Carnegie-Mellon University.

APPENDIX

We now derive the solution $\underline{p}$ to the optimization problem of (10) by using the Lagrange multiplier method. Defining

$$L = \frac{\underline{p}^T \underline{A} \underline{p}}{I + \underline{p}^T \underline{W} \underline{p}} - \lambda(\underline{p}^T \underline{p} - 1), \tag{A1}$$

where λ is a scalar, we observe the deviation δL of L with respect to the deviation $\delta \underline{p}$ of $\underline{p}$. From (A1), δL can be written as

$$\delta L = \frac{(\delta \underline{p}^T \underline{A} \underline{p})(I + \underline{p}^T \underline{W} \underline{p}) - (\delta \underline{p}^T \underline{W} \underline{p})(\underline{p}^T \underline{A} \underline{p}) - \lambda(\delta \underline{p}^T \underline{p})(I + \underline{p}^T \underline{W} \underline{p})^2}{(I + \underline{p}^T \underline{W} \underline{p})^2} \tag{A2}$$

If there exists a $\underline{p}$ such that $\delta L = 0$ for any $\delta \underline{p}$, then that $\underline{p}$ value determines the maximum value of L. By assuming the existence of such a $\underline{p}$, we obtain from (A2) with $\delta L = 0$ for any $\delta \underline{p}$,

$$(I + \underline{p}^T \underline{W} \underline{p}) \underline{A} \underline{p} - (\underline{p}^T \underline{A} \underline{p}) \underline{W} \underline{p} - \lambda(I + \underline{p}^T \underline{W} \underline{p})^2 \underline{p} = 0 . \tag{A3}$$

By premultiplying (A3) by $\underline{p}^T$, we obtain

$$\lambda(I + \underline{p}^T \underline{W} \underline{p})^2 = I \underline{p}^T \underline{A} \underline{p}. \tag{A4}$$

Using (A4) in (A3), we obtain

$$\{\underline{p}^T (I\underline{U} + \underline{W})\underline{p}\} \underline{A} \underline{p} - (\underline{p}^T \underline{A} \underline{p})(I\underline{U} + \underline{W})\underline{p} = 0 , \tag{A5}$$

where $\underline{U}$ is a unit matrix of size N × N. From (A5), we observe that

$$\underline{A} \underline{p} = \alpha(I\underline{U} + \underline{W})\underline{p} , \tag{A6}$$

where α is a scalar. (A6) can also be rewritten as

$$(\underline{A} - \alpha \underline{W})\underline{p} = \alpha I \underline{p}. \tag{A7}$$

Thus the optimal weight $\underline{p}$ in (10) can be obtained by solving (A7).

7. SYNTHETIC DISCRIMINANT FUNCTIONS FOR 3-D OBJECT RECOGNITION

360-24

SYNTHETIC DISCRIMINANT FUNCTIONS FOR 3-D OBJECT RECOGNITION

David Casasent, B.V.K. Vijaya Kumar and Vinod Sharma

Carnegie-Mellon University
Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

ABSTRACT

The synthetic discriminant function concept together with its modifications of maximum common information filters and decorrelation transformations are reviewed. We then advance a unified procedure for determining the coefficients for such linear combination filters for recognition of objects in different orientations and from different aspect views. Our formulation utilizes only deterministic techniques and a correlation matrix observation space. This formulation is most attractive for the realization of shift-invariant filters for use in correlator architectures. We then advance the highlights of our initial results on the performance of this new type of generalized shift-invariant filter.

1. INTRODUCTION

In the optical data processing community, considerable attention has recently been focused on the use of off-line matrix techniques applied to image training sets to derive filter functions that are capable of recognizing and distinguishing objects independent of geometrical distortions in the input image. In Section 2, we summarize much of the research in this area. To permit such a summary, we consider only deterministic techniques applied to image training set data and to techniques in which the resultant filter(s) is (are) linear combination (s) of the input training set data. We emphasize the differences between possible observation spaces used (Fourier transform coefficients [1-3], correlation matrices [4-7], etc.) and different applications (intra-class recognition of a target object of one class independent of geometrical distortions present in the input image [4-6], inter-class discrimination of different objects with no geometrical distortions [1] and prior work [3,7] and new research in which inter-class discrimination is achieved while retaining intra-class recognition). We consider only cases in which shift-invariance is retained and thus do not consider the wealth of research using other observation spaces and image features such as the moments [8-10], coded-phase processors [11-13], Mellin transforms [14] and techniques with no organized feature selection.

In Section 3, we review the concepts of a synthetic discriminant function (SDF) [4-5], maximum common information (MCI) filter [6] and the decorrelation transformation [7]. In our discussion, we include a review of the hyperspace formulation of an SDF and inter-class problems plus the effects of noise. These provide the basis for our description in Section 4 of a class of linear combination filters obtained by deterministic techniques. This formulation in Section 4 represents a new unified treatment that emphasizes how the coefficients for these SDFs are obtained. This differs from another unified treatment [15] that emphasized the philosophy of such filters and the fact that they are linear combinations of filters matched to each of the input training set objects. Our emphasis in Section 4 is on calculation of the weights in a linear combination filter (this issue is not addressed in [15]). In Section 5, we include initial results indicating the power of these SDF techniques for intra and inter-class pattern recognition for identification of objects in multiple classes with geometrical distortions present.

2. HISTORICAL DEVELOPMENT

In 1969, Caulfield and Maloney [16] considered the correlations of the letters of the alphabet with matched spatial filters (MSFs) matched to each letter. They noted that each input letter gave large cross-correlation outputs with many of the different MSFs. They then reasoned that if linear combinations of all 26 correlation outputs were used rather than a thresholded version of each correlation output, character recognition could be improved. Recent research has noted that an MSF that is a linear combination of the MSF of each character could be used [1] and that such a description was appropriate for intra-class and inter-class recognition [15]. The resultant MSFs are referred to as generalized matched filters (GMFs) [1-3], synthetic discriminant functions (SDFs) [4-7], and similar terms. The key feature in these approaches is that the SDF is a 2-D function through which the input image is projected or correlated. This differs from the classical feature extraction and image classification techniques in which scalar features are extracted from a segmented portion of an input image and then used in a classifier. Our SDFs are 2-D functions and are used in a correlator. Thus, they exhibit processing gain, shift-invariance, the ability to

recognize multiple targets and operate in noisy backgrounds without the need for segmentation and extensive preprocessing. For these reasons, such techniques have received considerable attention in recent years (especially within the optical processing community where correlation is easily achieved).

In [1], Caulfield and Haimes discussed GMFs and their similarity to linear discriminant functions in conventional pattern recognition. In [4-5], Hester and Casasent detailed and demonstrated the use of a matrix-vector technique by which an SDF could be obtained that gave the same correlation output intensity for any aspect view of a given object. We refer to this as an equal correlation peak (ECP) SDF and note that it addresses only the intra-class pattern recognition problem. Fourier transform and correlation matrix observation spaces have been the most used. The filter obtained from such observation spaces is easily fabricated as an MSF and is thus useable in an optical or digital correlator. Thus, we restrict attention to such cases. Many techniques are possible by which to obtain the weights. Those used and documented thusfar include: Foley-Sammon (F-S) [2,3,17] which optimizes the Fisher ratio [18]; Gram-Schmidt (G-S) [5]; optimization of correlation output SNR [5]; maximum common information (MCI) SDF [6]; decorrelation transformations [7]; and statistical Karhunen-Loeve (K-L) techniques [7] as used by Duvernoy and Leger [19] and Fukanaga-Koontz (F-K) [20] methods [12]. Each of these leads to an SDF that is a linear combination of the input training set data. However, each approach is intended for different purposes and depending upon how they are realized, shift-invariance is not always achievable.

3. MCI AND DECORRELATION TRANSFORMATION SDFs

Our initial formulation of an SDF [4,5] considered only the intra-class problem or an ECP SDF. To describe this problem, we consider an input training set of images $\{f_n\}$ of objects of one class taken from different aspect views. We desire to derive an SDF $\underline{h}$ that is a linear combination of the $\{f_n\}$ such that $\underline{f}_n \cdot \underline{h} = c = 1$ (where we arbitrarily choose unity as the constant c output from the correlation of $\underline{h}$ with any input image in the data set). To obtain $\underline{h}$, we write each $\underline{f}_n$ as a linear sum of a set of basis functions $\underline{\phi}_m$

$$f_n(x,y) = \sum_m a_{nm} \phi_m(x,y). \tag{1}$$

In this description, the ϕ_m are the axes of a hyperspace (each is a 2-D function) and each $\underline{f}_n$ is an M-dimensional vector defining a point in this hyperspace with the projections on each $\underline{\phi}_m$ axis being described by the coefficients a_{nm}. We write $\underline{h}$ as a linear sum of the $\underline{\phi}_m$,

$$h(x,y) = \sum_m b_m \phi_m(x,y). \tag{2}$$

Our ECP condition requires (for an orthonormal basis function set ϕ_m)

$$f \circledast h = \underline{f}_n \cdot \underline{h} = \sum_m a_{nm} b_m = c = 1. \tag{3}$$

We determine the ϕ_m and a_{nm} by diagonalizing the correlation matrix

$$\underline{R} = \underline{R}_{f_i f_j} = \underline{f}_i \cdot \underline{f}_j = \{r_{ij}\}. \tag{4}$$

We originally used a Gram-Schmidt technique to achieve diagonalization of R by selection of a set of orthogonal basis functions. Once this has been done, the ϕ_m and a_{nm} are then given by

$$\underline{\phi}_m = \sum_n d_{mn} \underline{f}_n, \tag{5}$$

$$a_{nm} = \underline{f}_n \cdot \underline{\phi}_m, \tag{6}$$

where the d_{mn} are the elements of the G-S coefficient matrix (assuming a G-S decomposition is used, as we have initially employed). With the $\underline{\phi}_m$ and a_{nm} so determined, the b_m are then obtained from (3) and $\underline{h}$ is then constructed. Denoting the elements of the target matrix $\underline{F}_T$ by the a_{nm} in (1), we described [4-6] condition (3) by

$$\underline{F}_T \cdot \underline{h} = c\,\underline{u}, \tag{7}$$

where $\underline{u}$ is the unit vector and the solution for the SDF $\underline{h}$ is

$$\underline{h} = \underline{F}_T^{-1}\, c\,\underline{u}. \tag{8}$$

This describes $\underline{h}$ in terms of $\underline{\phi}_m$. Substituting (5) into (8), we obtain an expression for $\underline{h}$ in terms of the training set data $\underline{f}_n$. In Section 4, we advance a direct derivation of this ECP filter function $\underline{h}$ from the correlation matrix rather than including the intermediate basis function description.

To improve the performance of this SDF, we [6] analyzed the effects of noise in a hyperspace and showed that maximizing the correlation plane SNR required maximizing the filter energy. In [6], we achieved this by shifting each of the images in our training set to maximize cross-correlations (the elements of the covariance matrix). This MCI SDF yielded excellent results. As a data base, we used several IR images of a tank from different aspect views (typical images are shown in Figure 1). The resultant correlation outputs exhibited correlation peak intensity values that were equal within 5%.

f_1 (RIGHT SIDE)

f_5 (REAR/SIDE)

FIGURE 1 Typical images used in initial maximum common information synthetic discriminant function synthesis and correlation tests.

To extend this SDF technique to inter-class discrimination, we [7] suggested and demonstrated the use of decorrelation techniques, K-L techniques and multiple SDFs. We review our decorrelation technique below as it forms the basis for our mutual orthogonal function SDF technique described in Section 4 and our method to realize the orthogonalized correlations of [1,15] from a correlation matrix observation space. We first [7] formed an MCI SDF from two tank images and then tested its discrimination ability against an APC input object. The correlation peak intensity for the APC was 7dB below that for the tank images. This indicated the inherent structure and discrimination ability of the MCI SDF. To improve this discrimination ability, we devised the decorrelation transformation in which the system was trained with N images of one target and then N images of a second target object. We used a G-S basis function generation technique and the property of the G-S decomposition that basis function ϕ_m is a function of the input images up to f_m only. We formed a 2N x 2N target matrix $\underline{F}_T$. We then multiplied this by a decorrelation transformation matrix $\underline{\mu}$ that retained only the last N rows and columns of $\underline{F}_T$. This new $\underline{F}_T$ target matrix was then inverted and used in (8) to derive a decorrelation SDF $\underline{h}'$. This $\underline{h}'$ has all of the information of the first class of targets removed and contains only those portions of the second class of objects not present in the first class of targets. It should thus be very successful in producing a zero output for any input image in class 1 and a unit output for any image in class 2. We successfully demonstrated this filter in a simple test using two objects $\underline{f}_1$ and $\underline{f}_5$ from one class and an object $\underline{f}_f$ from another class. In Figure 2, we show the resultant matrices. The final filter function used was a linear combination of all training set images

$$\underline{h}' = \underline{f}_1 + 0.42\underline{f}_5 - 0.1\underline{f}_f. \tag{9}$$

This filter gave equally large outputs for inputs $\underline{f}_1$ and $\underline{f}_5$ and zero outputs for input object $\underline{f}_f$.

	f_f	f_1	f_5
f_f	2.68	0.54	0.91
f_1	0.54	8.79	1.94
f_5	0.91	1.94	18.9

(2a)

	f_f	f_1	f_5
ϕ_1	0.61	0	0
ϕ_2	-0.06	0.33	0
ϕ_3	-0.07	-0.04	0.23

(2b)

	ϕ_1	ϕ_2	ϕ_3
f_f	1.63	0	0
f_1	0.32	2.98	0
f_5	0.54	0.05	4.26

(2c)

	ϕ_2	ϕ_3
f_1	2.98	0
f_5	0.05	4.26

(2d)

	54.6	0
	-0.69	38.2

(2e)

FIGURE 2 Data matrices for the initial demonstration of the decorrelation transformation. (a) Auto-correlation matrix, (b) G-S coefficient matrix, (c) full target matrix $\underline{F}_T$, (d) reduced target matrix $\underline{F}_T'$ and (e) inverse of reduced target matrix $\underline{F}_T'^{-1}$.

4. UNIFIED SDF SYNTHESIS TECHNIQUE

From our discussions in Sections 1 and 2, we see that various SDFs and GMFs exist for diverse purposes. In this section, we describe four types of SDFs and show how the synthesis of each can be described in one general expression involving the correlation matrix. As our observation space, we choose the correlation matrix of the training set data. Such an approach leads directly to an SDF that is a linear combination of the training set data and to evaluation of the coefficients necessary.

We first consider an ECP SDF $\underline{h}$ that will yield an equal correlation peak intensity output $\underline{h} \cdot \underline{f}_n = 1$ for any aspect view $\{\underline{f}_n\}$ of an object of one class. As in Section 3, we write $\underline{f}_n$ and $\underline{h}$ as

$$\underline{f}_n = \sum_j a_{nj} \underline{\phi}_j, \quad \underline{h} = \sum_j b_j \underline{\phi}_j \tag{10}$$

and we require the correlation coefficient of $\underline{h}$ with any $\underline{f}_n$ to equal a constant 1, i.e.

$$h \circledcirc f_n = \underline{h} \cdot \underline{f}_n = \Sigma a_{nj} b_j = 1, \tag{11}$$

where the orthonormality of the $\underline{\phi}_j$ was used in (11). To find $\underline{h}$, we first rewrite (10) as

$$\underline{\phi}_j = \sum_n d_{jn} \underline{f}_n \tag{12}$$

and $\underline{h}$ as

$$\begin{aligned} h &= b_1\phi_1 + b_2\phi_2 + \ldots \\ &= b_1 \textstyle\sum_n d_{1n} f_n + b_2 \sum_n d_{2n} f_n + \ldots \\ &= e_1 f_1 + e_2 f_2 + \ldots \\ &= \textstyle\sum_n e_n f_n, \end{aligned} \tag{13}$$

where the first expression is (10) rewritten, the second expression follows by substitution of (12) and the last expression is obtained by grouping f_n terms. To obtain $\underline{h}$, we must now solve $h = \sum_n e_n f_n$ for the coefficients e_n. Substituting (13) into (11), we obtain our ECP condition as

$$h \cdot \underline{f}_m = 1 = (\sum_n e_n \underline{f}_n) \cdot \underline{f}_m = \sum_n e_n (\underline{f}_n \cdot \underline{f}_m) = \sum_n e_n R_{mn} \tag{14}$$

or

$$\underline{R}\underline{e} = \underline{u}, \tag{15}$$

where $\underline{R}$ is the correlation matrix of the training set, $\underline{e}$ are the desired coefficients and $\underline{u}$ is the unit vector. The solution for $\underline{h}$ (or its coefficients $\underline{e}$) is thus

$$\underline{e} = \underline{R}^{-1}\underline{u}. \tag{16}$$

Thus, to derive an ECP SDF, we form the correlation matrix $\underline{R}$ of the training set data of different aspect views $\{f_n\}$ of one class of object. We then invert $\underline{R}$ and multiply it by the unit vector $\underline{u}$ to obtain the coefficients $\underline{e}$ from which to synthesize the linear combination SDF $\underline{h}$. This follows the ideas used earlier [4-6] with the correlation matrix in (16) replacing the target matrix in (8) and without consideration of the details of the basis function selection used. Our original technique [4-6] simply specified a Gram-Schmidt technique by which to find amatrix to invert. This ECP SDF is useful for intra-class pattern recognition.

Next, we consider a combined inter-class discrimination problem that still retains intra-class recognition. We consider three classes of objects $\{f_a\}$, $\{f_b\}$ and $\{f_c\}$. We desire three filters $\underline{h}_1$, $\underline{h}_2$ and $\underline{h}_3$ such that: for any input $\underline{f}_n$, $\underline{h}_1 \cdot \underline{f}_n$ gives a one output if f_n is in class $\{f_a\}$ and zero otherwise; $\underline{h}_2 \cdot \underline{f}_n = 1$ if $\underline{f}_n$ is in class $\{f_b\}$, etc., i.e.

$$\underline{h}_m \cdot \underline{f}_n = \delta_{mn}. \tag{17}$$

We write these three filters as three separate linear combinations of the entire training set $\{f_n\} = \{f_a, f_b, f_c\}$

$$\underline{h}_1 = \Sigma\, a_n \underline{f}_n,\ \underline{h}_2 = \Sigma\, b_n \underline{f}_n,\ \underline{h}_3 = \Sigma\, c_n \underline{f}_n. \tag{18}$$

In terms of the correlation matrix $\underline{R}$ of the full $\{f_n\}$ data set, we can describe the three filters (i.e. the coefficient vectors $\underline{a}$, $\underline{b}$, $\underline{c}$) by

$$\underline{R}\underline{a} = \underline{u}_1 = [1\ 0\ 0]^T,\ \underline{R}\underline{b} = \underline{u}_2 = [0\ 1\ 0]^T,\ \underline{R}\underline{c} = \underline{u}_3 = [0\ 0\ 1]^T, \tag{19}$$

where the number of 1's and 0's in the output vectors $\underline{u}$ depend upon the number of elements N in each class in our training set. The solutions for these three filters are thus from (19) and (16)

$$\begin{aligned}
\underline{a} &= \underline{R}^{-1}\underline{u}_1 = \underline{R}^{-1}[1\cdots1,\ 0\cdots0,\ 0\cdots0]^T \\
\underline{b} &= \underline{R}^{-1}\underline{u}_2 = \underline{R}^{-1}[0\cdots0,\ 1\cdots1,\ 0\cdots0]^T \\
\underline{c} &= \underline{R}^{-1}\underline{u}_3 = \underline{R}^{-1}[0\cdots0,\ 0\cdots0,\ 1\cdots1]^T.
\end{aligned} \tag{20}$$

Note that these two filters provide inter-class discrimination and intra-class recognition. The concept used to synthesize them follows directly from our decorrelation transformation, but is applied to the correlation matrix. Note that these filters are also equivalent to the orthogonalized correlation functions described in [1,15]. Moreover, note that the form for synthesis of these filters in (20) is the same as used in (16) with the exception of the exogenous vector. In (16), $\underline{u}$ is all 1's, whereas in (20) it contains a 1 only for those members of the input training set that we wish the given mutual orthogonal function (MOF) SDF to recognize.

This MOF technique can be extended to an N-class problem by devising N MOF SDFs and analyzing the N output correlation planes. This can become quite complex if a large multi-class problem is involved. In such cases, use of non-binary correlation output threshold levels can be employed. In this case, a single filter can be designed to perform multi-class pattern recognition. To describe this and to formulate synthesis of such a filter in our general form in (16) and (20), we consider the three class problem described above. However, now we require one filter $\underline{h}$ such that

$$\begin{aligned}
\underline{h} \cdot \underline{f}_n &= 1 \text{ if f is in class } \{f_a\} \\
\underline{h} \cdot \underline{f}_n &= 2 \text{ if f is in class } \{f_b\} \\
\underline{h} \cdot \underline{f}_n &= 3 \text{ if f is in class } \{f_c\}.
\end{aligned} \tag{21}$$

We write this filter as a linear combination of the entire reference set $\{f_n\}$ as

$$\underline{h} = \sum_n a_n \underline{f}_n. \tag{22}$$

To find the coefficients a_n, we solve

$$\underline{R}\underline{a} = \underline{u}_4 = [1\cdots1,\ 2\cdots2,\ 3\cdots3]^T. \tag{23}$$

This follows directly from (21) and (15). As before, the number M of 1's, 2's and 3's in the vector $\underline{u}_4$ depends upon the number of images present in each class in the training set. The solution for the coefficients $\underline{a}$ of this nonredundant SDF (NRF) is thus

$$\underline{a} = \underline{R}^{-1}\underline{u}_4 = \underline{R}^{-1}[1\cdots1,\ 2\cdots2,\ 3\cdots3]^T. \tag{24}$$

Fleuret and Maitre [21] described an algorithm which used K SDFs (with binary outputs, i.e. 1 or 0) to achieve recognition of $N = 2^K$ classes of objects. In this case, the K correlation outputs are viewed as a K-bit digital word whose decoded output tells us which of the N classes of data the input image belongs as also described by Braunecker et al [22]. We now adapt this technique into a coherent correlator (only noncoherent correlators were addressed in [22]) and we describe it in our general form. We refer to this as a multi-class MOF. We consider its use for a 4-class problem $\{f_a\}$ to $\{f_d\}$. We require two filters $\underline{h}_1$ and $\underline{h}_2$ such that the correlations of the input with the two filters yields the two correlation outputs: 00, 01, 10, and 11 respectively (the binary combination that occurs determines which of the four classes of data is present in the input). To describe the $\underline{h}_1$ and $\underline{h}_2$ solution, we denote the full data set by $\{f_n\} = \{f_a, f_b, f_c, f_d\}$, its correlation matrix by $\underline{R}$ and the two filters by

$$\underline{h}_1 = \sum a_n \underline{f}_n,\ \underline{h}_2 = \sum b_n \underline{f}_n. \tag{25}$$

To determine $\underline{a}$ and $\underline{b}$ and hence the two filters, we solve a matrix equation of the same form as (16), (20) or (24) with yet a different vector $\underline{u}$ used.

We have thusfar seen how a single unified formulation involving the inversion of the correlation matrix of the data and multiplication by a simple vector can be used to describe the coefficients required in many linear combination SDFs. The cases considered above include intra-class, inter-class and multi-class pattern recognition. As the number of images used in each of the classes is increased, these filters can provide both inter-class discrimination and intra-class recognition in the face of various geometrical and other distortions of the input data.

5. INITIAL RESULTS AND CONCLUSIONS

We have used the general deterministic formulation from a correlation matrix observation space for synthesis of various SDFs for intra and inter-class recognition. In our initial new tests, we have used images of four different objects with 36 images of each object available (taken at 10° intervals from an 0° depression angle). As our training set, only 6 images of each class of objects (out of 36 possible images) were used. In each case, a particular SDF was synthesized as described in Section 4. This SDF was then correlated against all 36 images in each of the indicated object classes. An ECP SDF was produced for the class one objects and a second one for the class two objects. Each of these SDFs was capable of recognizing all 36 objects in each of their particular classes with no errors and with less than a 3% variation in the correlation peak intensity obtained. This initial experiment demonstrated the ability of an ECP SDF to recognize objects of a different class not present in the training set. A 2-class MOF filter was produced. This filter was tested against all 72 images in each of the two classes. In demonstrated inter-class discrimination and intra-class recognition with over 90% probability of correct recognition. The 4-class MOF we formed achieved similar results. These initial tests were most encouraging.

In this paper, we have reviewed several different types of SDFs (equal correlation peak, maximum common information, decorrelation transformation, mutual orthogonal function, nonredundant, and multi-class). Each is a linear combination of the input training set. All have been unified into a direct technique to determine the coefficients by a similar matrix-vector equation with the matrix being the correlation matrix and with a different vector used for each type of SDF. Each SDF is intended for a different application (intra-class, inter-class or both). We have described and experimentally demonstrated an SDF technique that achieves both inter-class discrimination and intra-class recognition and that in this technique only a few images (17% of the data was used in our experiments) can be used for training and yet recognition and correct classification (in multi-class recognition problems) of over 90% correct recognition can be achieved.

REFERENCES

1. H. Caulfield and R. Haimes, Applied Optics, 19, 181 (1980).
2. H. Caulfield, R. Haimes and J. Horner, Israel Journal of Technology, 18, 263 (1980).
3. H. Caulfield and M. Weerberg, Applied Optics, 21, 1699 (1982).
4. C. Hester and D. Casasent, SPIE, 201, 77 (1979).
5. C. Hester and D. Casasent, Applied Optics, 19, 1758 (1980).
6. C. Hester and D. Casasent, SPIE, 292, 25 (1981).
7. C. Hester and D. Casasent, SPIE, 302, 108 (1981).
8. D. Casasent and D. Psaltis, Opt. Lett., 5, 395 (1980).
9. D. Casasent, J. Pauly and D. Fetterly, SPIE, 302, 126 (1981).
10. D. Casasent, L. Cheatham and D. Fetterly, SPIE, 360 (1982).
11. J. Leger and S. Lee, Applied Optics, 21, 274 (1982).
12. J. Leger and S. Lee, JOSA, 72, 556 (1982).
13. Z. Gu, J. Leger and S. Lee, JOSA, 72, 787 (1982).
14. D. Casasent and D. Psaltis, Proc. IEEE, 65, 77 (1977).
15. H. Caulfield, Applied Optics, 19, 3877 (1980).
16. H. Caulfield and W. Maloney, Applied Optics, 8, 2354 (1969).
17. D. Foley and J. Sammon, IEEE, C-24, 281 (1975).
18. R. Fisher, Ann. Eugenics, 7, 179 (1936).
19. J. Duvernoy and J. Leger, Optics Communication, 32, 39 (1980).
20. K. Fukunaga and W. Koontz, IEEE, C-19, 311 (1970).
21. J. Fleuret and H. Maitre, Optics Communication, 17, 64 (1976).
22. B. Braunecker et al, Applied Optics, 18, 2746 (1979).

8. PRINCIPAL COMPONENT IMAGERY FOR STATISTICAL PATTERN RECOGNITION CORRELATORS

Principal-component imagery for statistical pattern recognition correlators

B. V. K. Vijaya Kumar
D. Casasent
H. Murakami*
Carnegie-Mellon University
Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

Abstract. Concepts, measures, and models of image quality are shown to be quite important in pattern recognition applications. Pattern recognition of imagery subjected to geometrical differences (such as scale and rotational changes) and intensity differences (such as arise in multispectral imagery) are considered. After modeling these image differences as a stochastic process, the optimal filter is derived. This filter is shown to be the principal component of the data. This pattern recognition algorithm is verified using multi-sensor imagery, and the results are found to compare favorably to those obtained using other candidate techniques.

Keywords: image quality; pattern recognition.

Optical Engineering 21(1), 043-047 (January/February 1982)

CONTENTS

1. INTRODUCTION

Efforts have recently been made[1,2] to collect research on image information content and image quality measures and to discuss their use in photographic and printing processes as well as in image processing. In this paper, we consider the effects of image quality in pattern recognition applications. Specifically, we consider how to extract the optimum information from a data set and how to use it for selection and synthesis of the optimum filter for pattern recognition. We consider a statistical correlator and the problem of recognizing a reference object in the presence of various geometrical and intensity differences present in the real-time sensed input image. As we will show: Karhunen-Loeve (K-L) transform techniques[3,4] previously used in bandwidth compression,[5,6] techniques for computing the primary K-L components of an image data set,[7] and image or system quality measures such as space-blur bandwidth product[8] are of use in pattern recognition.

Our major concern is how to select and synthesize the optimal filter for a statistical correlator from the image data sets given. The quality of the imagery and its common information clearly affects filter selection. In Sec. 2, we describe our statistical correlator model. It is similar to the conventional one,[9] except that the reference object is not deterministic. In Sec. 3 we show that the principal component of the K-L expansion is the optimal linear filter that maximized the signal-to-noise ratio (SNR) of the output correlation. Simulation results are included in Sec. 4 to verify the performance of such a filter.

2. STATISTICAL CORRELATORS

We consider only correlation techniques for such scene-matching problems, because such methods have proven useful in many cases. A correlation can be realized by both optical[10] and digital[11] methods; however, we make no judgment at this time on which is preferable. A correlation is known to be the optimal operation by which a deterministic reference function can be extracted from additive white Gaussian noise. However, when random geometrical and intensity differences exist between the input and reference imagery, the performance of a correlator rapidly degrades.[12] Although much research exists on general statistical pattern recognition,[3,4] little effort has been devoted to generalized correlators with random image distortions. Prior work in this area has focused attention on performance improvement by deterministic methods such as coordinate transformations[13] and image plane weighting.[14] Most prior statistical pattern recognition work[3,4] has used the divergence between two classes as the performance measure for the system. In this paper, we use correlation SNR since we are considering correlator systems and since this parameter is easily measured for such systems.[8-14]

In Fig. 1, we show the model for the conventional correlator with a deterministic input signal x(t) corrupted by additive noise n(t)

*Present address: Toshiba Corporation, 70, Yanagi-Cho, Saiwai-Ku, Kawasaki, 210 Japan.

Paper IQ-107 received July 10, 1981; revised manuscript received Aug. 14, 1981; accepted for publication Aug. 18, 1981; received by Managing Editor Aug. 21, 1981.

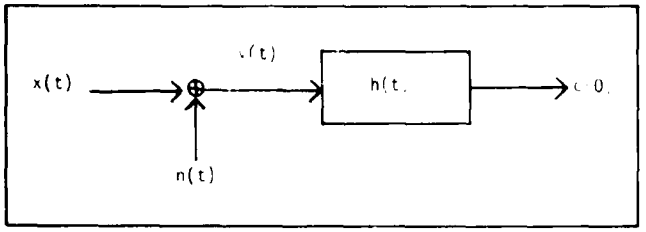


Fig. 1. Conventional correlator model.

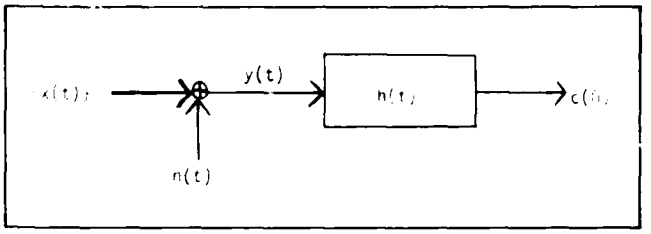


Fig. 2. Statistical correlator model.

producing an observed signal y(t). We will consider only 1-D time domain data for notational simplicity. The extension to the case of 2-D spatial data follows analogously. We denote the correlation output by $c(\tau)$ where τ is the correlation plane shift variable. We assume (with no loss of generality) that the correlation peak occurs at $\tau = 0$. The linear filter h(t) is designed so that the output c(0) emphasizes the signal x(t) and reduces the noise n(t). We can observe that the output $c(\tau)$ contains a deterministic part due to the signal x(t) and a noisy part and that the optimal filter h(t) is designed to maximize the output SNR at $\tau = 0$. The output SNR is expressed as

$$\text{SNR} = \left. \frac{[\int x(t)h(\tau - t)dt]^2}{E[\int n(t)h(\tau - t)dt]^2} \right|_{\tau = 0} . \tag{1}$$

This represents the ratio of filter output powers when the input is only signal and only noise. The optimal filter h(t) is chosen to maximize this SNR because this emphasizes the output power due to the signal while deemphasizing that due to the additive noise. This SNR is also an easily measured parameter. We assume that the target and the correlation peak occur at $\tau = 0$ with no loss of generality. The optimal filter which maximizes this SNR can be shown to be equivalent to the maximum likelihood detector, thus yielding the minimum false alarm. This optimal filter is[9]

$$[h(t)]_{\text{optimal}} = x(-t) .$$

Since this is a time-reversed replica of the signal x(t), this optimal filter is referred to as matched filter. With this choice for the filter, the output $c(\tau)$ is

$$c_{yx}(\tau) = \int y(t)x(t - \tau)\, dt , \tag{2}$$

which is the cross-correlation of x(t) and y(t). Thus the presence of the signal x(t) will be optimally detected by correlating the observed signal y(t) with the true signal x(t).

Our concern is to maintain recognition and to select the best filter h(t) to be used when the observed signal y(t) is a noisy version of a *distorted* signal x(t). These distortions will be known *a priori* only in a statistical manner; e.g., we may be able to bound the range of magnification and orientational differences. To model such a problem and the resultant correlator, we describe the signals (images) of interest by the signal class {x(t)}, where {x(t)} can only be characterized in a statistical manner and where its parameters depend on the maximum amount of distortion that we wish to consider in x(t). The model for the resultant statistical correlator is shown in Fig. 2. The inputs {x(t)} are the class of signals to be detected and all other parameters are as in Fig. 1. The input y(t) to the filter in Fig. 2 is a possible signal x(t) in the class {x(t)} with additive noise n(t).

To derive the optimal filter h(t) for this statistical correlator, new signal and correlation models are necessary as well as a new output correlation SNR measure. We will model the input signal x(t) as a sample realization of the stochastic process {x(t)} as in Ref. 15. We will then define the correlation function in terms of both the time average as in Eq. (2) and in terms of an ensemble average. We describe the time correlation function for the signal x(t) by

$$c_{xx}(\tau) = \int x(t)x(t - \tau)dt \tag{3}$$

as in Eq. (2), where the averaging is performed for different times along a single sample realization of the signal (i.e., a *time-average correlation function* results). We denote the ensemble-average correlation function of the stochastic process {x(t)} by

$$K_{xx}(t,s) = E[x(t)x(s)] . \tag{4}$$

In Eq. (4), the average is over the ensemble of signals (or images) x(t). This is a new and quite different statistical averaging from the one that is usually performed. In normal statistical image processing, the statistics of the pixels of a given image, or blocks of pixels in a given image, are considered and the statistical averaging is performed over these pixels or these blocks of pixels as in Ref. 16. In our pattern recognition, the randomness in the data is between different distorted versions of an image. Thus, in Eq. (4) we form the average over the ensemble of signals (or images) rather than over the pixels in one image. This model and performance measure using the ensemble of signals and SNR has not been used in prior work. We feel that this is a vital step in the design of optimal correlators. As our correlation output SNR measure for the statistical correlator of Fig. 2, we thus use

$$\text{SNR} = \left. \frac{E[\int x(t)h(\tau - t)dt]^2}{E[\int n(t)h(\tau - t)dt]^2} \right|_{\tau = 0} . \tag{5}$$

The major difference between Eq. (1) and Eq. (5) is in the numerator where the $E[\cdot]$ operator is used in Eq. (5) to account for the random inputs. The SNR in Eq. (1) is appropriate only when x(t) is deterministic. In our pattern recognition applications this is not the case since the x(t) inputs are randomly distorted functions. The optimal statistical filter is defined as the filter h(t) which maximizes this output SNR in Eq. (5). If we define $f(t) = h(-t)$, we can rewrite

$$\text{SNR} = \frac{E[\int x(t)f(t)dt]^2}{E[\int n(t)f(t)dt]^2} = \frac{\int\int f(t)f(s)K_{xx}(t,s)\,dtds}{\int\int f(t)f(s)K_{nn}(t,s)\,dtds} , \tag{6}$$

where $K_{xx}(t,s)$ and $K_{nn}(t,s)$ are the ensemble auto-correlation functions of random processes {x(t)} and {n(t)}, respectively. For the case when {n(t)} is white noise with uniform power spectral density N, the auto-correlation function $K_{nn}(t,s)$ can be expressed as[15]

$$K_{nn}(t,s) = N\delta(t - s) , \tag{7}$$

where $\delta(t)$ denotes a Dirac delta function. The SNR in Eq. (6) can then be written as

$$\text{SNR} = \frac{\int\int f(t)f(s)K_{xx}(t,s)\,dtds}{N\int f^2(t)\,dt} . \tag{8}$$

The white noise assumption in Eq. (7) is commonly made. In cases for which it is not appropriate, pre-whitening operators can be applied as preprocessing functions prior to correlation.

3. OPTIMAL STATISTICAL FILTER

We now consider how to determine the optimal filter function f(t) that maximizes the statistical correlation output SNR expression

Eq. (8) developed in Sec. 2. We note that the integral in the denominator of Eq. (8) is the energy of the filter. It is thus acceptable to consider a normalized filter energy or

$$\int f^2(t)\,dt = 1 . \tag{9}$$

To optimize Eq. (8), we thus find the f(t) that maximizes the numerator in Eq. (8) subject to the condition in Eq. (9).

We achieve this by expanding the correlation function K_{xx} in terms of K-L basis functions ϕ_n, where $\{\phi_n(t)\}$ are the set of orthonormal eigenfunctions of the integral kernel $K_{xx}(t,s)$, as

$$K_{xx}(t,s) = \sum_n \lambda_n \phi_n(t)\phi_n(s) , \tag{10}$$

where the $\{\lambda_n\}$ are the eigenvalues corresponding to the $\{\phi_n\}$. Using inner product notation and the orthonormality of the $\{\phi_n\}$, we can restate the optimization problem as finding f that maximizes

$$R = \sum_n \lambda_n (f \cdot \phi_n)^2 , \tag{11}$$

subject to the condition

$$(f \cdot f) = 1 . \tag{12}$$

The filter function f can also be expanded in terms of the orthonormal set $\{\phi_n\}$ as

$$f = \sum_n \omega_n \phi_n . \tag{13}$$

where from Eq. (12) the coefficients ω_n in Eq. (13) must satisfy

$$\sum_n \omega_n^2 = 1 . \tag{14}$$

If we order the eigenvalues so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq \ldots$, we can easily show the quantity R has a maximum of λ_1 as below:

$$R = \sum_n \lambda_n (f \cdot \phi_n)^2 = \sum_n \lambda_n \omega_n^2$$
$$\leq \lambda_1 \sum_n \omega_n^2 = \lambda_1 . \tag{15}$$

This maximum R value is achieved only if we choose the filter function f to be

$$f(t) = \phi_1(t) , \tag{16}$$

since when Eq. (16) is substituted into Eq. (11) we find $R = \lambda_1$. Thus Eq. (16) describes the optimal filter, and we see that it is the principal-component or dominant eigenvector of the integral kernel $K_{xx}(t,s)$. This means that if the set of functions $\{\phi_n(t)\}$ are the eigenfunctions of the kernel $K_{xx}(t,s)$ with the corresponding set of eigenvalues $\{\lambda_n\}$, then the optimal filter is the eigenfunction $\phi_1(t)$ with the largest eigenvalue.

It can be shown[17] that the correlation function $K_{xx}(t,s)$ can be represented with arbitrary accuracy by a finite number of eigenfunctions

$$\{\phi_n(t)\}_{n=1,M}$$

as assumed in Eq. (10). From Eq. (8), Eq. (16), and the orthonormality of the ϕ_n, we then find that the maximum obtainable SNR is λ_1/N.

4. EXPERIMENTAL RESULTS

In Sec. 3, we showed that the optimal filter for a statistical correlator to detect stochastic signals (randomly distorted reference functions) was the principal component of the stochastic process $\{x(t)\}$ that characterizes these signals or equivalently the dominant eigenvector or eigenimage of the ensemble correlation function $K_{xx}(t,s)$ in Eq. (4). In this section, we provide digital simulation data comparing the performance of our statistical correlator to other types of correlators for the recognition of distorted versions of a reference image.

As our image data set, we used four multispectral images of an area south of Fresno, California (Fig. 3). We denote these images by P_1, P_2, P_3, and P_4. They were taken from the multispectral scanner on the Landsat satellite in the spectral bands 0.8 to 1.1 μm, 0.7 to 0.8 μm, 0.6 to 0.7 μm, and 0.5 to 0.6 μm, respectively. Each digital image was of size 128×128 pixels with 256 gray levels per pixel. The principal-component image was computed and is denoted by P_5.

For comparison, we also computed a synthetic filter image from P_1 to P_4 using the technique described by Hester and Casasent.[18] We denote this synthetic reference by P_6. The technique used to produce P_6 involved diagonalizing the cross-correlation matrix of the imagery in Fig. 3 using the Gram-Schmidt procedure. A filter P_6 was then produced from a linear combination of the basis functions (found from the matrix diagonalization) subject to the constraint that the correlation of all inputs with the filter yields a constant. We thus refer to P_6 as an equal-correlation peak (ECP) synthetic reference function.

Prior to any operations, the average intensity level of each image was computed and subtracted from the data, and the intensity variances of all four images were normalized to a fixed constant. Such image preprocessing eliminated the effects of image bias and modulation level variations from our data. The 4×4 correlation matrix for these preprocessed images was obtained, and its eigenvalues were computed. The four eigenvalues λ_1, λ_2, λ_3, and λ_4 were found to be 2218.9, 1594.1, 195.2, and 87.8, respectively. The principal component image P_5 corresponding to the largest eigenvalue $\lambda_1 = 2218.9$ is shown in Fig. 4. Use of this image as the filter should yield the optimum SNR when correlated with any of the four images. By comparison, correlations with image P_6 should yield equal-correlation peaks for all four input images.

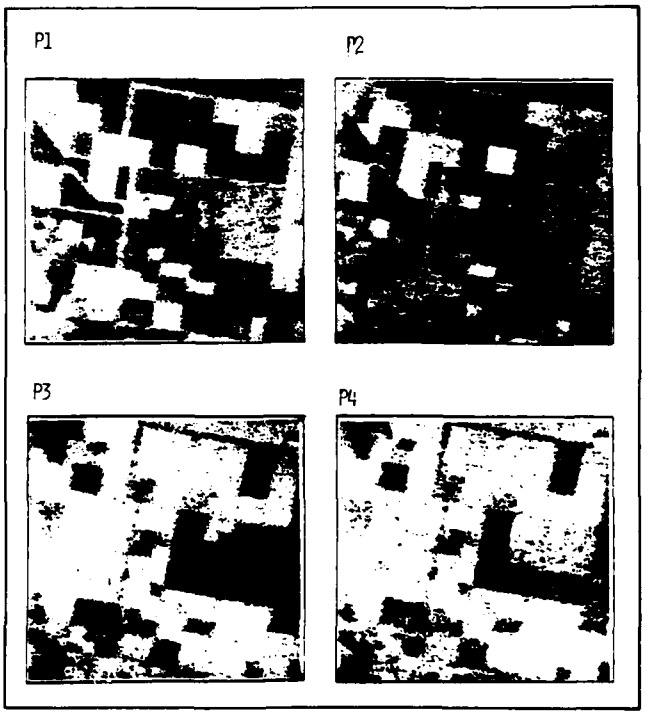


Fig. 3. Multispectral image data base used.

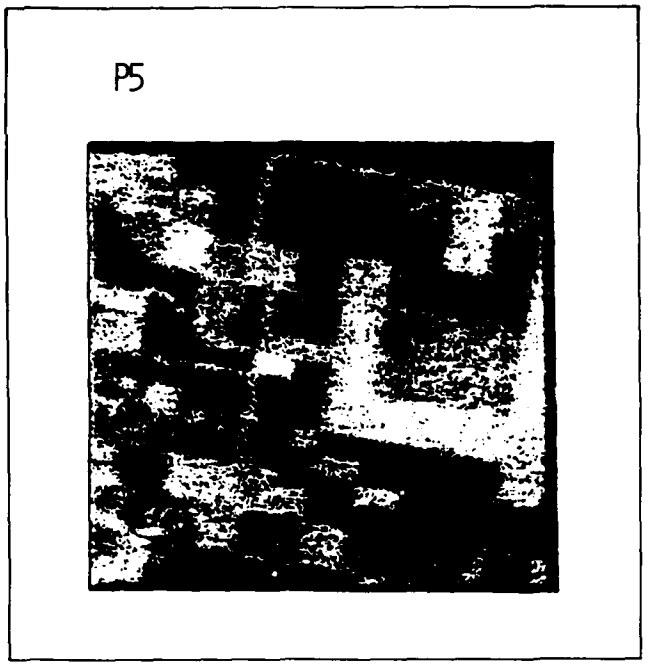


Fig. 4. Principal-component image P_5 obtained from the data set in Fig. 3.

Our simulation results are summarized in Tables I-IV. In Table I, we list the peak values of the output correlation for all 24 possible correlations of the original images P_1 to P_4, with P_1 to P_4 and the two synthetic filters P_5 and P_6. We have normalized the correlation peak values in each column in Table I by the auto-correlation peak value in each column. The correlation results for P_1 to P_4 correspond to conventional correlations. The results with P_5 and P_6 show how more advanced filters perform. From Table I, we note that images P_1 and P_2 correlate rather well with each other but quite poorly with P_3 and P_4 and vice versa. Conversely, filter P_5 correlates well with each input, with correlation values from 0.31 to 0.67. Similarly, filter P_6 correlates equally well with all four original inputs. We also note that our statistical filter P_5 performs better than the ECP filter P_6 for three of the four input images.

In Table II, we list the correlation plane peak-to-sidelobe levels for the 24 correlations. The peak-to-sidelobe level can be shown[8] to differ by a constant 3 dB from the SNR in Eq. (5). We computed it rather than Eq. (5) since it provided better sampling statistics. The results are similar to those obtained in Table I. Both agree with what one might expect from a visual inspection of the images in Fig. 3.

In Table III, we list the displacement of the correlation peak in pixels from its correct (0, 0) location, and in Table IV, we list the cross-sectional area of the correlation peak in pixels between its 3 dB points. From these data, we see that the conventional filters P_1 to P_4 perform well for autocorrelations, but have large errors in the location of the peak and have quite wide peaks for selected cross-correlations. This occurs because any two images from the set P_1 to P_4 are quite dissimilar even if they represent the same scene. Thus, even at correct registration, images such as P_1 and P_3 will have little in common, and thus a low correlation will result (as in Table I). Such conventional filters thus have a large probability of false alarm at incorrect registration (as Table III shows). The very wide 3 dB correlation peaks in Table IV indicate that there is no clear peak in the correlation output. Conversely, the statistical and synthetic filters P_5 and P_6 yield perfect registration results (Table III) and 3 dB correlation peak areas very close to those obtained from auto-correlations (Table IV).

Our emphasis in these initial experiments was to verify the appropriateness of our proposed statistical correlator. The experimental results obtained indicate the superiority of this new correlator to the conventional one and to the ECP correlator. Statistical data on the noise and randomness of the experimental data base used were not obtained, since accurate statistical estimates cannot be determined from only four images. Some geometrical distortions exist in this data base, and our model and correlator should be appropriate for such more general distortion cases. However, experimental verification of this is not included at this time, but will be the subject of a future paper in which the discrimination ability of our statistical correlator will be addressed.

TABLE I. Correlation Matrix for the Test Images in Fig. 3.

	P1	P2	P3	P4
P1	1.0000	0.6175	0.0622	0.0609
P2	0.6175	1.0000	0.0023	0.0026
P3	0.0622	0.0023	1.0000	0.8357
P4	0.0609	0.0026	0.8357	1.0000
P5	0.5211	0.3087	0.6688	0.6683
P6	0.3650	0.3650	0.3650	0.3650

TABLE II. Correlation Peak-to-Sidelobe Ratio for the Test Images in Fig. 3.

	P1	P2	P3	P4
P1	63.163	53.286	8.2671	9.9870
P2	53.287	93.628	11.060	12.533
P3	8.2671	11.060	44.912	51.848
P4	9.9870	12.533	51.848	76.739
P5	46.525	36.366	45.034	56.355
P6	35.310	43.845	28.418	39.949

TABLE III. Shift in the Location of the Correlation Peak from Its Correct (0,0) Location for the 24 Correlations in Tables I and II.

	P1	P2	P3	P4
P1	(0,0)	(0,−1)	(0,−1)	(−5,1)
P2	(0,1)	(0,0)	(*,0)	(*,−1)
P3	(0,1)	(*,0)	(0,0)	(0,0)
P4	(5,−1)	(*,1)	(0,0)	(0,0)
P5	(0,0)	(0,0)	(0,0)	(0,0)
P6	(0,0)	(0,0)	(0,0)	(0,0)

*Denotes a shift of more than ten pixels.

TABLE IV. Area of the Correlation Peak at the 3 dB Points for the 24 Correlations in Tables I and II.

	P1	P2	P3	P4
P1	29	48	129	92
P2	47	25	275	184
P3	140	275	79	66
P4	92	184	65	41
P5	49	40	75	55
P6	26	46	78	62

5. SUMMARY

In this paper, we have described one way to select and synthesize a filter to recognize distorted versions of a reference function from a

given data set. We considered the case of white noise and as our performance measure used a new ensemble SNR definition. For these cases, we found the principal-component or dominant eigenimage to be the optimal filter. Our experimental data verified the expected performance of this statistical filter. Both the statistical and synthetic filter were shown to perform much better than conventional matched spatial filters. Our new statistical principal-component filter was seen to provide better correlation peak and correlation SNR values than the synthetic filter in 75% of the cases and quite close performance in the remaining cases.

In future work, we will consider fast and efficient ways to compute the principal-component image as well as revised algorithms by which to improve the discrimination performance of such filters.

Image quality is of concern in pattern recognition systems as well as in systems and applications in which the output is an image. In pattern recognition problems, the SNR of the output correlation is the quality measure most commonly used to describe the imagery as well as the performance of the pattern recognition system. In this work, we have described a method to improve the performance of a correlator by using a statistical filter obtained from the eigenimage of the data base and have shown that such an image contains the image quality features of the data base needed for pattern recognition.

6. ACKNOWLEDGMENTS

The authors thank the Air Force Office of Scientific Research (GRANT AFOSR 79-0091) for support of this research, and Toshiba Corporation of Kawasaki for support of H. Murakami during the duration of this research.

7. REFERENCES

1. P. Cheatham, Ed., Proc. SPIE 310 (1981) in press.
2. P. Cheatham, Ed., Opt. Eng. 21(1) (1981).
3. K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, New York (1972).
4. K. Fukunaga and W. L. G. Koontz, IEEE Trans. Comput. C-19, 311(1970).
5. E. L. Hall, *Computer Image Processing and Recognition*, Academic Press, New York (1979).
6. A. Habibi and P. A. Wintz, IEEE Trans. Comm. Technol. COM-19, 948(1971).
7. H. Murakami, "Matched Filter Statistical Correlator Advances," M.S. Thesis, Carnegie-Mellon University (1981).
8. B. V. K. Vijaya Kumar and D. Casasent, J. Opt. Soc. of Amer. 70, 103(1980).
9. A. D. Whalen, *Detection of Signals in Noise*, Academic Press, New York (1971).
10. A. Van der Lugt, IEEE Trans. Inform. Theory IT-10, 139(1964).
11. R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley, Massachusetts (1977).
12. D. Casasent and A. Furman, Appl. Opt. 16, 1652(1977).
13. D. Casasent and D. Psaltis, IEEE Proc. 65, 77(1977).
14. H. Mostafavi and F. W. Smith, IEEE Trans. Aerosp. Electron. Syst., AES-14, 487(1978).
15. A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York (1965).
16. W. K. Pratt, *Digital Image Processing*, Sec. 19.3, J. Wiley and Sons, New York, (1978).
17. C. W. Helstrom, *Statistical Theory of Signal Detection*, Pergamon Press, New York (1968).
18. C. F. Hester and D. Casasent, Appl. Opt. 19, 1758(1980).

9. A MODIFIED HYPERPLANE METHOD FOR NULL SYNTHESIS IN AN ARRAY PATTERN

A Modified Hyperplane Method for Null Synthesis in an Array Pattern

B. V. K. VIJAYA KUMAR

Abstract—**It has recently been shown by Prasad that the problem of synthesizing nulls in an array pattern admits a geometrical formulation for which the alternative orthogonal projection (AOP) algorithm (previously used in image restoration applications) is a useful solution. We apply a modification to the original AOP suggested by Ramakrishnam *et al.* for image restoration purposes and show that it results in a computational savings in array pattern synthesis problems. A numerical example is presented to illustrate the computational advantages of this modification.**

I. INTRODUCTION

In many radar applications, it is desired that the receiving array of antennas exhibit a high gain in the direction of the "signal" and a very low gain in the directions of unwanted "interference" sources. Such a response from an arrray is usually obtained by weighting [1], [2] the signals received at each antenna by a different, but appropriate complex constant. Determining these complex weights to obtain a desirable array pattern is known as array pattern synthesis, and much research [3], [4] has been devoted to this topic. Polynomial approaches, which are computationally burdensome, have mostly been used for such pattern synthesis.

Consider an array of N identical omnidirectional receiving antennas located in three-dimensional space at vector positions $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$. Let $\mathbf{u}_s$ represent a unit vector in the "look" direction of the signal and let $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_M$ denote unit vectors in the directions of the M "interference" sources. The objective of the array of antennas is to maximize the signal arriving along $\mathbf{u}_s$ while minimizing the interference along $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_M$. Antenna pattern synthesis involves determining the complex array weights $W_1, W_2, W_3, \cdots, W_N$ so that the following constraints are satisfied.

$$\mathbf{C}^T\mathbf{W} = 1 \qquad \text{(look direction)} \tag{1a}$$

and

$$\mathbf{S}_i{}^T\mathbf{W} = 0 \qquad \text{(noise directions)}, \quad i = 1, 2, \cdots, M, \tag{1b}$$

where $[\cdot]^T$ denotes the *complex conjugate transpose*, $\mathbf{W}$ is the column vector of weights $W_1, W_2, \cdots, W_N$, and the vectors $\mathbf{C}^T$ and $\mathbf{S}_i{}^T$ defined as in [4] represent the received signals due to the "look" direction signal and due to the ith noise source, respectively.

For any general geometry of the array, the constraints in (1a) and (1b) can be solved to obtain the weight vector $\mathbf{W}$. This set of equations has no solutions if $\mathbf{M} \geqslant N$. Thus, we assume throughout this communication that the number of interference sources M is less than the number of antenna elements N. The $(M + 1)$ equations in (1) can be combined

Manuscript received March 27, 1981; revised June 17, 1981. This work was supported by the Air Force Office of Scientific Research under Grant AFOSR-79-0091.

The author is with the Department of Electrical Engineering, Carnegie-Mellon University, Schenley Park, Pittsburgh, PA 15213.

0018-926X/82/0500-0512$00.75

as

$$\mathbf{AW} = \mathbf{U}', \tag{2}$$

where $\mathbf{U}' = [1\ 0\ 0\ 0\ \cdots\ 0]^T$ is a $(M+1)$ element vector and $\mathbf{A}$ is a $(M+1) \times N$ matrix with the vectors $\mathbf{C}$ and $\mathbf{S}_i$ as its rows as below.

$$\mathbf{A} = [\mathbf{C}, \mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_M]^T. \tag{3}$$

When matrix $\mathbf{A}$ is not a square matrix (i.e., when $M < (N-1)$), pseudoinverses [5] should be used for solving (2) as below.

$$\mathbf{W} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{U}'. \tag{4}$$

Even when pseudo-inverses can be used, one is faced with the prospect of inverting an $N \times N$ matrix. This causes computational problems for large values of N. This computation aspect will be critical in scenarios where noise sources move continuously and thus new "adaptive" weights must be computed at a rapid rate. Recursive methods are needed to speed this weight determination.

Recently, Prasad [4] has shown that the problem in (2) admits a geometrical formulation and has used the alternating orthogonal projection (AOP) method suggested by Youla [6] for image restoration purposes. This approach enables the use of many other interesting image restoration algorithms in solving the null synthesis problem. We found the modified hyperplane method (MHP) suggested by Ramakrishnam *et al.* [7] to result in better computational efficiency in solving (2). MHP is based on modifying the $(M+1)$ equations in (2) so that the adjacent equations represent pairwise orthogonal hyperplanes in an N-dimensional space. This modification results in faster convergence of conventional projection methods to the solution vector $\mathbf{W}$.

We outline both the AOP and MHP methods in Section II and show in Section III that the MHP method is computationally more efficient. This is justified with the help of a numerical example in Section IV.

II. PROJECTION METHODS

A. Alternate Orthogonal Projections (AOP)

The AOP derives its name from the fact that we project our estimates orthogonally onto subspaces perpendicular to $\mathbf{S}_m$ and $\mathbf{C}$ in an alternating fashion. The recursion for the weight vector $\mathbf{W}_{K+1}$ can be summarized by

$$\mathbf{W}_{K+1} = (\mathbf{Q}\mathbf{P}_M\mathbf{P}_{M-1}\cdots\mathbf{P}_2\mathbf{P}_1)\mathbf{W}_K + \mathbf{g}, \quad K = 1, 2, 3, \cdots \tag{5}$$

where

$$\mathbf{g} = \mathbf{C}/(\mathbf{C}^T\mathbf{C}) \tag{6}$$

and

$$\mathbf{W}_1 = \mathbf{g}. \tag{7}$$

The projection operators $\mathbf{Q}$ and $\mathbf{P}_m$ given by

$$\mathbf{Q} = (\mathbf{I} - \mathbf{C}\mathbf{C}^T/\mathbf{C}^T\mathbf{C}), \tag{8}$$

and

$$\mathbf{P}_m = (\mathbf{I} - \mathbf{S}_m\mathbf{S}_m{}^T/\mathbf{S}_m{}^T\mathbf{S}_m), \tag{9}$$

project the weight vector to the subspaces orthogonal to $\mathbf{C}$ and $\mathbf{S}_m$, respectively. In the above equations, $\mathbf{I}$ represents the identity operator. Further details regarding this recursive method and its convergence behavior can be obtained from Prasad's paper [4].

B. Modified Hyperplane Method (MHP)

For the general case of N weights and $(M+1)$ constraints as in (1), we can modify each of the $(M+1)$ equations so that each one is orthogonal to all others. This is equivalent to diagonalizing the $\mathbf{A}^T\mathbf{A}$ matrix in (4). As a result, $\mathbf{W}$ can be easily obtained from (4) by simple matrix multiplication. The major obstacle in such a procedure is our ability to rapidly rearrange the equations in the desired manner. One can show that rearranging the equations to make them mutually orthogonal is computationally as burdensome as evaluating $\mathbf{W}$ from (4) directly. A practical, but suboptimal, procedure suggested by Ramakrishnam *et al.* [7] involves modifying these so that each equation is orthogonal to its *adjacent* one. As we will see in Section III, such an approach considerably reduces the computational complexity of this method. By the MHP method, we refer to this suboptimal procedure. This MHP procedure consists of the following two steps.

1) *Equation Rearrangement:* The Gram–Schmidt procedure [8] is used to produce the rearranged (primed) equations from the original ones. The equations defining the operations are

$$\mathbf{S}_M{}' = \mathbf{S}_M, \tag{10a}$$

$$\mathbf{S}_m{}' = \mathbf{S}_m - \left\{\frac{\mathbf{S}_m{}^T\mathbf{S}_{m+1}{}'}{\mathbf{S}_{m+1}{}'^T\mathbf{S}_{m+1}{}'}\right\}\mathbf{S}_{m+1}{}',$$

$$m = (M-1), (M-2), \cdots, 3, 2, 1, \tag{10b}$$

and

$$\mathbf{C}' = \mathbf{C} - \left\{\frac{\mathbf{C}^T\mathbf{S}_1{}'}{\mathbf{S}_1{}'^T\mathbf{S}_1{}'}\right\}\mathbf{S}_1{}'. \tag{10c}$$

When we modify the left side of (1) as above, we also need to change the right side of (1) in an exactly identical manner to obtain the correct solution. The right side of (1) in terms of the primed equations can be easily shown to still equal $\mathbf{U}' = [1\ 0\ 0\ \cdots\ 0]$. Thus the new equation to be solved is

$$\mathbf{A}'\mathbf{W} = \mathbf{U}', \tag{11}$$

where the prime denotes the result of the equation rearrangement.

2) *Projection Algorithm:* Projection methods [9] can be used on the modified equations to solve for $\mathbf{W}$ in (11). The projection method begins with an initial guess $\mathbf{W}_0$. This point $\mathbf{W}_0$ in the N-dimensional space is then orthogonally projected onto the first hyperplane ($\mathbf{C}'$ in this case). Let $\mathbf{W}_1$ represent the result of this projection. This is subsequently projected orthogonally onto the second hyperplane ($\mathbf{S}_1{}'$ here) to obtain $\mathbf{W}_2$. This orthogonal projection is carried out on successive hyperplanes unitl $\mathbf{W}_{M+1}$, the projection onto $\mathbf{S}_M{}'$, is obtained.

In our notation, obtaining $\mathbf{W}_{M+1}$ from $\mathbf{W}_0$ represents *one complete iteration*. $\mathbf{W}_{M+1}$ is then orthogonally projected onto the first hyperplane ($\mathbf{C}'$) and a complete second iteration begins. This algorithm can be described by

$$\mathbf{W}_1 = \mathbf{W}_0 - \left[\frac{\mathbf{W}_0{}^T\mathbf{C}' - 1}{\mathbf{C}'^T\mathbf{C}'}\right]\mathbf{C}', \tag{12a}$$

and

$$\mathbf{W}_{m+1} = \mathbf{W}_m - \left[\frac{\mathbf{W}_m{}^T\mathbf{S}_m'}{\mathbf{S}_m'^T\mathbf{S}_m'}\right]\mathbf{S}_m',$$

$$m = 1, 2, 3, \cdots, M. \tag{12b}$$

The additional unity term in (12a) can be understood from the unity in the right side of (11). In the next section, we compare these two methods from their computational loading requirements.

III. COMPUTATIONAL CONSIDERATIONS

As can be seen from (5)–(10), both the AOP and MHP methods require initial computation before the recursions can be started. We will now compare the two methods based on this computational burden as well as on the number of computations required per iteration.

A. AOP Method

Given the direction vectors $\mathbf{C}$ *and* $\mathbf{S}_m$, *we determine* the projection operators $\mathbf{Q}$ and $\mathbf{P}_m$ as in (8) and (9). For an N element array, this computation requires approximately $(M + 1)N^2$ complex operations. These $(M + 1)$ projection operators are then combined to produce a single projection operator T, where

$$\mathbf{T} = \mathbf{Q}\mathbf{P}_M\mathbf{P}_{M-1}\cdots\mathbf{P}_2\mathbf{P}_1. \tag{13}$$

The computation of $\mathbf{T}$ involves M matrix multiplications each involving N^3 complex operations. For large values of N, MN^3 complex operations are required for determining the operator $\mathbf{T}$.

As noted in (5), $\mathbf{W}_{K+1}$ can be obtained from $\mathbf{W}_K$ by a matrix/vector multiplication and vector addition. This involves N^2 complex operations. Thus, the AOP requires MN^3 complex operations for initial set-up and N^2 complex operations for every iteration thereafter.

B. MHP Method

The initial equation rearrangement for the MHP method is outlined in (10). From (10), we see that $\mathbf{S}_m'$ can be obtained from $\mathbf{S}_m$ and $\mathbf{S}_{m+1}'$ using approximately $2N$ complex operations. Then the hyperplane rearrangement requires a total of $2MN$ complex operations. This is the initial computational load for the MHP algorithm equations.

The projection algorithm of (12) is used on the rearranged equations to obtain the solution. $\mathbf{W}_{m+1}$ is obtained from $\mathbf{W}_m$ by determining the dot products $\mathbf{W}_m{}^T\mathbf{S}_m'$ and $\mathbf{S}_m'^T\mathbf{S}_m'$ and then using the ratio of these products to multiply $\mathbf{S}_m'$ by that scalar. The dot products involve N complex operations each and (ignoring the scaler multiplication) we need $2N$ complex operations to obtain $\mathbf{W}_{m+1}{}^T$ from $\mathbf{W}_m$. Thus, one complete iteration of MHP (i.e., obtaining $\mathbf{W}_{M+1}$ from $\mathbf{W}_0$) requires a total of $2MN$ complex operations. Thus, MHP requires $2MN$ initial complex operations and $2MN$ complex operations per iterations.

Comparing the two schemes, we note that MHP requires only $2MN$ operations while AOP needs MN^3 initially. For large values of N, this represents a substantial reduction. For every iteration after the initial set-up, the AOP requires N^2 operations while the MHP needs only $2MN$ operations. Usually $N > 2M$ and as a result the MHP technique is computationally more efficient than the AOP method.

An equally important consideration in evaluating the two methods is the number of iterations each requires to approach the true solution. Since both methods are based on projections, it is expected that they will perform similarly. However, because of the rearrangement of the equations, MHP will converge faster. No rigorous comparisons of this aspect have been considered because the convergence rate is very dependent on the $\mathbf{A}$ matrix in (2). In the next section, we consider a numerical example to compare the convergence behavior of the two techniques.

IV. NUMERICAL RESULTS

We consider the special case of a linear array of $N = 17$ elements, uniformly spaced at $\lambda/2$ intervals, where λ is the wavelength of the radiation. The signal "look" direction is along boresight (i.e., $\theta = 0°$) and $M = 5$ interference sources are assumed at angles of $-15°$, $-20°$, $5°$, $45°$, and $75°$ (from boresight), respectively. The array problem is then formulated as solving for 17 unknowns subject to six constraints as in (1).

For both schemes, the initial estimate of the weight vector $\mathbf{W}_0$ has unity for elements because $\mathbf{C}$ has unity for all elements. Fig. 1 shows the array pattern $E(\theta)$ obtained using the AOP algorithm after just one iteration. No nulls can be easily identified in this response pattern. Fig. 2 shows $E(\theta)$ obtained using MHP after just one iteration. MHP is seen to yield very sharp nulls at the two noise source locations, $45°$ (-175 dB) and $75°$ (-172 dB), after just one iteration. The null depths beyond 100 dB are not shown in our figures because they are practically not significant even though they are computationally possible. A detailed comparison of Figs. 1 and 2 shows that MHP also outperforms AOP at three other interference locations. Fig. 3 shows the array pattern obtained by AOP after five iterations. This response exhibits good nulls (with depths of 113 dB, 94 dB, 94 dB, 108 dB, and 98 dB). But, after five iterations, MHP (see Fig. 4) yields even better null depths (115 dB, 133 dB, 125 dB, 181 dB, and 184 dB) at the five noise source locations. After ten iterations, both methods are found to perform equally well. This supports Prasad's claim [4] that AOP requires approximately eight to ten iterations for convergence.

The null depths obtained with both methods at each of the five noise source locations after each iteration are listed in Table I. This table also contains the average null depths as well as the worst null depths obtained by both methods after each iteration. From Table I, we see that an average of 99 dB is obtained with MHP in just one iteration, whereas AOP requires four iterations to yield a 100 dB average null depth. Inspection of this column also indicates that MHP converges faster than AOP by at least a factor of two. The worst-case null depths obtained by the MHP are also seen to be consistently better than those of the AOP by at least 10 dB. More quantitative comparisons are not valid because the convergence behavior is very data dependent. From this limited ex-

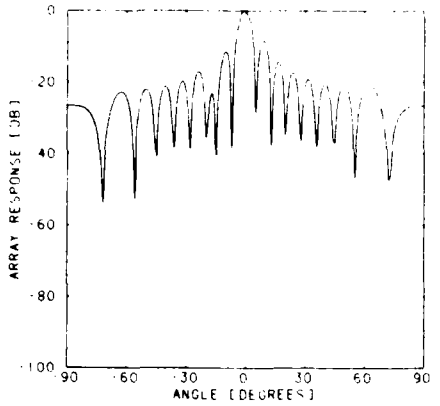


Fig. 1. Array response after one iteration of the AOP Method.

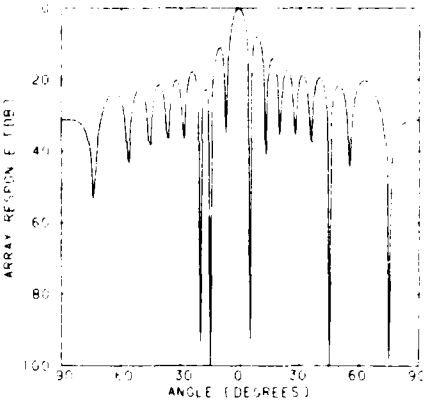


Fig. 3. Array response after five iterations of the AOP Method.

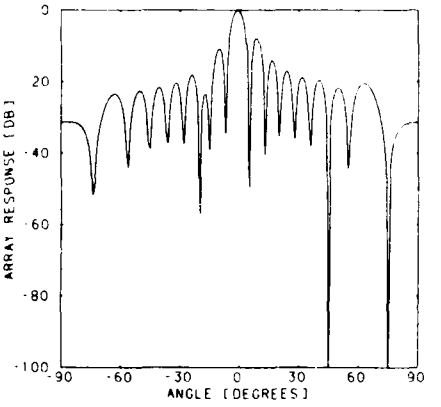


Fig. 2. Array response after one iteration of the MHP Method.

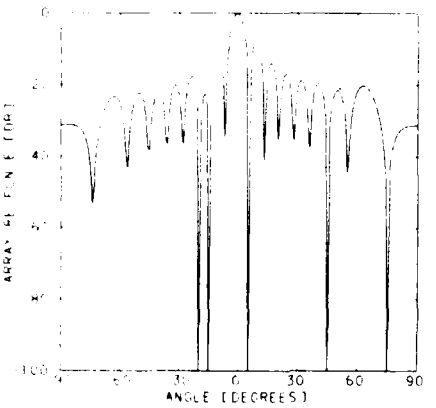


Fig. 4. Array response after five iterations of the MHP Method.

TABLE I
NULL DEPTHS (dB) WITH AOP AND MPH METHODS

Iteration Number	Null Depths AOP Method −15°	−20°	5°	45°	75°	Average	Worst Null	MPH Method −15°	−20°	5°	45°	75°	Average	Worst Null
1	41	36	29	38	34	35.6	29	39	57	50	175	172	98.6	39
2	59	49	51	55	63	55.4	49	59	76	69	180	174	111.6	59
3	80	62	117	77	76	82.4	62	78	95	87	174	173	121.4	78
4	141	78	83	114	85	100.2	78	96	114	106	170	175	144.3	96
5	113	94	94	108	98	101.4	94	115	133	125	181	184	147.6	115

periment and the analysis in Section III, we can say that MHP converges faster than AOP in almost all cases.

V. CONCLUSION

Prasad's [4] geometrical formulations allow us to use the modified hyperplane projection technique [7] in the generalized pattern synthesis problem. This MHP algorithm has been shown to be computationally more efficient than the alternating orthogonal projection method used earlier [4]. Such computational time savings are important in an environment where the interference locations are rapidly changing and in which the array weights must "adapt" to them at high rates.

ACKNOWLEDGMENT

The author wishes to acknowledge Professor David Casasent for his critical reading of this manuscript and for providing valuable comments on it.

REFERENCES

[1] D. E. N. Davies, "Independent angular steering of each zero of the directional pattern for a linear array," *IEEE Trans. Antennas Propagat.*, vol. AP-15, pp. 296–298, 1967.

[2] C. Drane and J. McIlvena, "Gain maximization and controlled null placement simultaneously achieved in aerial array systems," *Radio Electron. Eng.*, vol. 39, pp. 49–57, 1970.

[3] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, 1972.

[4] S. Prasad, "Generalized array pattern synthesis by the method of alternating orthogonal projections," *IEEE Trans. Antennas Propagat.*, vol. 28, pp. 328–332, 1980.
[5] C. R. Rao, *Linear Statistical Inference and its Applications*. New York: Wiley, 1965, p. 24.
[6] D. C. Youla, "Generalized image restoration by the method of alternating orthogonal projections," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 694–702, 1978.
[7] R. S. Ramakrishnam, S. K. Mullick, R. K. S. Rathore, and R. Subramanian, "Orthogonalization, Bernstein polynomials and image restoration," *Appl. Opt.*, vol. 18, pp. 464–468, 1979.
[8] T. S. Huang, D. A. Barker, and S. P. Berger, "Iterative image restoration," *Appl. Opt.*, vol. 14, pp. 1165–1168, 1975.

10. NEW RESEARCH IN HOLOGRAPHIC PATTERN RECOGNITION

353-02

NEW RESEARCH IN HOLOGRAPHIC PATTERN RECOGNITION

David Casasent

Carnegie-Mellon University
Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

ABSTRACT

The major use of holography in optical pattern recognition systems has been in matched spatial filter correlators. We discuss new uses of holographic techniques in optical pattern recognition applications. These include: optical pattern recognition architectures of reduced size, weight and volume; optical pattern recognition architectures with reduced component tolerances; new nonlinear local optical preprocessing architectures; and optical pattern recognition systems that operate when geometrical and other differences are present in the input image.

1. INTRODUCTION

This SPIE Proceedings Volume addresses holography and its applications. One of the attractive uses of holography is optical pattern recognition (OPR). In this paper, we review recent progress in new OPR algorithms, architectures and applications employing holography. Such research has been supported at Carnegie-Mellon University for several years by a grant from the Air Force Office of Scientific Research. In this paper, we review some of the aspects of this research. To remain within the context of this issue, we restrict the research reported upon to work involving holography.

When one considers OPR and holography, the optical matched spatial filter (MSF) correlator [1] and the joint transform correlator [2] immediately come to mind. We review these basic OPR architectures in Section 2. A disadvantage of optical processors frequently noted is their large size. In Section 3, we discuss one technique by which holographic optical elements (HOEs) and laser diodes (LDs) can be used to fabricate OPR systems of reduced size, weight and volume [3]. These architectures also demonstrate new system concepts that can be realized with holographic optical components as well as system architectures with reduced positional tolerance requirements of the elements [3]. MSF correlators are well-known to be sensitive to geometrical distortions in the input image. In Section 4, we review weighted MSFs [4], Mellin transforms (MTs) [5] and hybrid optical/digital processor techniques [6]. These techniques can reduce the sensitivity of an MSF correlator to geometrical distortions, but they do not fully overcome intra-class pattern recognition problems (recognition of an object independent of geometrical distortions between the input and reference object views). One approach to OPR has been to apply digital preprocessing to enhance the input image to an optical processor. As input image resolution and frame rates increase, such preprocessing operators are becoming very computationally burdensome. In Section 5, we discuss new nonlinear local optical preprocessing operators that operate in parallel on all image pixels [7]. We then conclude in Section 6, with a review of research on how off-line operations on a training set of images can produce a synthetic discriminant function (SDF) MSF capable of inter-class discrimination while retaining intra-class object recognition [8-11]. Because of space limitations, we restrict attention in this paper to AFOSR supported holographic pattern recognition research at CMU. Other holographic OPR research is described elsewhere [12-16].

2. CONVENTIONAL OPTICAL PATTERN RECOGNITION CORRELATORS

The basic OPR correlator architectures are well-known and are thus only briefly reviewed here. In the classic frequency plane correlator (Figure 1), a holographic MSF (formed by the coherent interference of the Fourier transform (FT) of the reference object and a plane wave reference beam) H^* is formed at P_2 with the desired reference object $h(x,y)$ placed at P_1. This holographic MSF differs from conventional holograms and specifically FT holograms in the beam-balance ratio used (i.e. the purpose of a hologram is to produce an aesthetically pleasing reconstructed image, whereas an MSF is intended for pattern recognition and thus conventionally low spatial frequencies are saturated during MSF synthesis). With H^* formed as noted above, the plane wave reference beam is no longer used and an on-line input image $f(x,y)$ (possibly containing the reference object $h(x,y)$) is placed at P_1. The light distribution incident on P_2 is $F(u,v)$(the FT of the input data) and the light distribution leaving P_2 is FH^*. At P_3, lens L_2 forms the FT of the light distribution leaving P_2 and at P_3 we find

$$u_3 = [FH^*] = f \circledast h, \qquad (1)$$

or the correlation of the input and reference patterns.

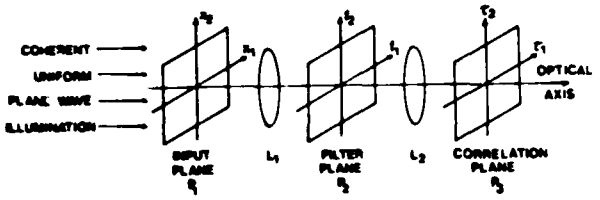


FIGURE 1 Schematic diagram of the conventional holographic matched spatial filter frequency plane correlator.

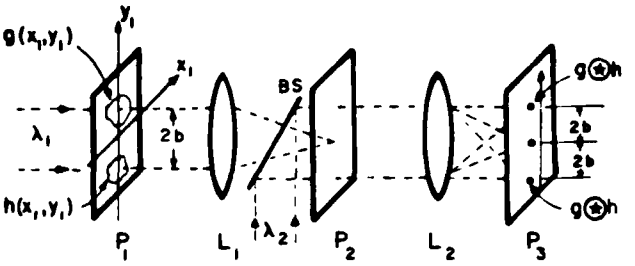


FIGURE 2 Schematic diagram of the conventional holographic joint transform correlator system.

In the joint transform correlator (Figure 2), the two objects to be correlated, g and h, are placed side-by-side in the input plane and their joint transform is formed at P_2 on an intensity sensitive material (film or a spatial light modulator). Plane P_2 is then illuminated (in transmission or reflection) and the FT of its transmittance is formed at P_3 where again we find the correlation of g and h. Any such optical correlation plane output contains peaks of light whose locations correspond to the positions of all occurrences of the reference object within the field-of-view of the input image. Thus, an optical correlator can recognize multiple occurrences of a reference object and determine the locations of each. An optical processor achieves these operations in parallel at high speed and thus such systems are quite attractive for many applications. However, they are now without shortcomings, as we now discuss.

3. HOLOGRAPHIC OPTICAL ELEMENTS AND LASER DIODES

It is possible to form multiple MSFs at P_2. We can form these multiple MSFs by superimposing MSFs with different carrier spatial frequencies or by placing the multiple MSFs at spatially separated locations in P_2. With such architectures, one can achieve multiple correlations in parallel. Researchers at Gruman Aerospace [17], Miradcom [18] and CMU are pursuing such approaches. The use of spatially-multiplexed multiple-MSFs is quite attractive. However, one must then form multiple replicas of the FT of the input image at the locations of each MSF. This can be accomplished by multiple point source HOEs.

Holographic optical elements (HOEs) also enable one to fabricate optical systems with new architectures and of reduced size. We detail one such system [3] in Figure 3. As a light source, we use a laser and we image it onto the FT plane P_2. We view lens L_1 as an FT lens and we note that the FT of the input data at P_1 is formed at P_2 even when P_1 is placed behind L_1. This scaling FT system allows one to adjust the size of the input and FT plane by adjusting the distance between P_1 and P_2 As the second new feature associated with Figure 3, we note that to form an MSF H^* at P_2, we place $h(x,y)$ at P_1. However, instead of a plane wave reference beam, we employ a converging reference wave. The resultant pattern recorded at P_2 is the MSF, plus an FT lens (a lensless MSF or the MSF plus an HOE FT lens). With this filter formed at P_2, if the input image at P_1 contains the reference object h, then a converging plane wave will leave P_2 and self focus at the correlation plane P_3. The location of this correlation peak will correspond to the location(s) of the reference object $h(x,y)$ within the entire field-of-view of the image $f(x,y)$ present at P_1.

We have assembled and demonstrated such a system [3] in the laboratory. This architecture is very attractive since with the MSF and the second FT lens of Figure 1 encoded on one plate, we have simultaneously satisfied all positioning requirements between these two planes. This greatly reduces the positioning requirements of the system as well as its over all length and size and number of components required. Although space permits only this one HOE architecture to be discussed, many other approaches are possible. The system depicted in Figure 3 is typical of such architectures and it demonstrates the flexibility that HOEs provide for new architectures such as the lensless MSF holographic pattern recognition system.

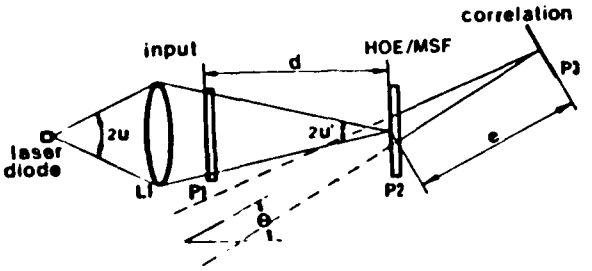


FIGURE 3 Schematic diagram of a lensless matched spatial filter holographic correlator using holographic optical elements.

4. WEIGHTED MATCHED SPATIAL FILTERS, MELLIN TRANSFORMS AND HYBRID OPTICAL/DIGITAL ARCHITECTURES

In [4], Casasent and Furman described how control of the beam balance ratio K during synthesis of an MSF could be used to control the spatial frequency region of the input pattern used during the correlation. We refer to such a filter as a weighted MSF. This concept is easily seen by considering the interference of two beams. When the two beams are of equal intensity, the modulation of the fringe pattern formed is a maximum. Because the FT of the input data is monotonically decreasing with spatial frequency or distance (in the transform plane), maximum modulation will occur at one spatial frequency region. During MSF synthesis, low spatial frequencies are saturated (because of the high intensities of the dc and low spatial frequencies in the FT of the reference signal). Similarly, high spatial frequencies have low modulation and emphasis (because of the low beam balance ratio). The intermediate spatial frequency region where the intensities of the FT of the reference object and the plane wave reference beam are approximately equal thus received maximum modulation. By adjusting this spatial frequency region, we can tune the MSF or correspondingly weight different spatial frequencies of the data. A similar phenomenon occurs during synthesis of the lensless MSF in Figure 3.

In [4], we showed that if this spatial frequency region were optimized for images at perfect registration, then the performance of the correlator would be optimum only for images with no distortions between the input and reference object. For such an MSF, a severe degradation in the system's performance would result if small scale and rotational differences were present. Similar remarks apply to small translational difference and positioning accuracies in the components of this processor. However, we also demonstrated [4], that emphasis of lower spatial frequencies would significantly improve the tolerance of the system to such error sources without appreciable degradation in its discrimination ability. However, this approach appears to be appropriate only for satisfying small variations in scale and rotational geometrical differences. For larger variations, advanced techniques (Section 6) are needed.

The Mellin transform (MT) and its variations such as the polar coordinate transformation represent techniques by which one can obtain invariance of the correlation peak intensity for selected image distortions. These approaches to distortion invariant pattern recognition are summarized in [5]. These techniques result in a space-variant OPR system in which the correlation of coordinate transformed versions of the input and reference object are produced. If the expected distortion can be described mathematically, then an appropriate coordinate transformation can be found that will yield a correlation peak intensity that is invariant to the given deformation. For scale invariance, a logarithmic coordinate transformation is the solution. The FT of a logarithmic coordinate transformed image is the MT of the original image. For rotational invariance, we find that a polar coordinate transformation is necessary. It is possible to combine several of these coordinate transformations with a cyclic operation of the system. However, one is restricted to producing a system with only two invariant distortion parameters on each cycle and to the application of these techniques for cases in which only one occurrence of the reference object is present in the input field-of-view. Thus, whereas this approach is very attractive for certain applications (e.g. for a top-down view of a single object), it is not as powerful as other techniques (Section 6) when multiple distortions or multiple objects are present.

The use of hybrid optical/digital processors is another approach by which the flexibility of an optical system can be increased. This concept [6] and many detailed versions of it can significantly improve the performance and repertoire of operations achievable on various optical processors.

5. PARALLEL OPTICAL NONLINEAR LOCAL PREPROCESSING OPERATORS

Many nonlinear local operators can and have been used to preprocess image data prior to feature extraction, pattern recognition and classification. These operators have thusfar been implemented digitally. A most attractive such nonlinear operator is the Sobel transform. This operator performs a nonlinear edge-enhancement and smoothing of the input image f(x,y) described by

$$g(x,y) = [\langle \partial f(x,y)/\partial x \rangle_y^2 + \langle \partial f(x,y)/\partial y \rangle_x^2]^{1/2}. \tag{2}$$

In recent research [7], we have devised a technique to fabricate holographic filters capable of performing the general class of nonlinear local operators. To describe the necessary filter, we first write (2) as

$$g = [X^2 + Y^2]^{1/2}, \tag{3}$$

where X and Y are best described as the appropriate 3 x 3 local masks

$$X = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \quad Y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}. \tag{4}$$

These local 3 x 3 operators in (4) are convolved with each image pixel and then nonlinearly combined as in (3). From (2) - (4), we see that the Sobel operator performs the local second differencing operations in the x and y directions (averaged over three rows or columns in y and x respectively). It thus achieves a nonlinear local edge-enhancement operation plus smoothing to reduce noise effects.

To achieve (2) or (3) optically, we rewrite (3) for input image pixel (j,k) as

$$g(i,j) = (X^2 + Y^2)^{1/2} = [|X + jY|^2]^{1/2} = |X + jY|. \tag{5}$$

Next, we recall [19] that the spatial derivative in the first approximation can be realized by the convolution of the input image with an MSF whose impulse response h is as below

$$\text{Output} = f * h = f(x,y) * [\delta(x+d,y) - \delta(x,y)] = \delta f/\delta x. \tag{6}$$

To realize (5) using a technique such as that in (6), we require an MSF with an impulse response h that satisfies

$$\begin{aligned} \text{Output} = g(x,y) &= X + jY = f(x,y) * h(x,y) \\ &= f(x,y) * \sum_{m,n} k_{m,n}\delta(x-md,y-nd) \\ &= f(x,y) * [(1+j)\delta(x-d,y-d) + 2\delta(x-d,y) \\ &\quad + (1-j)\delta(x-d,y+d) + 2j\delta(x,y-d) \\ &\quad - 2j\delta(x,y+d) - (1-j)\delta(x+d,y-d) \\ &\quad - 2\delta(x+d,y) + (1+j)\delta(x+d,y+d)]. \end{aligned} \tag{7}$$

In (6) and (7), the distance d is the spacing between pixels in the input image and the weights and locations of the eight delta functions in (7) were obtained from (4) and (5). We have realized the MSF described by (7) using both multiple exposure holographic techniques and use of a computer-generated hologram. We have also successfully demonstrated this technique for various 3 x 3 and 5 x 5 nonlinear local preprocessing operators [7]. An example of optical edge-enhancement preprocessing using these techniques is shown in Figure 4. The preprocessed output image shown was obtained in parallel using the nonlinear local operators described above. Extensions of this technique to larger window operators and to other preprocessing functions appears to be a quite attractive area for future research.

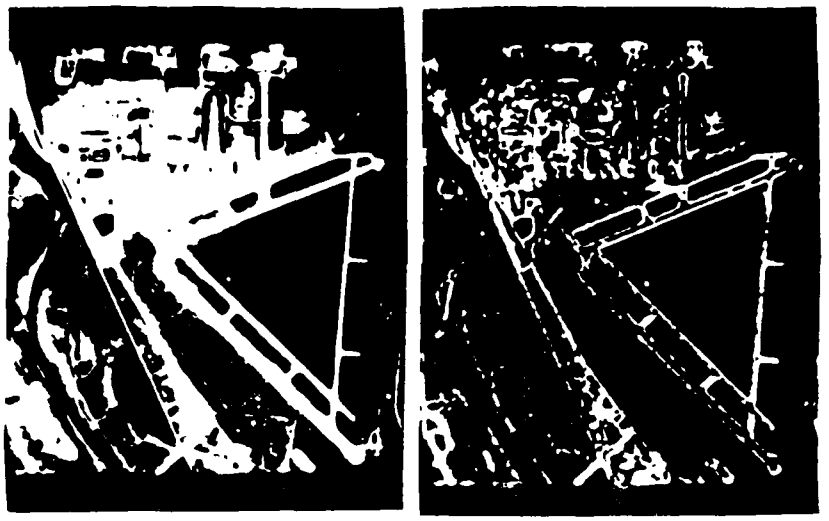


FIGURE 4 Input and output images demonstrating our new parallel optical nonlinear local edge-enhancement holographic convolution operator for the case of a 3 x 3 Sobel operator.

6. SYNTHETIC DISCRIMINANT FUNCTION HOLOGRAPHIC MATCHED SPATIAL FILTERS

A most attractive new holographic MSF concept is the use of synthetic discriminant functions (SDFs) to synthesize the MSFs for use in an optical correlator. These SDFs are linear combinations of an input image training set data base. When the SDFs are properly produced, they can provide a correlator whose output is invariant to any geometrical distortions of an input object. We refer to this as an intra-class SDF or as an equal correlation peak (ECP) SDF since it yields a correlation output whose intensity is a constant value independent of any geometrical distortions present in the input object [8-11]. Advanced versions of these SDFs can provide correlation outputs that enable one to discriminate between different objects (we refer to these as inter-class SDFs) and in more advanced cases to SDFs that retain intra-class recognition while enjoying inter-class discrimination as well.

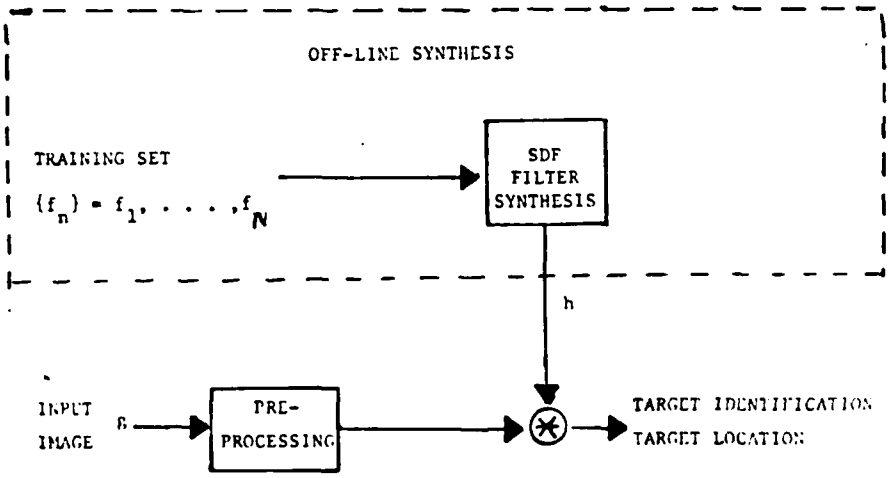


FIGURE 5 General functional block diagram for a holographic synthetic discriminant function matched spatial filter system for intra-class and inter-class pattern recognition.

The general approach to the synthesis of such SDFs is shown in Figure 5. In this general system block diagram, we use a training set of images $\{f_n\}$ which can be different aspect views of an object. From these $\{f_n\}$, we form an ECP SDF filter h such that $h \circledast f_n = c = 1$ (a constant c) for any aspect view of the object being search for. The ECP SDF $h(x,y)$ is a linear combination of the input training set images f_n,

$$h(x,y) = \sum_n a_n f_n(x,y). \tag{8}$$

We determine the coefficients a_n by forming the correlation matrix $\underline{R}$ of the training set data and inverting it. In matrix-vector notation, the coefficients a_n are a vector $\underline{a}$ that satisfy

$$\underline{a} = \underline{R}^{-1}\underline{u}, \tag{9}$$

where $\underline{u}$ is the unit vector equal to the constant unity. This ECP SDF and the advanced inter-class and the combined intra and inter-class SDF synthesis techniques together with their realization as holographic MSFs are detailed elsewhere in this conference [20].

7. SUMMARY AND CONCLUSION

As these brief pages have noted, OPR research now enjoys many new techniques and architectures. Many of these new concepts utilize holographic techniques as we have just described. These techniques and algorithms are addressing practical pattern recognition problems and the results of this basic research program are increasing the practicality of OPR systems and the repertoire of operations achievable on such processors. The recent research results described in this paper have included new holographic optical element architectures, nonlinear local optical operators and advanced matched spatial filters using synthetic discriminant functions and other techniques.

ACKNOWLEDGMENTS

Th support of this research at Carnegie-Mellon University by the Air Force Office of Scientific Research (Grant AFOSR 79-0091) is gratefully acknowledged. The author thanks all present and past members of the Hybrid Optical and Digital Image and Signal Processing Laboratory at Carnegie-Mellon University whose research results were highlighted in this paper.

REFERENCES

1. A. Van der Lugt, IEEE, IT-10, 139 (1964).
2. J. Rao, JOSA, 56, 1490 (1966).
3. F. Caimi et al, Applied Optics, 19, 2653 (1980).
4. D. Casasent and A. Furman, Applied Optics, 16, 1652 (1977) and Applied Optics, 16, 1662 (1977).
5. D. Casasent and D. Psaltis, Proc. IEEE, 65, 77 (1977) and pp. 291-356 in Progress in Optics, editor E. Wolf, Volume XVI (North Holland Publishing Company, New York, 1979).
6. D. Casasent, Chapter in Optical Information Processing, S. Lee editor (Springer-Verlag, Heidelberg, 1981).
7. J. Chen and D. Casasent, Conference on Applications of Lasers and Electro-Optics, LIA (Boston, September 1982).
8. C. Hester and D. Casasent, Applied Optics, 19, 1758 (1980).
9. C. Hester and D. Casasent, SPIE, 292, 25 (August 1981).
10. C. Hester and D. Casasent, SPIE, 302, 108 (August 1981).
11. D. Casasent et al, SPIE, 360 (August 1982).
12. Optical Engineering, Special Issue on "Applications of Holography", 14 (September/October 1975).
13. H. J. Caulfield editor, Handbook of Holography (Academic Press, New York, 1979).
14. D. Casasent editor, Optical Data Processing: Applications (Springer-Verlag, Heidelberg (1978).
15. S. Lee editor, Optical Information Processing (Springer-Verlag, Heidelberg, 1981).
16. SPIE, 201, Optical Pattern Recognition (August 1979).
17. C. Calderone, Final Report on Contract DAAK 40-77-C-0089, Gruman Aerospace Corporation (17 December 1979).
18. J. Duthie et al, SPIE, 231 (April 1980).
19. S. Yao and S. Lee, JOSA, 61, 474 (1971).
20. D. Casasent, B. Kumar and V. Sharma, SPIE, 360 (1982).

11. MICROPROCESSOR-BASED FIBER-OPTIC ITERATIVE OPTICAL PROCESSOR

Reprinted from APPLIED OPTICS, Vol. 21, page 147, January 1, 1982

Microprocessor-based fiber-optic iterative optical processor

Mark Carlotto and David Casasent

The design and fabrication of an iterative optical vector–matrix processor are described. Microprocessor feedback is used to produce an iterative processor capable of solving simultaneous linear equations. It also facilitates scaling and biasing of the data and the handling of bipolar and complex-valued data as well as correction for selected system error sources. Fiber-optic interconnections are used to improve the system's alignment and to reduce its size, weight, and errors. The design, fabrication, and performance of the system are analyzed.

I. Introduction

Optical vector–matrix multipliers[1-3] represent a general class of optical processors since many data processing problems can be formulated as vector–matrix equations or as a set of simultaneous linear algebraic equations. One of the most attractive ways to realize an optical vector–matrix multiplier with present-day hardware is to image a linear array of LEDs through a 2-D mask and onto a linear photodetector array.[3] Both serial systems[4,5] using one LED (whose output is time-sequentially modulated) and parallel systems[6,7] (with a linear array of input LEDs) have been described to achieve an optical vector–matrix multiplication. In both cases, the LED outputs describe the elements of a vector, the transmittances of the 2-D mask describe the elements of a matrix, and the system's output is a vector–matrix product. With a 2-D output CCD shift register detector, one can perform convolutions on such a system.[5] When the transmittances of the mask elements correspond to the Fourier kernels, the outputs on the linear photodetector array are the discrete Fourier transform of the sampled input data present on the linear LED input array.[6]

In Ref. 8 we described an iterative optical processor (IOP) in which the linear photodetector outputs from a vector–matrix multiplier were combined with an external vector, and the result fed back to the linear LED inputs. The use of this system in adaptive phased array radar processing,[9] eigenvalue and eigenvector computation,[10,11] and for optimal control applications[12] has been described. In the course of these application studies increasingly complex operations and control were required in the electronic feedback loop, and more attention to the system's accuracy was necessitated. In this paper, we describe the microprocessor-based fiber-optic IOP system we designed and fabricated to address future applications. In Sec. II we describe the new iterative algorithm we use with emphasis on the algorithm's convergence. Scaling, biasing, and how bipolar and complex-valued data are handled on the system are described in Sec. III. Following a description of the microprocessor-based fiber-optic IOP system we designed and fabricated (Sec. IV), an error analysis and quantification of the laboratory system's performance are advanced in Sec. V.

II. Convergence of the Iterative Algorithm

In Fig. 1 we show a simplified schematic diagram of the IOP. Bipolar-valued matrices are denoted by $\mathbf{H}$ and bipolar-valued vectors by $\mathbf{y}$ and $\mathbf{x}$. These are distinguished from the vector and matrix quantities $(\mathbf{a},\mathbf{c},\mathbf{B})$ in the physical optical system. This is necessary, since the latter must be real and positive. Complex-valued quantities will be denoted by a tilde above the variable. Considering the physical system first, we denote the light distribution leaving the linear input LED array at $\mathbf{P}_1$ at iteration j by the vector $\mathbf{a}^T(j)$ with elements $a_m(j)$. The light distribution leaving $\mathbf{P}_1$ is imaged vertically and expanded horizontally to uniformly illuminate the rows of the mask at $\mathbf{P}_2$ with light from the corresponding input LED. We denote the transmittance of the mask at $\mathbf{P}_2$ by the matrix $\mathbf{B}^T$ with elements b_{mn}. The light

When this work was done both authors were with Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, Pennsylvania 15213; M. Carlotto is now with Analytical Sciences Corporation, 1 Jacob Way, Reading, Massachusetts 01867.

Received 24 July 1981.
0003-6935/82/010147-06$01.00/0.

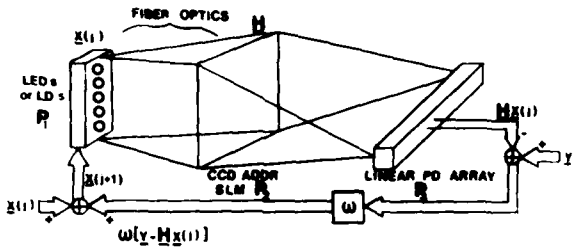


Fig. 1. Schematic diagram of the IOP emphasizing the iterative algorithm.

distribution leaving each column of $\mathbf{P}_2$ is summed on the corresponding photodetector and the linear output detector array at $\mathbf{P}_3$. The system's output is thus the vector–matrix product $\mathbf{c}^T(j) = \mathbf{a}^T(j)\mathbf{B}^T$. For notational simplicity, we will suppress the use of transposed vectors and matrices and thus describe the vector–matrix product by

$$\mathbf{c}(j) = \mathbf{B}\mathbf{a}(j), \tag{1}$$

with elements

$$c_n = \sum_{m=1}^{M} a_m b_{mn}, \tag{2}$$

where $\mathbf{c}$ has N elements, $\mathbf{a}$ has M elements, and $\mathbf{B}$ is an $M \times N$ matrix.

If the light distribution leaving the LEDs at $\mathbf{P}_1$ at iteration j is $\mathbf{x}(j)$, and the mask transmittance is described by $\mathbf{H}$, the output from the photodetectors at $\mathbf{P}_3$ is $\mathbf{H}\mathbf{x}(j)$. We form the difference between this output and a fixed external vector $\mathbf{y}$, multiply the difference by an acceleration parameter ω, and add the result to the original input vector $\mathbf{x}(j)$ to form a new $\mathbf{x}(j+1)$ input for iteration $j+1$. The system thus realizes the Richardson algorithm[13]

$$\mathbf{x}(j+1) = \mathbf{x}(j) + \omega[\mathbf{y} - \mathbf{H}\mathbf{x}(j)]. \tag{3}$$

When $\mathbf{x}(j) \cong \mathbf{x}(j+1) = \mathbf{x}$, Eq. (3) reduces to

$$\mathbf{H}\mathbf{x} = \mathbf{y}, \tag{4}$$

and the system's output

$$\mathbf{x} = \mathbf{H}^{-1}\mathbf{y} \tag{5}$$

is the desired solution to the vector–matrix equation in Eq. (4).

To insure convergence of Eq. (3) for all initializations $\mathbf{x}(0)$ of the system, the N roots $s_n(\omega)$ of the characteristic equation:

$$\text{determinant}|s\mathbf{I} - (\mathbf{I} - \omega\mathbf{H})| \tag{6}$$

of the discrete time system must lie strictly within the unit circle in the z-plane.[14] If λ_i are the eigenvalues of $\mathbf{H}$, we must satisfy

$$0 < |1 - \omega\lambda_i| < 1 \tag{7}$$

to insure convergence of Eq. (3). An obvious choice for ω is

$$\omega = -1/\lambda_{max}, \tag{8}$$

where λ_{max} is the absolute value of the largest eigenvalue of $\mathbf{H}$. With ω chosen as in Eq. (8), Eq. (7) is satisfied for the largest λ_i and thus is easily satisfied for smaller λ_i values. In practice, we can increase ω slightly from the value in Eq. (8) to insure that $|1 - \omega\lambda_i| > 0$. In special cases, when highly oscillatory poles of $\mathbf{H}$ occur, we can select ω to be a multiple of $1/\lambda_{max}$. (We have yet to consider cases when this situation arises.)

Let us now consider how to compute the choice of ω noted in Eq. (7). We generally use the conservative upper-bound[15]

$$\lambda_{max} < ||\mathbf{H}|| = \left(\sum_m \sum_n h_{mn}^2\right)^{1/2} \tag{9}$$

for λ_{max}. However, it is also possible to use the IOP itself to estimate λ_{max} as we now describe.

In this case, we let $\mathbf{y} = 0$, place $\mathbf{H}$ at $\mathbf{P}_2$, and describe the initial input vector $\mathbf{x}(0)$ at iteration $j = 0$ by

$$\mathbf{x}(0) = a_1\boldsymbol{\phi}_1 + a_2\boldsymbol{\phi}_2 + \ldots + a_M\boldsymbol{\phi}_M, \tag{10}$$

where the $\boldsymbol{\phi}_m$ are the eigenvectors of $\mathbf{H}$. We feed the output at each iteration directly back to the input and thus after j iterations find

$$\mathbf{x}(j) = \mathbf{H}^j\mathbf{x}(0). \tag{11}$$

We can write $\mathbf{H}$ in terms of its eigenvalues λ_m and its eigenvectors $\boldsymbol{\phi}_m$ by singular value decomposition as

$$\mathbf{H} = \sum_m \boldsymbol{\phi}_m \lambda_m \boldsymbol{\phi}_m^T. \tag{12}$$

Multiplying both sides of Eq. (10) by $\boldsymbol{\phi}_m^T$ and using the orthonormality of eigenvectors, we find $a_m = \boldsymbol{\phi}_m^T\mathbf{x}(0)$. Using this in Eq. (11), we find that $\mathbf{x}(j)$ can be rewritten as

$$\mathbf{x}(j) = \sum_m \boldsymbol{\phi}_m \lambda_m^j a_m. \tag{13}$$

After a sufficiently large number j of iterations, the eigenvector $\boldsymbol{\phi}_d$ with the largest eigenvalue λ_{max} will dominate the summation in Eq. (13), and the system's output will be

$$\mathbf{x}(j) \cong \boldsymbol{\phi}_d \lambda_{max}^j a_d. \tag{14}$$

From the ratios $x_m(j+1)/x_m(j)$ for j large, we find λ_{max}.

It is also possible to extend this conventional power method[16] to allow computation of all the eigenvalues and eigenvectors of $\mathbf{H}$ on the IOP as noted in Refs. 10 and 11. In practice, we normally use Eq. (9) to estimate λ_{max}. Since the calculation of the Euclidean norm of $\mathbf{H}$ in Eq. (9) is easily achieved in the microprocessor system, and since the calculation need only be done once and the same acceleration parameter ω used for all iterations, the technique in Eq. (9) is used in preference to the one in Eq. (14). When ω is properly chosen, the system's iterative algorithm monotonically converges, and the stability of the algorithm and the IOP system are assured. This solution in Eqs. (8) and (9) has worked well for all vector–matrix and matrix–matrix problems to which we have applied the IOP. Even when the matrix is ill-conditioned, use of this acceleration parameter insures convergence of the algorithm, although many iterations may be required.

III. Bipolar and Complex-Valued Data

The LED and photodetectors outputs as well as the transmittances of the mask in the system of Fig. 1 must be real and positive. Since noncoherent light is used, the IOP cannot handle bipolar or complex-value data directly. This is a severe limitation of the system, and thus much work has been done to allow processing of such data on a vector–matrix system.[17] When the vector–matrix multiplier in Ref. 6 is used to compute the discrete Fourier transform of the spatial data present across the LED array, complex- valued data must be handled by the system. This is accomplished by formatting the input vector and the fixed elements of the matrix mask in terms of the bipolar real (Re) and imaginary (Im) parts of the vectors and matrix as

$$\begin{bmatrix} y_{Re} \\ y_{Im} \end{bmatrix} = \begin{bmatrix} H_{Re} & -H_{Im} \\ H_{Im} & H_{Re} \end{bmatrix} \begin{bmatrix} x_{Re} \\ x_{Im} \end{bmatrix}. \tag{15}$$

This requires $2M$ input LEDs, a $2M \times 2N$ mask, and a $2N$ element output detector. One can handle bipolar data on the system in many ways.[17] One technique that has been used[6] is to bias all vector and matrix elements so that they are positive. To obtain the bipolar vector–matrix product from the measured outputs, electronic postprocessing is needed. To perform this, additional factors such as the product of known bias matrices and the unknown input vector are necessary. These can be obtained by adding a column to the matrix mask at $\mathbf{P}_2$ that contains all constant elements and by including an additional detector element in the output.[17]

In our IOP we handle complex data as in Eq. (15), but we accommodate the bipolar data in Eq. (15) differently. We first decompose the bipolar input vector $\mathbf{x}$ into its positive $\mathbf{x}^+$ and negative $\mathbf{x}^-$ parts. The optical system's input vectors corresponding to each of these are $\mathbf{a}_1$ and $\mathbf{a}_2$, respectively, with elements

$$a_{1m} = 0.5(x_m + |x_m|) \qquad a_{2m} = -0.5(x_m - |x_m|). \tag{16}$$

The elements of the optical mask $\mathbf{B}$ are a scaled and biased version of the elements of the bipolar matrix $\mathbf{H}$. Specifically

$$b_{mn} = (h_{mn} - \underline{h})/(\bar{h} - \underline{h}), \tag{17}$$

where $\underline{h}$ and $\bar{h}$ are the minimum and maximum elements of $\mathbf{H}$. With the $\mathbf{P}_2$ mask described by Eq. (17), we see that the elements of B satisfy $0 \le b_{mn} \le 1$ as is necessary. We then achieve a bipolar vector–matrix multiplier by operating the system twice, once with $\mathbf{a}_1$ as the input vector and once with $\mathbf{a}_2$ as the input vector, with the same fixed $M \times N$ mask $\mathbf{B}$ present in both cases. The microprocessor forms the difference $\mathbf{Ba}_1 - \mathbf{Ba}_2$ between the system's outputs in the two cycles and scales and biases the difference according to

$$\mathbf{y} = \mathbf{Hx} = (\bar{h} - \underline{h})(\mathbf{Ba}_1 - \mathbf{Ba}_2) + \underline{h} \sum_m x_m (1, \ldots, 1)^T. \tag{18}$$

All the required operations in Eq. (18) are easily performed in the microprocessor support system since only additions, subtractions, and multiplications by fixed constants are required. The two-cycle algorithm in Eq. (18) by which bipolar data are handled on the IOP has two other attractive features worth noting. First, the size of the input LED array is only M (rather than $2M$) and the size of the output photodetector array is N (not $2N$). Likewise the matrix mask must be $M \times N$ (rather than $2M \times 2N$). Thus, larger vectors and matrices can be handled on a given system at the expense of using the two cycles of the system. Since the potential speed of the IOP is so high, trading a factor of 2 in speed for a factor of 2 or 4 in the sizes of the vectors and matrices that can be handled appears to be a useful trade off for the applications with which we are concerned. A second practical feature resulting from the use of the algorithm in Eq. (18) is that all fixed pattern detector noise is automatically cancelled. In Fig. 2 we show the system's outputs ($\mathbf{c}_1 = \mathbf{Ba}_1$ and $\mathbf{c}_2 = \mathbf{Ba}_2$) on two successive iterations with no input present (i.e., $\mathbf{a}_1 = \mathbf{a}_2 = 0$) and the electronically calculated difference $\mathbf{c}_1 - \mathbf{c}_2$. The outputs are thus caused by detector noise only. As seen in the $\mathbf{c}_1$ and $\mathbf{c}_2$ outputs, the detectors have a large leakage current ($\cong$8% of full scale) in a fixed spatial pattern. However, after subtraction all fixed pattern noise and leakage current effects are canceled (as is seen in the $\mathbf{c}_1 - \mathbf{c}_2$ difference output), and we are left with only the temporal noise variations (Johnson noise) of the detector. In Fig. 2, this noise is measured to be <0.4% of full scale.

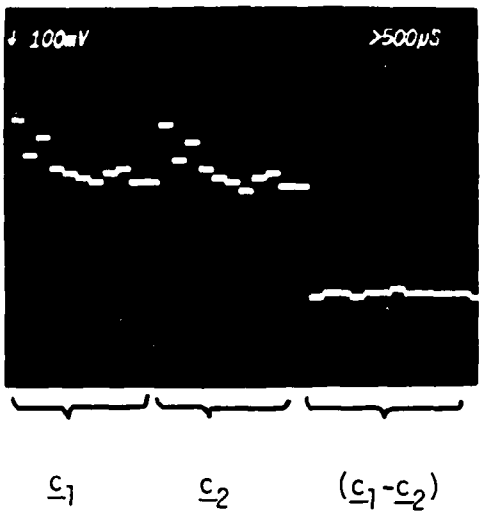


Fig. 2. Photodetector output $\mathbf{c}_1$ and $\mathbf{c}_2$ at two successive iterations and their difference showing suppression of detector leakage current and fixed-pattern noise.

IV. System Fabrication

Let us now consider the laboratory IOP system we fabricated with attention to the microprocessor system and the fiber-optic interconnections and how these features are used to overcome many potential system error sources. A detailed analysis of the system's error sources and quantitative data on the system's performance are included in Sec. V.

The electronic feedback system was to compute: the running iterative sum on the right-hand side of Eq. (3), the acceleration parameter ω in Eq. (9), the LED preprocessing in Eq. (16), the matrix scaling and biasing in Eq. (17), and the detector postprocessing in Eq. (18). It must also properly sync, control, and format the LED inputs and system cycling for bipolar and complex-value data handling. In practice, LED and photodetector correction factors are also included in the preprocessing and postprocessing (Sec. V). All these operations can be hard-wired and performed at high speed in a dedicated system. However, to enable the use of the IOP to be studied for many diverse problems and applications, a flexible rather than a dedicated electronic support system was desired. We achieved this with a microprocessor support system. A schematic of the IOP emphasizing the microprocessor electronic feedback and support system is shown in Fig. 3.

The electronic feedback system contains four subsystems:

(1) An LED board that performs the necessary preprocessing in Eq. (16) for handling bipolar and complex-valued data, provides the multiplexed pulsewidth-modulated current drive for the LEDs and corrections for nonuniform LED saturation levels.

(2) A detector board with parallel resettable operational-amplifier (op-amp) integrators to allow variable detector integration times and correction for nonuniform photodiode responsivity.

(3) Analog to digital (ADC) and digital to analog (DAC) converters for input to and output from the microprocessor controller.

(4) a microprocessor controller subsystem. This controls the scheduling of all IOP operations and performs the operations in Eqs. (3), (9), (17), and (18).

The microprocessor subsystem contains a control section with a Fairchild 9408 LSI microprogram sequencer to execute various microprograms stored in a 26K random access memory (RAM) program memory and a thirty-two line instruction decoder to activate various control points in the system. It also contains an arithmetic data section containing a custom-designed arithmetic unit consisting of a 16-bit 300-nsec multiplier, a 16-bit arithmetic logic unit (ALU), and a 16K RAM with a special row-column address structure. In this data section all arithmetic operations are performed at high speed. The 16K RAM is used to store fixed data such as LED and photodetector response correction factors ω, $(\bar{h} - \underline{h})$, etc. The system is also provided with a capability of storing up to fifty-four different selected iterative data outputs for future display on a scope or for input to a microcomputer for analysis. The laboratory IOP system also contains a front panel console from which the operator can load any programs into the 26K or 16K memory depending on the IOP application being considered. It also includes all necessary operator controls to start, stop, and reset the microprocessor IOP.

A photograph of the full microprocessor-based IOP is shown in Fig. 4. The front panel is shown at the top. Below this is the microprocessor system. The optical

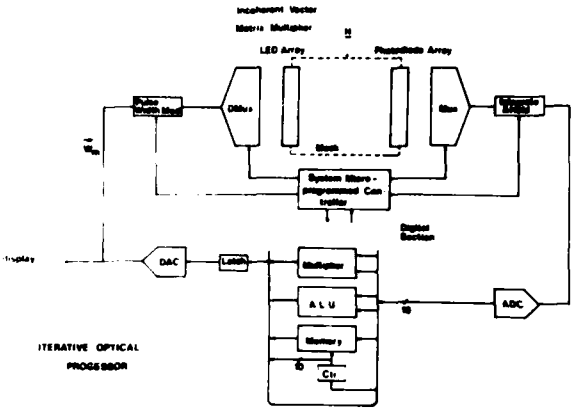


Fig. 3. Schematic diagram of the IOP emphasizing the microprocessor-based electronic support system.

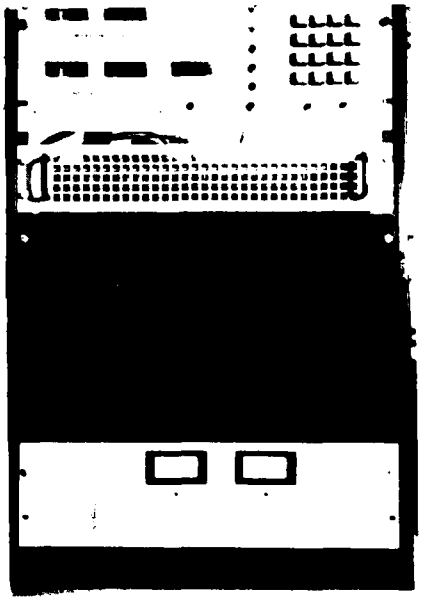


Fig. 4. Laboratory IOP and its entire electronic support system (microprocessor, power supply, front panel console).

IOP system is placed behind the central dark panel, and the system's power supplies are on the bottom of the rack. The present electronic support system contains 160 integrated circuits, requires 50 W of power, and has a 300-nsec cycle time. Higher speed is possible, but the flexibility and cost of the assembled system just described were more compatible with our goals. Moreover, it provided us with a sufficiently powerful system to be used on many different applications and problems, and sufficient complexity to allow unforeseen problems in the design and fabrication of larger systems to be uncovered.

The laboratory IOP system uses a linear input array of ten RCA SG-1002 LEDs at plane P_1 of Fig. 1 with 1-mW output at 940 nm for a 50-mA drive current. The LEDs are mounted on 0.375-cm centers along a copper block 3.75 cm long. They are held in place by silver epoxy and sealed in white RTV compound. The inner connections from P_1 to the mask at P_2 are accomplished by a specially fabricated fiber-optic system. It contains a linear array of ten apertures at one end into which the LEDs are placed. Each aperture contains ten glass fibers, each 25 μm in diameter, that branch outward to form a line of ten fibers. The output from the fiber-optic system is thus a 10×10 array of 100 fibers whose locations match the 10×10 elements of the mask at P_2. The fiber outputs have a center-to-center spacing of 0.35 mm vertical and 0.94 mm horizontal. The P_2 mask is placed between these fiber-optic outputs in the detector array. In the present system, the mask used is a fixed pattern recorded on film. An advanced IOP system we are presently designing will use a real-time light modulator (such as the CCD-addressed liquid crystal light valve[18]) as a real-time adaptive mask element. The detector used in the present system is a Centronics LD-20 silicon photodiode array containing twenty elements each measuring 4×0.9 mm on 0.94-mm centers. The spacings and sizes of the outputs from the fiber-optic system were chosen to match the size of the elements in this detector. The horizontal spacing between fibers equals the spacing between detector elements, and the height of the ten vertical fiber outputs equals the height of a detector element. This allowed us to sandwich the mask between the output from the fiber-optic element and the detector with no imaging optics necessary between P_2 and P_3 in Fig. 1. A photograph of the optical vector–matrix multiplier is shown in Fig. 5. From right to left are the LED array, fiber-optic connector, mask, and photodetector array. The photodetector board is also visible on the left. The components in the system have been separated for clarity in the photograph. In practice, the entire system is less than 5 cm long.

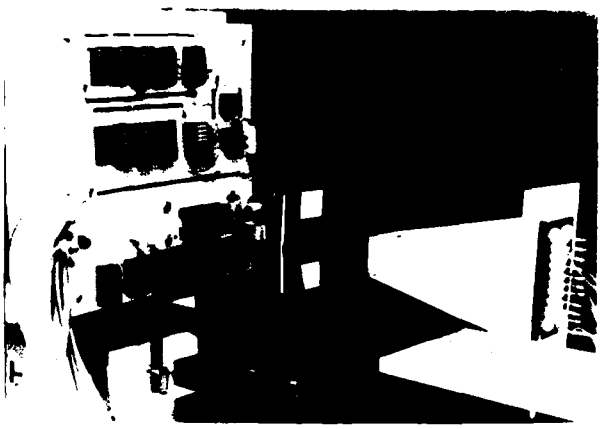


Fig. 5. Laboratory optical vector–matrix multiplier.

V. System Performance

In this section we discuss the performance of the laboratory IOP system we fabricated and emphasize how many system designed features were chosen to improve the accuracy and stability of the system. The first IOP that we fabricated[9] used cylindrical lenses for imaging from P_1 to P_2 and from P_2 to P_3. Experiments and simulation analysis on this system showed that two major error sources were cross talk in the vertical imaging from P_1 to P_2 and nonuniform illumination of each row of the P_2 mask. The fiber-optic system (Sec. IV) effectively removes both of these error sources. Similar problems were found to occur in the required imaging from P_2 to P_3. By placing the detector, mask, and fiber-optic element in contact, these error sources were similarly removed, and a rugged stable system (Fig. 4) of greatly reduced size and weight resulted.

Amplitude nonlinearities in the light outputs from the LEDs is a well-known problem. These components are thus usually operated at a fixed bias current and amplitude-modulated over a restricted range to decrease nonlinear effects. This results in a decreased useful linear dynamic range. Correction for LED nonlinearities is possible but was not included in the laboratory system. Rather, we use pulse width modulation (PWM) of the LEDs when linear performance and large dynamic range data are needed. When operated in the PWM mode, the present laboratory system has a minimum to maximum pulse-width ratio of 256 and thus a 256:1 input dynamic range. This has proven adequate for all applications with which we are concerned. Use of laser diodes rather than LEDs for the input source array allows amplitude modulation with a large linear dynamic range. However, linear laser diode arrays are not yet commercially available, and thus our present system is operated with an LED input source array.

When operated in the PWM data mode, a large spatial nonuniformity in the output power from the LEDs of $\pm 25\%$ was measured. This fixed error is corrected for by multiplying the input signal to the mth LED by the reciprocal of its response. Spatial nonuniformities of $\pm 7\%$ were measured for the responsivity of the output photodetector elements. These output nonuniformities were similarly corrected for by multiplying each photodiode output by its appropriate reciprocal responsivity correction factor. These multiplicative source and detector corrections are directly included in the preprocessing and postprocessing with no additional overhead, since the correction factors can be measured once and stored in the microprocessor system's 16K data memory. As noted in Sec. III, the bipolar data handling algorithm in Eq. (18) automatically cancels all fixed-pattern detector noise and detector leakage current effects.

Residual spatial nonuniformity errors caused by the source and detector may still remain. In addition, differences in the coupling loss from the different LEDs to the different fiber-optic elements may exist together with spatial variations in the outputs from the 100

fiber-optic elements (resulting mainly from differences in the polishing of the ends of each fiber). After applying the preliminary LED and photodiode corrections, we measured the 2-D spatial variations of the entire system and found a residual nonuniformity with a standard deviation of only 0.8%. This level of accuracy was sufficient for our purposes. It can be reduced further by placing a fixed correction mask (with transmittances inversely proportional to the system's 2-D spatial nonuniformity) in contact with the data mask at $\mathbf{P}_2$. In practice, we include these fixed corrections on the data mask itself when it is recorded.

From the discussion thus far, we find that all fixed spatial errors in the system can be reduced to nearly any desired level, and nearly any desired input data dynamic range can be achieved (by pulse width modulation with an associated loss in speed, by amplitude modulation of a laser diode source array, or by a combination of amplitude and PWM modulation). The major errors in the system are thus the time-varying thermal noise in the detector, and noise in the data recorded on the mask at $\mathbf{P}_2$. In the present system, the temporal detector noise is <0.4% of full scale. This can be further reduced by use of cooled detectors, advanced detector fabrication techniques, and chopper-stabilized operational amplifiers if necessary. Noise in the recorded data at $\mathbf{P}_2$ thus appears to represent the major limitation in the system's performance. A general analysis of the effect of this error source on the performance of the iterative algorithm is not possible. Rather, specific case studies and applications must be individually addressed. In general, we have found that the accuracy of the final answer in the iterative algorithm will be less than or equal to the error in a given vector–matrix product with final errors of 1% being quite easily achieved. In some cases, the $\mathbf{P}_2$ matrix mask can be row or column biased to reduce its dynamic range requirements. Such issues are best treated for specific applications. We are in the process of completing such analyses for adaptive phased array radar[9] and linear quadratic regulator control[12] applications on the IOP. These will be published and this issue addressed more fully when sufficient data and funding are available.

VI. Summary and Conclusion

In this paper, we have described the design and fabrication of an iterative optical vector–matrix processor and the performance possible from such a system. The use of a microprocessor electronic support system was shown to provide extensive flexibility in the laboratory system assembled. The use of fiber-optic interconnections was found to result in a rugged and stable system of small size and weight and remove cross talk and nonuniform illumination error sources. A new algorithm for handling bipolar data on the system was shown to provide cancellation of spatial fixed-pattern system noise. Source and detector nonuniformities are also easily corrected by RAM look-up tables. The residual spatial system error on the laboratory IOP assembled was <0.8%, and its temporal noise was below 0.4%.

This present system description has only emphasized its use in solving linear algebraic equations or vector–matrix equations. In Sec. II we noted that the system can also be used to compute the eigenvalues and eigenvectors of a matrix. Multiplication of two $M \times M$ matrices is also possible on the system by vectorizing one matrix or by running the system M times with the M columns of one matrix as inputs. Matrix inversion is similarly possible by describing the problem as N problems each of the form of Eq. (3) with $\mathbf{y} = 1$. We have also[12] used the system to solve nonlinear matrix–matrix problems using a modified Newton-Rhapson algorithm. In this latter application the solution involves an inner and an outer iterative loop, with the output from the inner loop fed to the outer loop after N iterations and a different mask necessary for each outer loop iteration. The IOP system thus appears to be a viable, powerful, flexible, and quite general purpose processor with many potential applications.

The authors thank Rome Air Development Center for initial support of this research, Air Force Office of Scientific Research (grant 79-0091) for interim support, and NASA Lewis Research Center (grant NAG 3-5) for present support of our IOP research.

References

1. L. Cutrona, in *Optical and Electro-Optical Information Processing*, J. Tippett *et al.*, Eds. (MIT Press, Cambridge, 1965), pp. 97–98.
2. A. Edison and M. Noble, Optical Analog Matrix Processors, A. D. 646060 (Nov. 1966).
3. P. Mengert *et al.*, U.S. Patent 3,525,856 (6 Oct. 1966).
4. M. Monahan, in *Digest of the International Optical Computing Conference*, IEEE Catalog 75-CH0941-5C (IEEE, New York, 1975), pp. 25–33.
5. M. Monahan, K. Bromley, and R. Bocker, Proc. IEEE **65**, 121 (1977).
6. J. W. Goodman, A. R. Dias, and L. M. Woody, Opt. Lett. **2**, 1 (1978).
7. J. Goodman *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. **190**, 484 (1979).
8. D. Psaltis, D. Casasent, and M. Carlotto, Opt. Lett. **4**, 348 (1979).
9. D. Psaltis *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. **180**, 114 (1979).
10. H. J. Caulfield, D. Dvore, J. W. Goodman, and William Rhodes, Appl. Opt. **20**, 2263 (1981).
11. B. Kumar and D. Casasent, Appl. Opt. **20**, 3707 (1981).
12. D. Casasent *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. **295** (1981).
13. L. F. Richardson, Philos. Trans. R. Soc. London, Ser. A **210**, 307–357 (1910).
14. E. Jury, *Theory and Applications of the z-Transform Method* (Kreiger, New York, 1973).
15. E. Kreyszig, *Advanced Engineering Mathematics* (Wiley, New York, 1972).
16. G. Stewart, *Introduction to Matrix Computations* (Academic, New York, 1973).
17. J. Goodman *et al.*, "Incoherent Optical Matrix-Vector Multiplier," Stanford U. Tech. Rept. L-723-1 (Feb. 1979).
18. J. Grinberg *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. **128**, 253 (1977).

12. MULTIDIMENSIONAL ADAPTIVE RADAR ARRAY PROCESSING USING AN ITERATIVE OPTICAL MATRIX-VECTOR PROCESSOR

Multidimensional adaptive radar array processing using an iterative optical matrix-vector processor

David Casasent
Mark Carlotto*
Carnegie-Mellon University
Department of Electrical Engineering
Pittsburgh, Pennsylvania 15213

Abstract. An iterative optical matrix-vector processor that computes the adaptive weights for a phased array radar is described. Multidimensional adaptivity in both target angle and velocity is achieved by lexicographically ordering the antenna elements as they are fed to the optical processor. Complex weights are computed by spatial multiplexing of the vector and matrix inputs to the system. The error sources of the optical system and the convergence of the iterative algorithm are analyzed, and experimental demonstration of the accuracy and performance of the system is included. This novel processor is found to perform quite adequately and to be most appropriate for advanced multidimensional adaptive phased array radars.

Keywords: two-dimensional signal processing; multidimensional adaptivity; radar array processing; iterative optical matrix-vector processor; adaptive phased array signal processing; angle adaptivity.

Optical Engineering 21(5), 814-821 (September/October 1982).

CONTENTS

1. INTRODUCTION

Adaptive phased array radar (APAR)[1-3] represents a formidable signal processing problem of considerable current interest[4,5] and one for which advanced signal processing concepts and algorithms are necessary. The real-time and parallel processing features of optical systems make them attractive candidates for this application. However, the nature of the APAR problem requires a new optical processing system that performs more general functions besides the Fourier transform and correlation operations normally realized in such systems.[6] In this paper, we describe a new and general purpose optical processor, discuss its application for APAR processing, provide experimental demonstrations of its use in APAR processing, and analyze the accuracy and performance of the system for this application.

In Sec. 2, we describe the APAR signal processing problem. Computation of the optimum set of adaptive weights to apply to the receiving elements of the antenna to steer it in a desired direction and to null all noise sources in other directions is formulated as the solution of a matrix-vector equation requiring the inversion of a matrix. In Sec. 3, we describe an iterative optical matrix-vector processor (IOP) that we have fabricated[7] to address this problem.[8] We also discuss how complex values are accommodated on this system and how convergence of the iterative algorithm is achieved. We also advance an error source model for the processor. The experimental use of the IOP to cancel noise sources distributed in angle is then demonstrated in Sec. 4. In Sec. 5, we extend our theory to the case of multidimensional adaptive antennas. Experimental demonstration of the use of the IOP for an antenna adaptive in both space and time is included in Sec. 6 together with an initial error source and accuracy analysis of this new optical processor for this application.

2. ADAPTIVE PHASED ARRAY SIGNAL PROCESSING

For simplicity, we initially consider a linear phased array antenna with adaptive steering and noise null cancellation in angle (θ) only. In Sec. 5, we extend this theory to the case of multidimensional adaptive antennas. Consider the linear (1-D) phased array antenna system of Fig. 1 with N isotropic elements spaced $d = \lambda_R/2$ apart (where λ_R is the wavelength of the radar). In the far field of the antenna (i.e., at ranges $R \gg (Nd)^2/\lambda_R$), we assume a signal $s(t)\exp(j\omega t)$ at an angle θ_0 (this represents the direction in which we wish to steer the antenna and obtain maximum response) and M uncorrelated, zero-mean, narrow-band interference sources $r_m(t)$

*Present address: The Analytical Sciences Corporation, 1 Jacob Way, Reading, MA 01867.

Invited Paper TD-102 received Apr. 27, 1982; revised manuscript received May 18, 1982; accepted for publication May 21, 1982; received by Managing Editor June 1, 1982.

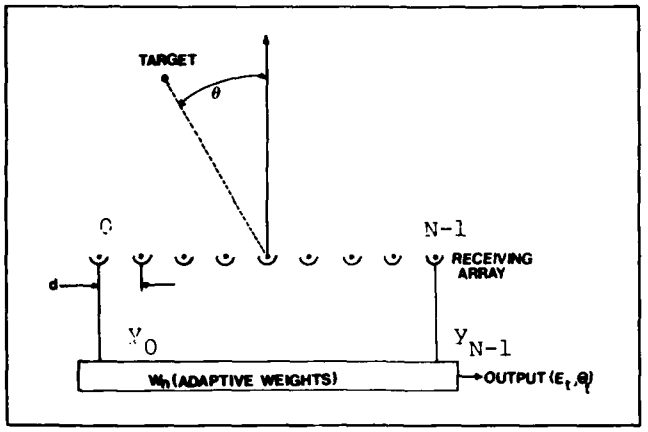


Fig. 1. Simplified pictorial block diagram of an adaptive phased array radar processor.

$\exp(j\omega t)$ at angles $\theta_1, \ldots, \theta_M$. All angles are measured with respect to boresight (the normal to the array). The objective of an APAR is to point the antenna in the direction θ_0 and to null the antenna pattern in the directions θ_m of the interference sources. With the signal and interference sources in the far field of the antenna, the radiation incident on the array can be described by the superposition of plane waves from the directions of each source. Since the path difference between two antenna elements is $d \sin\theta = (\lambda_R \sin\theta)/2$, the signal received at the n-th antenna element is

$$z_n(t) = s(t)e^{j(\omega t + \pi n \sin\theta_0)} + \sum_{m=1}^{M} r_m(t)e^{j(\omega t + \pi n \sin\theta_m)} . \quad (1)$$

Each of these N antenna outputs, $\underline{z}(t) = \{z_n(t)\}$, is multiplied by a complex weight $\underline{w} = \{w_n\}$, and the output from the receiver is the coherent summation of the products of the weights and the received signals:

$$v_{out}(t) = \sum_{n=0}^{(N-1)} w_n z_n(t) = \underline{w}^T \underline{z}(t) . \quad (2)$$

In Eq. (2) and in our future descriptions, we employ vector and matrix notation to describe the various signal components of the system. Lower (upper) case letters with an underbar denote vectors (matrices).

The antenna pattern that is obtained from such a receiver is described by an angular response $E(\theta)$ which is the inverse Fourier transform of the weighting pattern $\{w_n\}$. The attractive feature of a phased array radar is the ease with which one can steer the antenna. To direct the antenna to $\theta = \theta_0$, we simply weight the antenna outputs by the conjugate phase pattern $w_n = \exp(-j\pi n \sin\theta_0)$. When uncorrelated noise is present, due to uniform background radiation or thermal noise in the antenna itself, this weighting maximizes the signal-to-noise power ratio (SNR) at the antenna's output.[2] However, when directional interference is present, this simple weighting is not optimal, and the weights must be computed adaptively as a function of the changes in the rf noise environment. This is the APAR signal processing problem with which we are concerned. In Ref. 9, we show that the vector $\underline{w}$ which minimizes the mean square error between the signal and array output satisfies the matrix-vector equation

$$\underline{M}\underline{w} = P_0 \underline{s}^* , \quad (3)$$

where $\underline{M} = \overline{\underline{z}^*(t)\underline{z}^T(t)}$ the covariance matrix of the received signal plus interference; $P_0 = \overline{s^*(t)s(t)}$ the signal power; $\underline{s} = \exp(j\pi n \sin\theta_0)$ is the steering vector. In Ref. 9, we also show that the solution $\underline{w}$ to Eq. (3) also maximizes the SNR at the output of the antenna. These results are in agreement with the conclusions in Refs. 1, 2, and 4.

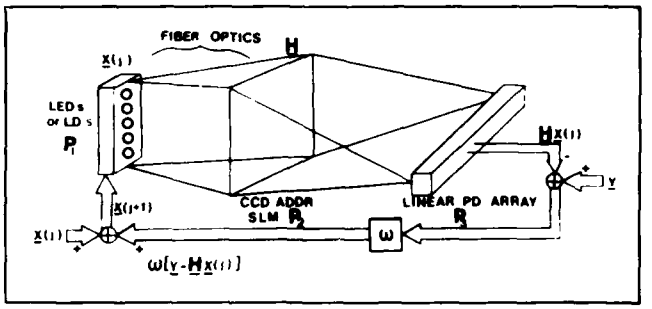


Fig. 2. Schematic diagram of the iterative optical matrix-vector processor.[7]

3. ITERATIVE OPTICAL PROCESSOR

From Sec. 2, we found that the solution of the optimal adaptive weights $\underline{w}$ that will steer an antenna in a direction $\underline{s}^*$ and null directional interference noise described by the covariance matrix $\underline{M}$ must satisfy the vector-matrix equation

$$\underline{M}\underline{w} = \underline{s}^* . \quad (4)$$

The constant multiplicative factor P_0 in Eq. (3) does not affect the computed solution $\underline{w}$. In Refs. 7-10, we described an iterative optical processor (IOP) that can solve matrix-vector equations or systems of linear algebraic equations such as Eq. (4). An improved version of this IOP is shown schematically in Fig. 2. The input at P_1 is a linear array of LEDs or laser diodes (LDs), whose outputs at time j describe a vector $\underline{x}(j)$. This vector output is imaged vertically and expanded horizontally to uniformly illuminate row m of P_2 with the input x_m. A 2-D mask whose transmittance is described by a matrix $\underline{H}$ is placed at P_2. The light distribution leaving each column of $\underline{H}$ is collected on a separate photodetector at P_3. The output from the linear photodetector array at P_3 is thus the matrix-vector product $\underline{H}\underline{x}(j)$.

With such a matrix-vector processor (as described in Refs. 11, 12 and more recently in Ref. 13) as the basic element of our system, we subtract an external vector $\underline{y}$ from $\underline{H}\underline{x}(j)$, multiply the difference by a constant acceleration parameter ω, and add the result to the prior $\underline{x}(j)$ input to obtain a new iterative input $\underline{x}(j+1)$ for time $j+1$. This IOP thus realizes the Richardson algorithm[14] in the form

$$\underline{x}(j+1) = \underline{x}(j) + \omega[\underline{H}\underline{x}(j) - \underline{y}] . \quad (5)$$

When $\underline{x}(j) = \underline{x}(j+1) = \underline{x}$, Eq. (5) reduces to

$$\underline{H}\underline{x} = \underline{y} , \quad (6)$$

and the system's output is

$$\underline{x} = \underline{H}^{-1}\underline{y} \quad (7)$$

or the solution of the matrix-vector equation in Eq. (6). Such a system can be directly used for the APAR problem described in Sec. 2 and summarized in Eq. (4). We simply use the covariance matrix $\underline{M}$ as the matrix $\underline{H}$, the steering vector $\underline{s}^*$ as the exogenous vector $\underline{y}$, and the solution vector $\underline{x}$ is then the desired set of adaptive weights $\underline{w}$ to be computed.

In Ref. 7, we detailed the fabrication and performance of the laboratory IOP system we assembled. As the inputs, we used ten LEDs which were interconnected to the mask by a fiber optic system. A film transparency is used as the mask in our present laboratory system, although a real-time 2-D light modulator such as the CCD-addressed liquid crystal light valve[15] can provide a real-time adaptive

mask for this system. Alternatively, a new optical systolic array architecture[16] using a 1-D acousto-optic cell at P_2 of Fig. 2 with feedback of the photodetector outputs to the acousto-optic cell's inputs[17] can also be used to provide a fully real-time iterative optical matrix-vector processor. For now, we consider only the system of Fig. 2. The height of the matrix mask used at P_2 is 4 mm, and it was chosen to match the height of the detector elements in our linear photodetector array at P_3. We also chose the horizontal spacings between detector elements and thus, in our present system, we simply place the output photodetector array in direct contact with the mask at P_2.

Since the outputs from the LEDs and the transmittances of the mask elements are real and positive, this system can multiply and add only positive numbers. In the APAR problem, the elements of the matrices and vectors are complex valued. Thus, to provide the complex-valued matrix-vector product in Eq. (4), we employ spatial and temporal multiplexing. We realize a bipolar matrix-vector product $\underline{y} = \underline{H}\underline{x}$ by scaling, biasing, and partitioning $\underline{H}$ such that it is a unipolar matrix. We then operate the IOP twice. On the first cycle, the positive values $\underline{x}_1$ of $\underline{x}$ are the inputs, and on the second cycle the negative values $\underline{x}_2$ of $\underline{x}$ are the LED inputs. In the postprocessor, we form the difference $\underline{H}\underline{x}_1 - \underline{H}\underline{x}_2$ of the two successive matrix-vector outputs and scale and bias the result to provide the new inputs for the next iterative cycle. This procedure is detailed in Ref. 7, where we discuss how this procedure reduces the required space bandwidth product of the mask and enables all fixed pattern detector noise to be canceled. To enable the system to perform complex-valued matrix-vector multiplications, we partition the matrix and the vectors in the system as

$$\underline{y} = \begin{bmatrix} \underline{s}^*_{re} \\ \underline{s}^*_{im} \end{bmatrix}, \quad \underline{H} = \begin{bmatrix} \underline{M}_{re} & -\underline{M}_{im} \\ \underline{M}_{im} & \underline{M}_{re} \end{bmatrix}, \quad \underline{x} = \begin{bmatrix} \underline{w}_{re} \\ \underline{w}_{im} \end{bmatrix}. \tag{8}$$

where the subscripts re and im denote the bipolar real and imaginary parts of the indicated vectors and matrices. To accommodate complex-valued matrix-vector operations on the system, we bias $\underline{M}$, form a unipolar matrix as before, partition it as indicated in Eq. (8), format the vectors $\underline{x}$ and $\underline{y}$ as in Eq. (8), and operate the system for two cycles with the positive- and negative-valued vector elements as the inputs. We detail this complex-valued algorithm and demonstrate its use for a specific example in Sec. 6. In Sec. 4, we describe an alternate complex-valued algorithm and demonstrate its use on the IOP system of Fig. 2.

A new feature included in the system of Fig. 2 is the presence of the acceleration parameter ω. (We retain the standard notation ω for the acceleration factor. This will cause no confusion with the radian frequency ω in practice.) Proper selection of ω ensures convergence of the iterative algorithm and speeds the convergence, as we now discuss. Since $\underline{M}$ is a positive definite Hermetian matrix, its N eigenvalues λ_n are positive, and thus to ensure convergence of Eq. (5), we require ω to satisfy[7]

$$|1 - \omega\lambda_n| < 1. \tag{9}$$

This is ensured by the choice[7]

$$\omega = ||\underline{H}||^{-1} = \left(\sum_m \sum_n h_{mn}^2 \right)^{-1/2}, \tag{10}$$

where the Euclidean norm of $\underline{H}$ (the square root of the sum of the squares of the elements h_{mn} of $\underline{H}$) is represented by the symbol shown. When the spatial multiplexing in Eq. (8) is used, $||\underline{H}|| = \sqrt{2}||\underline{M}||$ is used in Eq. (10).

The accuracy and performance obtainable for any analog or optical processor is an issue of primary concern. In the error source model we have developed for the IOP, we separate the errors of the system into spatially-fixed and temporal errors. In terms of these errors, we describe the observed output $\hat{\underline{z}} = \underline{H}\underline{x}$ obtained from the matrix-vector multiplier as the exact result $\underline{z}_{exact}$ plus two terms:

$$\hat{\underline{z}} = \underline{z}_{exact} + \underline{z}_{spatial} + \underline{z}_{temporal}. \tag{11}$$

The spatially fixed errors in the IOP are due to nonuniformity and nonlinearity in the LED and detector responses, spatial variations in the transmittances of the fiber optic interconnections, and errors in the transmittances of the elements of the mask. We can correct for the source and detector errors by multiplying the inputs to the LEDs by a fixed correction vector stored in read-only memory. The residual spatial nonuniformities that remain can all be transferred to the mask plane. This is quite attractive since we can then correct for all residual spatial errors by properly modifying the matrix data as they are recorded on the mask. For our laboratory system, the measured residual spatial error without a correction mask was $\pm 0.8\%$.[7] This represents $\underline{z}_{spatial}$ in Eq. (11) under uniform LED illumination. Since it is adequate for our applications, as we will see, no further corrections for it were included in our present system. The temporal time-varying component of the system noise $\underline{z}_{temporal}$ in Eq. (11) is due to the detector. It was measured to be $\pm 0.4\%$ for our system. As before, this is sufficiently small that cooled detectors and other measures to decrease this noise component were not used. This latter error source represents the fundamental limit and performance of the IOP. In Sec. 6, we use our error source model in Eq. (11) and present an initial analysis (with experimental confirmation) of the performance of the IOP for a multidimensional adaptive antenna.

4. ANGLE ADAPTIVITY USING THE IOP

As an initial example of the use of the IOP for APAR processing, we consider interference sources distributed only in angle as described in Sec. 2. We also use this initial example to detail an alternate method to process complex-valued data on the IOP. We consider a two-element array with one interference source at an angle θ_1 with noise power P_1 (per received channel) and with additive receiver noise of N_r watts per channel. We ignore the signal strength in this present treatment. The received signals at the two array elements are

$$z_1 = x_1 + y_1 ;$$

$$z_2 = x_2 + y_2 , \tag{12}$$

where x_n and y_n are the interference voltage and noise voltage in channel n. The voltage x_2 will lag x_1 by a phase angle $\gamma = \pi \sin\theta_1$ (where $d = \lambda_R/2$ is assumed), and the noise voltages y_n will be independent of each other and of the x_n signals. For this case, the covariance matrix is

$$\underline{M} = \begin{bmatrix} P_1 + N_r & P_1 \exp(-j\gamma) \\ P_1 \exp(+j\gamma) & P_1 + N_r \end{bmatrix}. \tag{13}$$

In this initial experiment,[8] we set $\omega = 1$ in Eq. (5), and to avoid the need to add the original input to the difference between the matrix-vector product and the exogenous vector, we place $[\underline{I} - \underline{M}]$ on the mask, where $\underline{I}$ is the identity matrix. The iterative algorithm of Eq. (5) now becomes

$$\underline{x}(j+1) = \underline{x}(j)\,[\underline{I} - \underline{M}] + \underline{y}.$$

For the specific case chosen, we used $P_1 = 0.1$ watts and $N_r = 0.5$ watts (these N_r and P_1 values ensure convergence, and thus the acceleration factor can be unity) and chose θ_1 such that $\lambda = 4\pi/3$. The required matrix mask is thus

$$[\underline{I} - \underline{M}] = \begin{bmatrix} 0.4 & 0.1\exp(-j4\pi/3) \\ 0.1\exp(j4\pi/3) & 0.4 \end{bmatrix}. \tag{14}$$

To realize the complex mask transmittance in Eq. (14), we introduce an alternate technique[18] in which a complex-valued number m′ (one component of the mask) is represented by three real and positive components (m_0', m_1', and m_2') which are the projections of the mask elements along axes at angles 0°, 120°, and 240° in complex space, i.e.,

$$m' = m_0' \exp(j0) + m_1' \exp(j2\pi/3) + m_2' \exp(j4\pi/3) . \quad (15)$$

For the specific matrix in Eq. (14), the phase angles of its four components are 0°, 120°, and 240°, and thus we can represent Eq. (14) by the three matrices

$$[\underline{1} - \underline{M}]_0' = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.4 \end{bmatrix} , \quad [\underline{1} - \underline{M}]_1' = \begin{bmatrix} 0 & 0.1 \\ 0 & 0 \end{bmatrix} ,$$

$$[\underline{1} - \underline{M}]_2' = \begin{bmatrix} 0 & 0 \\ 0.1 & 0 \end{bmatrix} . \quad (16)$$

We note that each of these matrices and all of their elements are positive. As the mask used at P_2 of Fig. 2, we thus spatially multiplexed these three matrices, and as the formatted matrix mask we used

$$[\underline{1} - \underline{M}] = \left[\begin{array}{cc|cc|cc} 0.4 & 0 & 0 & 0 & 0 & 0.1 \\ 0 & 0.4 & 0.1 & 0 & 0 & 0 \\ \hline 0 & 0.1 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.1 & 0 \\ \hline 0 & 0 & 0 & 0.1 & 0.4 & 0 \\ 0.1 & 0 & 0 & 0 & 0 & 0.4 \end{array}\right] . \quad (17)$$

For the case of two adaptive antenna elements, there will be two complex-valued weights. We represent these by w_a and w_b. Each of these weights will have three positive-valued elements with a decomposition similar to that used in Eq. (15); i.e., for w_a, its three positive projections are w_{a0}', w_{a1}', and w_{a2}', with a similar notation for w_b. The input vector $\underline{w}$ to the first six LEDs is formatted in terms of the six positive numbers corresponding to the three projections of each of the two adaptive weights as

$$\underline{w} = [w_{a0}, w_{b0}, w_{a1}, w_{b1}, w_{a2}, w_{b2}]^T . \quad (18)$$

The mask was arranged as described by Eq. (17), the input vector was formatted as described by Eq. (18), and the steering vector was chosen to be

$$\underline{s}^* = [0.3, 0.3]^T \quad (19)$$

for our experiments. The coefficients in the steering vector in Eq. (19) were chosen to simplify the solution. This steering vector corresponds to the boresight direction. Solving Eq. (4) for this case, we find

$$\underline{w} = (0.3\ 0.35) \left[0.6 - 0.1 \exp(-j\pi/3) , \right.$$
$$\left. 0.6 - 0.1 \exp(+j\pi/3)\right]^T \quad (20)$$

In Fig. 3, we show the outputs from the six relevant photodetector elements of concern in the output of Fig. 2 at iterations j = 0, 1, and 5. In Fig. 3, the six photodetector outputs shown correspond *right-to-left* to the six w components in Eq. (18). The six measured output voltages after the sixth iteration were found to be

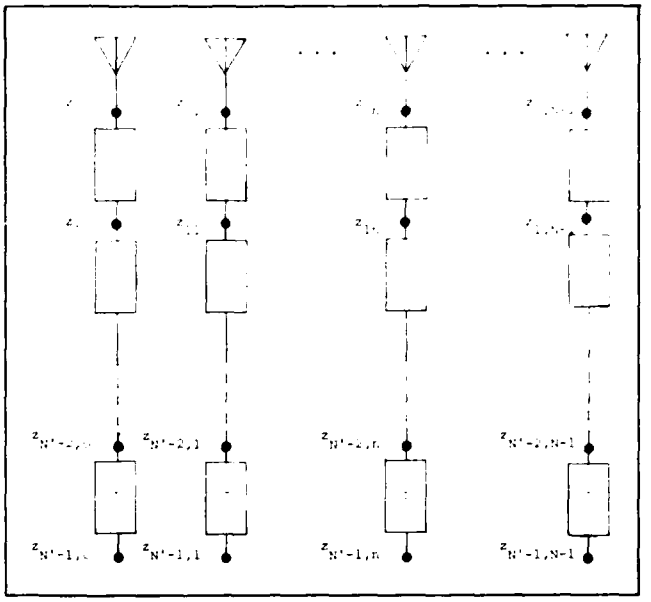


Fig. 3. Pictorial description of the antenna receiver for an array antenna with space and time adaptivity.

$$[0.5118, 0.5118, 0.0834, 0, 0, 0.0834] . \quad (21)$$

These compare very well to the exact results in Eq. (20) after an infinite number of iterations. In the order and format in Eq. (18), these exact results are

$$[0.5142, 0.5142, 0.0857, 0, 0, 0.0857] . \quad (22)$$

Comparing (21) and (22), we find that the IOP's experimentally calculated weights are within about 1% of the exact values. Thus, the performance of this optical processor appears to be excellent for this simple initial example.

5. MULTIDIMENSIONAL ADAPTIVITY

We now consider extending our adaptive antenna theory of Sec. 2 to include adaptivity in velocity or time (further extensions such as polarization and multipath compensation are possible using the techniques to follow, but the details of such formulations are beyond the present scope of this paper). The adaptive weights $\underline{w} = \{w_n\}$ described in Sec. 2 only affect the spatial frequency response of the antenna and hence the angular position of nulls in the antenna pattern. To control the temporal frequency response of the array, we require taps on the time-history outputs from each antenna element. In Fig. 3, we show a 2-D space-time antenna array. There are N antenna elements and N spatial weights as before. These provide adaptivity in space or in angle. To provide temporal frequency control, we include N′ taps on each antenna element with time delays τ between each. We choose τ to satisfy $\tau = \lambda_R / 4v_{max}$, where v_{max} is the blind speed of the radar (i.e., the maximum uniquely resolvable velocity of an object moving relative to the antenna). These provide adaptivity in time or target frequency or velocity. The processor for such an adaptive antenna thus requires the calculation and application of N×N′ weights $w_{n,n'}$ at the $z_{n,n'}$ tap points shown in Fig. 3. We first formulate the required processing as a 2-D extension of the theory of Sec. 2. We then provide an experimental demonstration of the IOP in the solution of such a problem and an analysis of the accuracy of the results obtained (Sec. 6).

We describe this multidimensional adaptive antenna problem by extending our model and analysis of Sec. 2 to include a target or signal at θ_0 with a velocity v_0 and M interference sources at angles θ_m

with M velocities $v_1, \ldots, v_M$. The output at the (n, n') element of the antenna in our 2-D model of Fig. 3 is thus

$$z_{n,n'}(t) = s(t)e^{j[\omega t + \pi n \sin\theta_0 + \pi(4\tau/\lambda_R)n'v_0]} + \sum_{m=1}^{M} r_m(t)e^{j[\omega t + \pi n \sin\theta_m + \pi(4\tau/\lambda_R)n'v_m]} , \quad (23)$$

where $-\pi/2 < \theta < \pi/2$ and $-(4\tau/\lambda) < v < (4\tau/\lambda)$. The output from the receiver is now more complex than in Eq. (2) and is a 2-D summation given by

$$v_{out}(t) = \sum_{n=0}^{(N-1)} \sum_{n'=0}^{(N'-1)} z_{n,n'}(t) w_{n,n'} . \quad (24)$$

The weights are now a 2-D function $w_{n,n'}$. They satisfy the more complex system of linear algebraic equations

$$s^*_{k,k'} = \sum_{n=0}^{(N-1)} \sum_{n'=0}^{(N'-1)} m_{k,k',n,n'} w_{n,n'} , \quad (25)$$

where $m_{k,k',n,n'}$ in Eq. (25) describes the elements of the new covariance matrix $\underline{M}$ for the signals in Eq. (23) and where the steering vector is now

$$s^*_{k,k'} = e^{-j\pi[k \sin\theta_0 + (4\tau/\lambda_R)k'v_0]} . \quad (26)$$

Equation (25) is in the general form of a matrix-matrix equation. Since the IOP (Sec. 3) can perform only matrix-vector multiplications, we must convert Eq. (25) into the form of a matrix-vector equation. We achieve this by performing a lexicographic mapping of the 2-D antenna outputs in Eq. (23) onto a vector $\underline{\tilde{z}}(t) = \{\tilde{z}_l(t)\}$. For a two-element antenna (N = 2) with two taps (N' = 2) per element, we can relate $\underline{\tilde{z}}$ to the elements $z_{n,n'}$ of $\underline{z}$ by

$$\begin{aligned} \tilde{z}_0(t) &= z_{00}(t) \\ \tilde{z}_1(t) &= z_{01}(t) \\ \tilde{z}_2(t) &= z_{10}(t) \\ \tilde{z}_3(t) &= z_{11}(t) . \end{aligned} \quad (27)$$

We describe a new covariance matrix $\underline{\tilde{M}}$ in terms of $\underline{\tilde{z}}$ and a new steering vector $\underline{\tilde{s}}$ that is ordered similar to $\underline{\tilde{z}}$. The resultant weights to be computed are similarly ordered and denoted by $\underline{\tilde{w}}$. With this new notation, we solve the new matrix-vector equation

$$\underline{\tilde{s}}^* = \underline{\tilde{M}}\underline{\tilde{w}} , \quad (28)$$

where $\underline{\tilde{s}}$ and $\underline{\tilde{w}}$ are lexicographically ordered and where $\underline{\tilde{M}}$ is the covariance matrix of the similarly lexicographically ordered received signals $\underline{\tilde{z}}$ in Eq. (27). The solution of Eq. (28) on the IOP of Fig. 2 now follows directly. The space bandwidth product required for the input LEDs, the mask, and the output detector are increased by the lexicographic ordering used. If this becomes prohibitive, one can operate the system successively with one column vector for one of the 2-D functions being the input at successive cycles. This alternate technique for performing matrix-matrix multiplication on a matrix-vector processor was detailed earlier in Ref. 19.

6. MULTIDIMENSIONAL ADAPTIVITY USING THE IOP

In this section, we provide an experimental demonstration of the use of the IOP for multidimensional antenna processing as formulated in Sec. 5. We also emphasize the accuracy of the resultant system with attention to the error source model formulated for the IOP in Sec. 3. We use two performance measures to describe the performance obtained from our adaptive radar processor. Our first performance measure used is the SNR of the resultant antenna pattern. For the spatially-adaptive antenna, the SNR of the output is

$$SNR(j) = \frac{P_0|E(\theta_0, j)|^2}{\sum_{m=1}^{M} P_m|E(\theta_m, j)|^2} , \quad (29)$$

where P_0 is the strength of the signal located at θ_0, and P_m is the strength of the interference source at θ_m. For the spatially adaptive antenna (Sec. 2), the output antenna pattern E is a function of angle and the iteration number j. The numerator in Eq. (29) describes the total power in the antenna pattern at the location θ_0 of the source, and the denominator is the sum of the total power in the antenna pattern at the location of the M noise sources after application of the adaptive weights. For the multidimensional antenna with adaptivity in space θ and time (or velocity v), the output SNR is a function of the iteration number j as well as angle θ and velocity v. It is described by

$$SNR(j) = \frac{P_0|E(\theta_0, v_0, j)|^2}{\sum_{m=1}^{M} P_m|E(\theta_m, v_m, j)|^2} . \quad (30)$$

As our second performance measure, we use the processing gain (PG) defined as

$$PG(j) = SNR(j)/SNR(0) . \quad (31)$$

The denominator in Eq. (31) describes the initial output SNR with no adaptive weighting (i.e., after iteration j = 0). The numerator denotes the SNR that results after j iterations. This PG parameter is thus a measure of the output SNR improvement obtained after j iterations. We expect it to increase with j. It is thus most useful in providing a measure of how various choices of the signal and noise scenario and the acceleration parameter ω affect the speed with which our iterative algorithm achieves convergence or a given performance (i.e., a prespecified antenna pattern SNR). To graphically present our results, we will plot the output antenna pattern obtained for the adaptive weights calculated from the IOP. We also compute the output SNR and PG defined above for each of the resultant antenna patterns as a function of the iteration index j and other system and scenario parameters of concern.

We first consider the effect of the acceleration parameter on the number of iterations required for the algorithm to converge to its steady-state value and on the performance obtained after a given number of iterations j. As our performance measure, we use PG(j) in Eq. (31), where this PG represents the amount by which the various interference sources are nulled by our adaptive algorithm. We found this to be a function of the strength P_m of the interference sources (for a fixed antenna or receiver noise N_r and signal strength P_0). To determine the importance of using Eq. (10) for the acceleration parameter rather than $\omega = 1$ as we used in Eq. (14), we considered various signal powers P_0, interference source powers P_m (we consider only one noise source of power P_I), and antenna or receiver noise powers N_r. In Fig. 4, we highlight our results by plotting PG versus the iteration index j for the two different acceleration parameter measures $\omega = 1$ and $\omega = ||\underline{H}||^{-1}$. In Fig. 4(a), we consider the case when $P_I \gg P_0 = N_r$, and in Fig. 4(b), we consider the case when $P_I = P_0 = N_r$. These data (and much additional testing not included in these drawings) show that SNR increases as in the interference power P_I is increased with a null depth of 40 dB obtained for a noise source of strength $P_I = 0.1$ [Fig. 4(a)] and a much poorer 9 dB null depth obtained for a noise source of lower strength $P_I = 0.001$ [Fig. 4(b)]. This is in agreement with the general performance of an

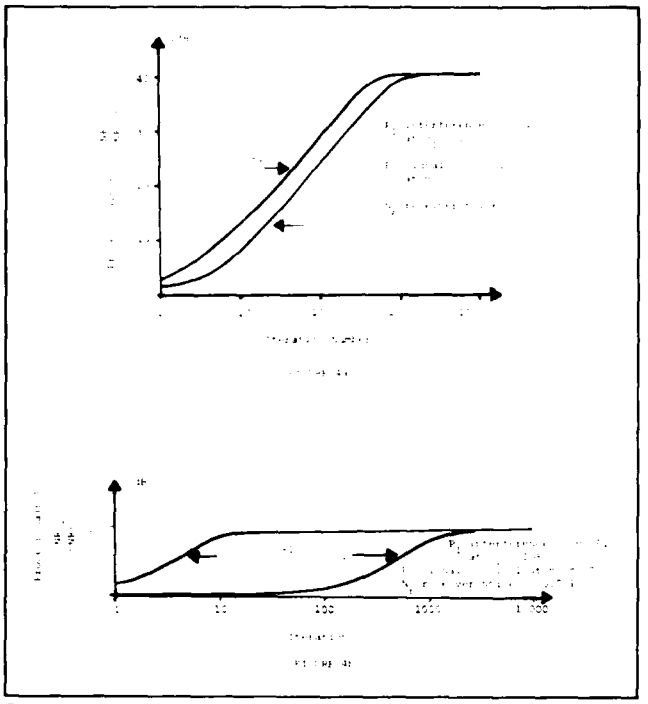


Fig. 4. The processing gain PG(j) = SNR(j)/SNR(0) describing the output SNR improvement obtained with different acceleration parameter choices as a function of the number of iterations j: (a) interference power $p_1 = 0.1$ watts; (b) interference power $p_1 = 0.001$ watts.

adaptive antenna, which provides deeper antenna pattern nulls for stronger interference sources than for weaker ones. From Fig. 4, we also note that the use of our acceleration parameter choice in Eq. (10) becomes increasingly important as the SNR at the input (P_0/P_1) increases and as P_1/N_r increases. This is in agreement with standard adaptive array antenna theory.[1-5] From Fig. 4(b), we notice that our adaptive algorithm converges to the final value in about ten iterations when the acceleration parameter is chosen according to Eq. (10), whereas over 100 times more iterations are needed if no acceleration parameter ($\omega = 1$) is used. Thus, to accommodate all possible ratios of signal, interference, and receiver noise, the use of Eq. (10) for the acceleration parameter is warranted.

In other tests, we studied how the processing gain varied with the angular separation between the signal and the interference source. We verified that our system could achieve super-resolution beyond the classical resolution limit as described further in Ref. 20. When the number of interference sources M is larger than the number of adaptive elements, we found that choosing the locations of the adaptive elements to be randomly distributed on the N×N′ grid in Fig. 3 improves performance very well.

One of our most important theoretical analysis and simulation results concerns the effect of the system's spatial Δb and temporal Δt errors on SN. of the output. We considered an N = 5 element antenna with receiver noise $N_r = 0.1$ and one interference source at $\theta_1 = 45°$ with $P_1 = 1.0$. In Fig. 5, we show SNR(j) as a function of the iteration index j for different Δb and Δt percent errors and noise. With no errors, a steady-state SNR of 44 dB was achieved after about 100 iterations. With a 2.5% spatial error and an 0.5% temporal error ($\Delta b = \pm 0.025$, $\Delta t = \pm 0.005$), we find less SNR than the ideal system can provide, but the SNR is still a very respectable 38 dB value after only 100 iterations. For increased spatial and temporal errors, the system's SNR performance is degraded even worse. The error values included in Fig. 5 are comparable to what the present laboratory IOP system can achieve, and, as seen, its performance is quite acceptable. To obtain the data in Fig. 5, we employed a random number generator with a uniform density function to produce mask errors and detector noise with the three-sigma variance value shown. Different sample realizations of each of these errors were added on each iteration of the IOP. The data in Fig. 5 are the average of five Monte Carlo runs.

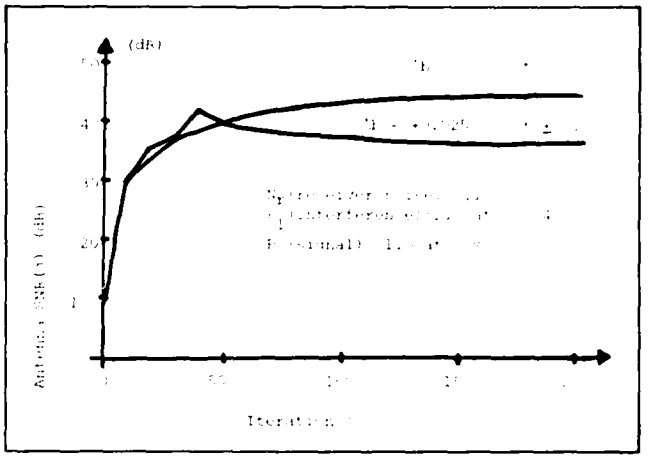


Fig. 5. Output antenna pattern SNR(j) as a function of the number of iterations j for no IOP system errors ($\Delta b = \Delta t = 0$) and for typical experimental IOP errors ($\Delta b = 0.025$, $\Delta t = 0.005$).

We conclude this section with a detailed description of a typical multidimensional adaptive antenna processing experiment performed on the laboratory IOP. For this case, we consider an N = 2 element array with N′ = 2 time taps. The signal source was of power $P_0 = 0.1$ and located at $\theta_0 = 45°$ and $v_0 = 0.5 v_{max}$. We used one interference source at $\theta_0 = 0°$ and $v_1 = 0$ with power 1.0. The N×N′ = 4 array elements each with receiver noise power $N_r = 1.0$ were lexicographically ordered as described in Eq. (27). The covariance matrix computed from the received signals for this scenario was

$$\underline{\tilde{M}} \simeq \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} . \tag{32}$$

The Euclidean norm of $\underline{H}$ calculated from Eq. (32) is 7.48. We used its reciprocal as our acceleration parameter $\omega = 0.13$ as described in Eq. (10). The complex-valued format in Eq. (8) was then used for $\underline{M}$.

For this case, $\underline{M}$ is real and thus it is arranged as the 8×8 matrix

$$\underline{\tilde{M}} = \begin{bmatrix} \underline{\tilde{M}}_{re} & \underline{0} \\ \underline{0} & \underline{\tilde{M}}_{re} \end{bmatrix} , \tag{33}$$

where $\underline{\tilde{M}}_{re}$ is described by Eq. (32). To obtain the optical mask used in the actual system, we divided each element of $\underline{M}$ by $(h_{max} - h_{min}) = 2$ and biased the entire matrix by $h_{min}/(h_{max} - h_{min}) = 0$. The resultant optical mask actually placed at P_2 of Fig. 2 was thus

$$\underline{H} = \begin{bmatrix} \underline{H}_{re} & 0 \\ 0 & \underline{H}_{re} \end{bmatrix} . \tag{34}$$

Where $\underline{H}_{re}$ is the same $\underline{\tilde{M}}_{re}$ in Eq. (32) with each element divided by two. The complex-valued steering vector corresponding to the signal direction $\theta_0 = 45°$ and $v_0 = 0.5$ has element values given by Eq. (26). When arranged in the lexicographic format of Eq. (27), it becomes

$$\underline{\tilde{s}}^{*} = \begin{bmatrix} -0.82 - 0.61j \\ 0.97 - 0.27j \\ 0.61 - 0.82j \\ 0.27 + 0.97j \end{bmatrix} . \quad (35)$$

We easily decompose Eq. (35) into its real and imaginary parts and obtain the eight element exogenous vector

$$\underline{y} = [\underline{s}^{*}_{re}\ \underline{s}^{*}_{im}]^{T} = [-0.82, 0.97, 0.61, 0.27, -0.61, -0.27, -0.82, 0.97]^{T} . \quad (36)$$

The two cycle complex algorithm described in Sec. 3 was employed with the positive elements of y used on odd iterations and the negative elements on even iterations. The $\underline{x}$ outputs were computed by the laboratory IOP for the first fifty iterations. The microprocessor support system and the dedicated high speed memory in the IOP were used to combine the positive and negative outputs from successive iterations and to store the resultant bipolar numbers $\underline{w}(j)$ computed at each iteration j. The eight relevant photodetector outputs corresponding to the eight elements of the bipolar and complex-valued output vector $\underline{x} = [\underline{x}^{*}_{re}\ \underline{x}^{*}_{im}]$ corresponding to the complex-valued weights $\underline{w}$ are shown in Fig. 6 after the first, fifth, and fiftieth iterations. We denote these outputs by $\underline{x}(1)$, $\underline{x}(5)$, and $\underline{x}(50)$, respectively, in the Fig. 6 caption. The complex-valued weights

$$\underline{w} = \underline{w}_{re} + j\underline{w}_{im} = [w_{00}, w_{10}, w_{01}, w_{11}]^{T} \quad (37)$$

were directly obtained from the eight $\underline{x}$ outputs ($\underline{x}_1, \ldots, \underline{x}_8$) according to

$$\begin{aligned} W_{00} &= x_1 + jx_5 \\ W_{10} &= x_2 + jx_6 \\ W_{01} &= x_3 + jx_7 \\ W_{11} &= x_4 + jx_8 \end{aligned} \quad (38)$$

After the fiftieth iteration, we obtained

$$\underline{w}(50) = \begin{bmatrix} -0.9 - 0.45j \\ 0.75 - 0.25j \\ 0.4 - 0.8\,j \\ 0.2 + 1.1\,j \end{bmatrix} . \quad (39)$$

To determine the accuracy of these results, we first calculated the rms errors between the exact weights and those computed after 50 iterations. This error was found to be 2.3% (it did not decrease appreciably when further iterations were performed). The true measure of the performance accuracy of the weights computed from the laboratory IOP lies in the SNR obtained in the output pattern that results when the weights in Eq. (39) are applied with the interference sources and receiver noise indicated. The resultant antenna pattern was obtained. Its SNR was 14.7 dB. This is nearly exactly equal to the SNR obtained (14.96 dB) if the exact weights were applied. We thus find the laboratory IOP system to be extremely accurate with less than 0.26 dB difference in the SNR of the output antenna for the cases of the optically computed weights and the exact weights. Initial simulations were performed to verify that these results were typical of those to be expected. The above results were found to be typical for the ten different cases we considered. We also produced initial theoretical expressions[21] from which upper bounds on the performance of the IOP can be predicted as a function of spatial and temporal system errors. In all cases, the results obtained in our experiments were well below the weak bounds we derived. Because of this, no derivation of these bounds is included at present.

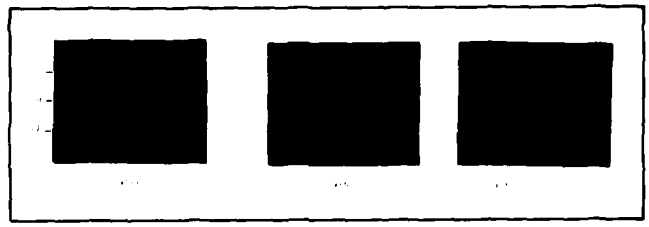


Fig. 6. Experimental outputs from the relevant eight photodetectors of the IOP of Fig. 2 in the computation of the complex-valued weights for a multidimensional antenna with space and time adaptivity.

7. CONCLUSION

We have reviewed the basic signal processing requirements for adaptive antennas and have provided a summary description of an iterative optical matrix-vector processor that appears most attractive for such advanced signal processing applications. Modifications to our initial IOP were described to allow incorporation of an acceleration parameter, and two techniques were described to allow the system to operate on complex-valued data (as required for the APAR application). Theoretical and experimental data and simulations showed that use of an acceleration parameter equal to the reciprocal of the Euclidean norm of the covariance matrix greatly reduced the number of iterations needed (by a factor of 100 or more), especially as the interference power approaches the signal power and the receiver noise. We described and experimentally demonstrated two different techniques by which the system can operate on complex-valued data. The technique (Sec. 4) in which each complex number is represented by its three positive projections on three axes in the complex plane requires more space bandwidth product with fewer iterations. The bipolar technique (Secs. 3 and 6) in which the positive- and negative-valued input data are used on successive iterations requires twice the number of iterations but less space bandwidth product than the technique used in Sec. 4. The choice between these two methods of handling complex-valued data depends upon the number of adaptive weights, the speed required, the space bandwidth product available on the IOP used, and the importance of canceling fixed-pattern detector noise.

We have also extended the use of the system to include multidimensional adaptivity and have experimentally demonstrated angular adaptivity and multidimensional space and time adaptivity on our laboratory IOP. The experimental performance obtained was quite excellent. Theoretical and simulation studies have shown that the performance of the present IOP with its 0.8% spatial errors and 0.4% temporal errors is quite adequate for APAR applications. Our experiments showed an rms error of only 2.3% in the computed weights, and, more important, that this resulted in less than an 0.26 dB difference in the SNR of the output antenna pattern. This novel and most general purpose optical processing architecture merits more research and analysis for the indicated APAR problem and for many other diverse applications that can be reduced to matrix-vector equations and matrix-inversion problems.

8. ACKNOWLEDGMENT

The support of the Air Force Office of Scientific Research on a supplement to Grant AFOSR-79-0091 which enabled us to complete the work herein is gratefully acknowledged.

9. REFERENCES

1. B. Widrow, P. Mantey, L. Griffiths, and B. Goode, Proc. IEEE 55(12), 2143(1967).
2. S. Applebaum, IEEE Trans. Antennas Propag. AP-24 (5), 585(1976).
3. L. Brennan and L. Reed, IEEE Trans. Aerospace and Electronic Systems AES-9 (2), 237(1973).
4. W. Gabriel, Proc. IEEE 64(2), 239(1976).
5. Special Issue on Adaptive Antennas, IEEE Trans. Antennas Propag. AP-24 (1976).
6. D. Casasent, IEEE Commun., 40 (Sep. 1981).
7. M. Carlotto and D. Casasent, Appl. Opt. 21, 147 (Jan. 1, 1982).
8. D. Psaltis, D. Casasent, and M. Carlotto, Proc. SPIE 180, 114(1979).
9. D. Casasent and M. Carlotto, "A Novel Optical Processor for Adaptive Phased Array Radar," Proc. SPIE 341 (May 1982).
10. D. Psaltis, D. Casasent, and M. Carlotto, Opt. Lett. 4, 348 (Nov. 1979).
11. A. Edison and M. Noble, "Optical Analog Matrix Processors," AD646060 (Nov. 1966).
12. P. Mengert et al., U.S. Patent 3,525,856 (Oct. 6, 1966).
13. J. Goodman, A. Dias, and L. Woody, Opt. Lett. 2, 1(1978).
14. L. Richardson, *Philos. Trans. R. Soc. London, Ser. A210,* 307(1910).
15. J. Grinberg, W. P. Bleha, P. O. Braatz, K. Chow, D. H. Close, A. D. Jacobson, M. J. Little, N. Massetti, R. J. Murphy, J. G. Nash, and M. Waldner, Proc. SPIE 128, 253(1977).
16. H. Caulfield et al., Opt. Commun. 40, 86(1981).
17. D. Casasent, Appl. Opt. 21, 1859(1982).
18. J. Goodman and L. Woody, Appl. Opt. 16, 2611(1977).
19. D. Casasent and C. Neuman, "Operations Achievable on an Iterative Optical Processor," Proc. NASA Conference on Applications of Optical Processing to Aerospace Needs (Aug. 1981).
20. G. Borgiotti and L. Kaplan, IEEE Trans. Antennas Propag. AP-27 (6), 842(1979).
21. M. Carlotto, "Iterative Electro-Optic Matrix Processor," PhD Thesis, Carnegie-Mellon University (April 1981).

13. EIGENVECTOR DETERMINATION BY ITERATIVE OPTICAL METHODS

Reprinted from APPLIED OPTICS, Vol. 20, page 3707, November 1, 1981

Eigenvector determination by iterative optical methods

B. V. K. Vijaya Kumar and D. Casasent

Three power methods to compute the eigenvalues and eigenvectors of a matrix on an iterative optical processor (IOP) are analyzed. Each is appropriate for a different eigenvalue and eigenvector application. When implementation on an IOP and the processing speed are considered, the second method is found to be preferable to others recently described.

I. Introduction

A noncoherent optical vector–matrix processor[1] and an iterative optical processor (IOP)[2] version of this system have recently been described. The IOP system enables one to solve sets of simultaneous linear equations at very high computations/second rates. This also represents a general purpose optical processor, since many problems can and are easily formulated as vector–matrix equations. In this paper we describe another operation, computation of the eigenvalues and eigenvectors of a symmetric matrix, that the IOP can realize easily.

In Sec. II we briefly review the basic IOP system and describe our initial[3] use of it to compute the dominant eigenvalue and eigenvector of a matrix that we need to determine the acceleration parameter used in our iterative algorithm. A recent publication[4] has detailed a similar power-law algorithm[5] and has discussed its use on an IOP to compute all the eigenvectors and eigenvalues of a matrix. Our concern in this paper is thus with implementational problems associated with the power-law algorithm[5] and the use of the IOP in three different eigenvalue and eigenvector computational problems. These applications are: calculation of the largest eigenvalue (Sec. II); calculation of the eigenvectors in order of decreasingly dominant eigenvalues (Sec. III); and computation of the eigenvector whose eigenvalue is closest to a given value (Sec. IV). In all three problem different modifications of the basic power method are preferable from implementation considerations. Our summary and concluding remarks follow in Sec. V.

Methods to accommodate biopolar[3,4,6] and complex-valued[2,3,6] vector and matrix elements in an IOP have been previously described and thus are not discussed explicitly in this paper, as all the techniques are equally appropriate for any of the algorithms to be described.

II. Iterative Optical Processor and Power Method

A new schematic diagram of the IOP, including the exact iterative algorithm we employ and the acceleration parameter ω, is shown in Fig. 1. The input vector data are described by $\mathbf{x}(k)$. It is provided as the spatially varying intensity output from a linear array of LEDs or laser diodes (LDs). The output from these input data at P_1 is imaged vertically and expanded horizontally to illuminate uniformly a matrix mask $\mathbf{H}$ at P_2. In the present IOP system this is achieved by fiber optics. The transmittance of each element at P_2 is proportional to each element h_{ij} of $\mathbf{H}$. In the final version of the IOP a CCD-addressed spatial light modulator (SLM) will be used at P_2 (see Ref. 7). The light leaving each column of P_2 is integrated vertically and imaged horizontally onto a linear photodetector array at P_3 with parallel readout. The mask at P_2 is presently a fixed photographic film, and the required imaging from P_2 to P_3 is achieved by placing the detector array in direct contact with the P_2 mask. The

The authors are with Carnegie-Mellon University, Department of Electrical Engineering, Pittsburgh, Pennsylvania 15213.
Received 20 June 1981.
0003-6935/81/213707-04$00.50/0.

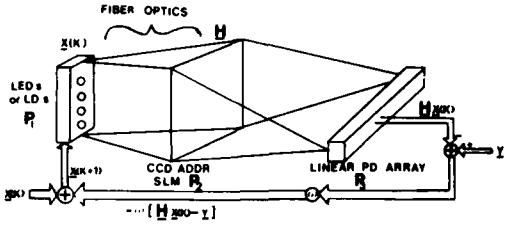


Fig. 1. Schematic diagram of the IOP.

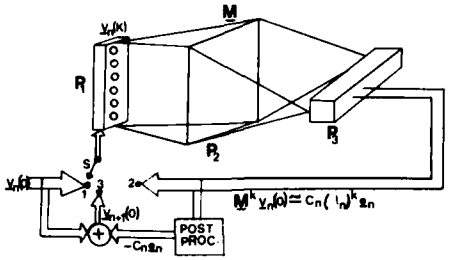


Fig. 2. Schematic diagram of an eigenvector and eigenvalue computational system using the IOP.

height of each detector is chosen to equal the height of the P_2 mask (3 mm) to achieve this.

We can thus describe the output from P_3 at iteration k as $\mathbf{Hx}(k)$, i.e., by the vector–matrix product of the P_1 and P_2 data. We subtract this output from an external vector $\mathbf{y}$ and multiply the difference by an acceleration factor ω. The result is then added to the prior $\mathbf{x}(k)$ input at iteration k to produce a new $\mathbf{x}(k+1)$ input vector for iteration $k+1$. The resultant system thus realizes the iterative algorithm

$$\mathbf{x}(k+1) = \mathbf{x}(k) - \omega[\mathbf{Hx}(k) - \mathbf{y}]. \tag{1}$$

When $\mathbf{x}(k+1) = \mathbf{x}(k)$, Eq. (1) reduces to

$$\mathbf{Hx} = y. \tag{2}$$

The IOP system of Fig. 1 thus solves Eq. (2) for the unknown vector $\mathbf{x}$ given $\mathbf{H}$ and $\mathbf{y}$. It thus effectively computes

$$\mathbf{x} = \mathbf{H}^{-1}\mathbf{y}. \tag{3}$$

To optimize convergence of the algorithm in Eq. (1), we select ω to be λ_{max} (the maximum eigenvalue of $\mathbf{H}$).[3,5] We have used two techniques to compute λ_{max}. The first is to determine rapidly an upper bound for λ_{max} from the norm (square root of the sum of the squares of the elements of $\mathbf{H}$),[8]

$$\lambda_{max} < \sqrt{\sum_{ij} h_{ij}^2}. \tag{4}$$

We have also used the power method[5] to compute λ_{max}. To describe this method refer to Fig. 2. If the switch is in position 1 (initialize), the original P_1 input vector $\mathbf{x}(k)$ at the first $k = 0$ iteration $\mathbf{v}_1(0)$ [$\mathbf{v}_n(0)$ for $n = 1$] is loaded into the source array at P_1. After initialization the switch is placed in position 2. On the mask $\mathbf{H}$ at P_2 we place $\mathbf{M}$, the matrix whose λ_{max} is desired. Then the system's output $\mathbf{v}_1(k)$ at P_2 after k iterations is

$$\mathbf{M}^k\mathbf{V}_1(0) = \mathbf{v}_1(k). \tag{5}$$

If the original input $\mathbf{v}_1(0)$ vector is described by

$$\mathbf{v}_1(0) = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots c_N\mathbf{e}_N, \tag{6}$$

after k iterations the P_3 output is

$$\begin{aligned}\mathbf{v}_1(k) &= \mathbf{M}^k\mathbf{v}_1(0)\\ &= \mathbf{M}^k(c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots c_N\mathbf{e}_N)\\ &= c_1(\lambda_1)^k\mathbf{e}_1 + c_2(\lambda_2)^k\mathbf{e}_2 + \ldots c_N(\lambda_N)^k\mathbf{e}_N\\ &\simeq c_1(\lambda_1)^k\mathbf{e}_1.\end{aligned} \tag{7}$$

In applying the IOP as described by Eq.(1) to the computation of λ_{max}, we denote the N eigenvectors by $\mathbf{M}$ by $\mathbf{e}_n$ and their eigenvalues by λ_n. Thus the third equation in Eq. (7) results. If the λ_n is ordered as $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_N|$, after a sufficient number of k iterations the sum in line 3 of Eq. (7) is dominated by the first term and the last expression in Eq. (7) results.

Thus with any arbitrary $\mathbf{v}_1(0)$ input vector at P_1, with the condition that the c_n's are not zero (we have achieved this condition by using a random number generator to produce the initial set of c_n values,[3] and with $\mathbf{y} = 0$, the P_3 output after a sufficiently large

number (k) of iterations will yield $c_1(\lambda_1)^k\mathbf{e}_1$. After one more iteration we obtain $c_1(\lambda_1)^{k+1}\mathbf{e}_1$. The quotient of these two outputs yields λ_1. From our assumptions on the ordering of the eigenvalues this is an approximation to $\lambda_1 = \lambda_{max}$ as shown in Eq. (7).

Caulfield *et al.*[4] have noted this use of the power method on the IOP to compute $\lambda_1 = \lambda_{max}$. They also extended this algorithm to allow computation of all N eigenvectors and eigenvalues of $\mathbf{M}$. The basic algorithm suggested in Ref. 4 to achieve this is to form a new matrix after each eigenvector and eigenvalue has been computed. The general form used for the mask after $n-1$ eigenvectors and eigenvalues have been obtained is

$$\mathbf{M}_n = (\mathbf{M} - \lambda_1\mathbf{I})(\mathbf{M} - \lambda_2\mathbf{I}) \ldots (\mathbf{M} - \lambda_{n-1}\mathbf{I}), \tag{8}$$

where $\mathbf{I}$ is the identity matrix. This new $\mathbf{M}_n$ matrix can then be placed at P_2 of Fig. 1, $\mathbf{v}_1(0)$ can be placed at P_1, and after k iterations the output at P_3 provides the largest eigenvalue and its eigenvector of $\mathbf{M}_n$ as before. This occurs because the eigenvectors of $\mathbf{M}$ and $\mathbf{M}_n$ are the same. For the mth eigenvector $\mathbf{e}_m$ we can show

$$\mathbf{M}_n\mathbf{e}_m = (\lambda_m - \lambda_1)(\lambda_m - \lambda_2) \ldots (\lambda_m - \lambda_{n-1})\mathbf{e}_m, \tag{9}$$

which is equivalent to Eq. (21) in Ref. 4. Since $n-1$ of the eigenvalues of the new $\mathbf{M}_n$ matrix is zero, the new λ_n eigenvalue can be found from Eq. (9) by postprocessing, given the prior λ_1 to λ_{n-1} eigenvalues obtained from $n-1$ prior operations of the system as described in Ref. 4. In Secs. III and IV we describe preferable eigenvalue and eigenvector computational algorithms for realization on the IOP for selected applications that are also preferable from implementation, speed of computation, and accuracy considerations.

III. Computation of Ordered Dominant Eigenvalues and Eigenvectors

In many cases one desires only the first two or three eigenvectors of $\mathbf{M}$ with the largest eigenvalues. This need arises in Karhunen-Loeve analysis,[9] bandwidth compression,[10] pattern recognition,[11] and other cases. In these instances we would like to compute the eigenvectors $\mathbf{e}_n$ of $\mathbf{M}$ orderd in terms of their largest eigenvalues (absolute values). The method[4] of Sec. II does not provide this ordered set of eigenvectors. Rather, it provides the eigenvector with the largest eigenvalue of the modified matrix $\mathbf{M}_n$. This is not necessarily the eigenvector of $\mathbf{M}$ with the next largest eigenvalue. Thus, to avoid computing all N eigenvalues and eigenvectors and then ordering them to determine the two or three most dominant ones, we suggest a modified algorithm.

A second reason for considering an alternate algorithm is that solving the nth-order polynomial equations of the form in Eq. (9) (as is required for each eigenvector) can become as computationally difficult as computing the eigenvectors of $\mathbf{M}$ by conventional methods. As a third motivational reason to consider an alternate eigenvector computational algorithm, we note that the prior power method (Sec. II) requires a new matrix mask, as described by Eq. (8), to be computed and generated for each new eigenvector $\mathbf{e}_n$. This computation requires n matrix–matrix multiplications and is thus prone to computational errors. Also, it requires a new P_2 mask $\mathbf{M}_n$ to be generated and hence a real-time electronically addressed light modulator at P_2. The most advanced device of this type is a CCD-addressed liquid crystal light valve.[7] This device and other 2-D light modulators can presently be cycled only once every 30 msec. This can represent a considerable decrease in the overall computational speed of the IOP for eigenvector calculations.

For these reasons an algorithm permitting the use of a fixed mask and requiring less complex postprocessing is preferable for many applications. The modified algorithm uses the original IOP concept that it is far easier to alter the photodetector's output than the matrix mask. We thus enter $\mathbf{v}_1(0)$ and $\mathbf{M}$ at P_1 and P_2 and estimate λ_1 and $\mathbf{e}_1$ as before with $\mathbf{y} = 0$ in Fig. 1. The P_3 output (after k iterations) is $c_1(\lambda_1)^k\mathbf{e}_1$. We determine λ_1 from the ratio of the outputs at the k and $k+1$ iterations as noted in Sec. II. We then compute the elements of $\mathbf{e}_1$ by dividing each measured photodetector output value by the square root of the sum of the squares of the photodiode outputs. This yields the elements of $\mathbf{e}_1$ and simultaneously normalizes $\mathbf{e}_1$. Then we can compute c_1 (again in dedicated output logic) by

$$c_1 = [\mathbf{v}_1(0) \cdot \mathbf{e}_1], \tag{10}$$

where $\cdot$ in Eq. (10) denotes the vector inner product operation.

With λ_1 and $\mathbf{e}_1$ determined by Eq. (7), we then subtract $c_1\mathbf{e}_1$ from the original $\mathbf{v}_1(0)$ input vector for iteration $k = 0$ for the first or $n = 1$ eigenvector. This yields a new initial (iteration $k = 0$) starting vector for computation of the $n =$ second eigenvector, given by

$$\mathbf{v}_2(0) = \mathbf{v}_1(0) - c_1\mathbf{e}_1. \tag{11}$$

Substituting Eq. (6) into Eq. (11) we find that

$$\mathbf{v}_2(0) = c_2\mathbf{e}_2 + c_3\mathbf{e}_3 + \ldots + c_N\mathbf{e}_N. \tag{12}$$

After k iterative cycles with the same $\mathbf{M}$ at P_2, and with the P_1 input described by Eq. (12), the P_3 output is

$$\mathbf{M}^k\mathbf{v}_2(0) = \mathbf{v}_2(k) \simeq c_2(\lambda_2)^k\mathbf{e}_2. \tag{13}$$

As before, we compute λ_2, $\mathbf{e}_2$, and then c_2 from

$$c_2 = [\mathbf{v}_2(0) \cdot \mathbf{e}_2]. \tag{14}$$

At the next set of k iterations we compute c_3 and $\mathbf{e}_3$ using the input vector

$$\mathbf{v}_3(0) = \mathbf{v}_2(0) - c_2\mathbf{e}_2. \tag{15}$$

Inspection of Eqs. (11), (12), and (15) shows that the required postprocessing to obtain the initial starting vector $\mathbf{v}_{n+1}(0)$ for computation of the $(n+1)$st eigenvalue and eigenvector simply requires subtraction of $c_n\mathbf{e}_n$ from the previous initial vector $\mathbf{v}_n(0)$. This is easily achieved as shown by switch position 3 in Fig. 2. We denote the iteration number by the index k and the eigenvector being computed by the subscript n. The three steps in this algorithm can be realized by a three-position switch as shown in Fig. 2. The opera-

tions performed by the system with the switch in each of the three positions can be summarized as below:

(1) *Initialize.* In this case $\mathbf{v}_n(0) = \mathbf{v}_1(0)$.

(2) *Iterate.* After k iterative cycles, the P_3 output (for $n = 1$) is

$$\mathbf{M}^k\mathbf{v}_1(0) = \mathbf{v}_1(k) \simeq c_1(\lambda_1)^k\mathbf{e}_1, \tag{16}$$

from which c_1 and $\mathbf{e}_1$ can be found by direct postprocessing.

(3) *Reinitialize.* The starting vector $\mathbf{v}_n(0)$ for computing the nth eigenvector is determined from the initial input $\mathbf{v}_{n-1}(0)$ used in computing λ_{n-1} and $\mathbf{e}_{n-1}$ and the $c_{n-1}\mathbf{e}_{n-1}$ product by

$$\mathbf{v}_n(0) = \mathbf{v}_{n-1}(0) - c_{n-1}\mathbf{e}_{n-1}. \tag{17}$$

The system's output in mode 3 after k iterations is

$$\mathbf{M}^k\mathbf{v}_n(0) = \mathbf{v}_n(k) \simeq c_n(\lambda_n)^k\mathbf{e}_n. \tag{18}$$

Equations (16)–(18) thus describe the general reinitialization and iterative algorithms for the use of the IOP in computing the ordered eigenvectors of a matrix **M**. As seen, the required postprocessing in Eqs. (10) and (11) is trivial compared with that in Eq. (9). In addition, the matrix mask at P_2 need not be changed as was required in Eq. (8). Furthermore, this new algorithm provides the eigenvectors of **M** ordered in terms of their decreasing eigenvalues (absolute values), whereas the prior power method does not. It is thus superior for the application indicated in terms of complexity of the postprocessing required, its speed (no P_2 mask changes are needed), and accuracy ($n - 1$ matrix multiplications are not needed to produce the new $\mathbf{M}_n$ mask).

IV. Determining an Arbitrary Eigenvector

The power method in Sec. II produces all eigenvectors of a matrix **M**, whereas the method in Sec. III produces the eigenvectors ordered in terms of their eigenvalues. However, it is often desired to determine the eigenvalue closest to a given value. This problem arises in physics[12] and in control applications. The former case is discussed by Gamba.[12] The latter case arises when a given RC time constant response is desired from a system. We can also use the IOP for this application by modifying the algorithm in Sec. III as described in Ref. 12. This is preferable to use of the technique in Sec. III since its cumulative errors can become excessive.

The algorithm we use is quite simple. We create a new matrix **M′** from **M**, where

$$\mathbf{M}' = (\mathbf{M} - \lambda\mathbf{I})^2 - (|\lambda| + K)^2\mathbf{I}. \tag{19}$$

In Eq. (19), λ is the eigenvalue whose eigenvector we desire, and K is a constant (that must be larger than the largest eigenvalue λ_1 of **M**). We can use Eq. (4) or the power method of Sec. II to estimate K. It is easy to show that **M′** and **M** have the same eigenvectors. Their eigenvalues differ, however. For the nth eigenvalue λ_n and eigenvector $\mathbf{e}_n$ we find that

$$\mathbf{M}'\mathbf{e}_n = (\mathbf{M} - \lambda\mathbf{I})^2\mathbf{e}_n - (|\lambda| + K)^2\mathbf{I}\mathbf{e}_n$$
$$= [(\lambda_n - \lambda)^2 - (|\lambda| + K)^2]\mathbf{e}_n. \tag{20}$$

Since K is larger than any $|\lambda_n|$, the term in brackets in Eq. (20) has a negative maximum when $\lambda_n \simeq \lambda$. Thus, since the power method yields the eigenvector with the largest absolute eigenvalue, we can use the IOP with **M′** at P_2 and the input initial vector in Eq. (6). The output at P_3 will then yield data for the eigenvector **e** with the eigenvalue closest to λ. A simple numerical example can be found in Ref. 12. We note in conclusion that computing **M′** in Eq. (19) from **M** requires only two matrix multiplications.

V. Summary and Conclusions

In this brief paper we have discussed three power algorithms by which the IOP can be used to compute the eigenvalues and eigenvectors of a matrix. We found the first method to be adequate for estimating the maximum eigenvalue. We suggested a new scheme (Sec. III) for computing the dominant-ordered eigenvectors that did not require altering the mask in the IOP. The second method requires less electronic processing and thus should be more accurate than the technique in Sec. II. However, the accuracy with which either technique can identify a nondominant eigenvector may not be adequate. We thus advanced a third power algorithm (Sec. IV) that can compute the eigenvector whose eigenvalue is closest to a given value.

These applications of the IOP serve to show how general purpose this IOP processing system architecture is for a diverse selection of applications in different disciplines.

We thank the Air Force Office of Scientific Research (grant 79-0091) for support of such extended applications of the IOP. We also thank NASA Lewis (grant NAG 3-5) for recent support of our continued IOP program.

References

1. J. W. Goodman, A. Dias, and L. Woody, Opt. Lett. **2,** 1 (1978).
2. D. Psaltis,D. Casasent, and M. Carlotto, Opt. Lett. **4,** 348 (1979).
3. M. Carlotto, Iterative Electro-Optic Matrix Processor, Ph.D. Thesis, Carnegie-Mellon U., Pittsburgh, Penn. (April 1981).
4. H. J. Caulfield, D. Dvore, J.W. Goodman, and W. Rhodes, Appl. Opt. **20,** 2263 (1981).
5. H. R. Schwartz, H. Rutishauser, and E. Stiefel, *Numerical Analysis of Symmetric Matrices*, translated by P. Hertlendy (Prentice-Hall, Englewood Cliffs, N. J., 1973), pp 188–193.
6. J. W. Goodman and L. M. Woody, Appl. Opt. **16,** 2611 (1977); see also J. Goodman *et al.*, BMD Technical Report. L-723-1 (1979).
7. J. Grinberg *et al.*, Proc. Soc. Photo-Opt. Instrum. Eng. **128,** 253 (1977).
8. E. Kreyszig, *Advanced Engineering Mathematics* (Wiley, New York, 1972), p. 686.
9. Y. T. Chien and K. S. Fu, IEEE Trans. Inf. Theory **IT-13,** 518 (1967).
10. A. Rosenfield and A. C. Kak, *Digital Picture Processing* (Academic, New York, 1976).
11. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic, New York, 1972).
12. A. Gamba, Am. J. Phys. **49,** 187 (1981).

14. A REVIEW OF OPTICAL SIGNAL PROCESSING

A REVIEW OF OPTICAL SIGNAL PROCESSING

DAVID CASASENT

The present, the future, and the potential of this discipline.

BULK acousto-optic and other optical processors for spectrum analysis, correlation, ambiguity function computation, and other signal processing functions are described. Basic architectures and algorithms are reviewed and several case studies are described. These include spread spectrum, wide-band signal processors, radar ambiguity function processors, passive ambiguity function computation, and adaptive phased array radar processing.

Advanced radar, sonar, and communication systems require increased processing capacity to keep abreast of the computation load produced by the larger bandwidth and longer time-bandwidth products of new sophisticated waveforms, as well as the denser target and signal environments to be handled. This paper contains a review of recent advances in optical signal processing for such applications. Major attention is given to bulk acousto-optic (AO) systems, since they have recently [1], [2] emerged as a major component with the necessary high bandwidth and large time-bandwidth product for advanced signal processing needs. This has occurred because of parallel advances in devices, algorithms, and system architectures. The repertoire of basic AO processor architectures and algorithms are briefly reviewed in Section I. Wide-band optical signal processing systems and techniques for folded-spectrum analysis are then described in Section II. Several spread spectrum OSP techniques for synchronization and decoding are treated in Section III. These involve new signal processors such as Spann correlators and hybrid time and space integrating and space-variant AO systems. Simultaneous determination of the range and Doppler of multiple targets represents one of the most demanding radar processing functions. In Section IV, two optical processors that compute the ambiguity function are described, one for active radar applications and the second for passive signal processing use. Adaptive phased array radar signal processing applications are the subject of Section V. Major attention is given to an iterative optical vector-matrix processor that represents a new class of an optical system of quite general use for any problem that can be formulated as a vector-matrix equation or a set of simultaneous linear equations. Concluding and summary remarks are then advanced in Section VI. Throughout the various sections, new architectures and signal processing techniques as well as complete optical systems that have been fabricated and delivered to several agencies are noted. This combination of new research techniques and optical system engineering has been selected to demonstrate the state of the art of optical signal processing, future research directions for this technology, and the quite major potential this discipline has to offer.

BASIC TECHNIQUES

The basic operation of AO transducers [2] can be briefly reviewed with attention to Fig. 1. An electrical signal of amplitude A_1 and frequency ω_1 is fed to the AO cell. We can describe the output pattern from this transducer in three ways. We first note that the angle $\theta_i + \theta_d$ at which the diffracted light leaves the cell is proportional to the frequency ω_1 of the input signal, and that the amplitude of the diffracted light is proportional to the amplitude A_1 of the input signal. Optical systems are linear. Thus, if multiple input signals are present at frequencies ω_n and with amplitudes A_n, multiple diffracted plane waves leave the cell at angles proportional to ω_n and with amplitudes proportional to A_n. A lens placed behind the cell focuses these plane waves to different spatial locations in

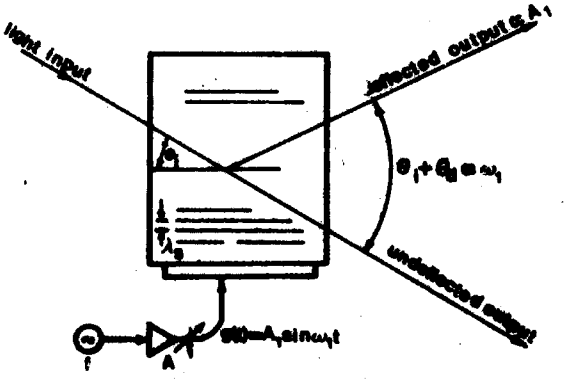


Fig. 1. Schematic diagram showing the operation of an acousto-optic transducer.

0163-6804/81/0900-0040 $00.75 © 1981 IEEE

the back focal plane of the lens, with the amplitude of each component proportional to the amount of each frequency present in the input data. Such a system is thus a spectrum analyzer.

We next consider the output light distribution when the input signal to the cell is a chirp or linear frequency modulated waveform whose frequency increases linearly with time. In this case, the angle of the deflected light distribution increases linearly with time and a scanner results. Use of a galvanometer or similar deflector on the second orthogonal axis yields a two-dimensional (2-D) deflector. AO cells operated in such a mode result in fast deflector systems for use in recording input data at high rates. The third mode of operation for AO cells, and the one most used in this paper, can be seen by noting that the input signal $g(t)$ to the cell moves along the length of the cell at a velocity v_s (the acoustic velocity of propagation of the AO material used). We can thus describe the transmittance τ of the AO cell with such an input signal for this mode of operation as a function of time (t) and space (x) by

$$\tau(x,t) = g(t\text{-}x/v_s) \tag{1a}$$

or

$$\tau(x,t) = g(x\text{-}v_s t). \tag{1b}$$

In this summary presentation, we assume that the response of the AO cell is linear, and we note that this can and is achieved in practice by restricting the range of the input voltage.

Acousto-optic transducer technology has produced cells with 1 GHz bandwidth (and 1 μs aperture transit times), 100 MHz bandwidths (and 20 μs aperture times), or similar intermediate values with effective time-bandwidth product values of 1×10^3-2×10^3. These performance specifications make the AO spectrum analyzer attractive for many applications. Bragg receivers with 1 MHz frequency resolution over a 1 GHz bandwidth have been fabricated and used in many diverse applications to determine the frequency content of radar signals and other sources. Such information is also of use in determining the frequency and pulse repetition frequency of operation and the coding employed in a given signal transmission. This information is of subsequent use for *jamming, countermeasures*, and other techniques.

Another technique of immense use in discriminating the multitude of signals entering an advanced receiver is to determine the direction of arrival and frequency (f) of all input signals and to display the full spectrum of wide-band signals in a 2-D direction-of-arrival versus f plot. A technique to achieve such a display is shown [3] in Fig. 2. Such a system has been fabricated by GTE Sylvania for the Naval Research Laboratory. In this AO processor, the received signals from different elements of a phased array receiver are fed to a multichannel AO processor, and the 2-D Fourier transform of the input signal pattern is produced on a detector system. The vertical Fourier transform provides frequency information on the input data, and the horizontal Fourier transform produces a deflected beam with deflection angles proportional to the time delays between the signals in successive channels, and hence proportional to the angle from which the input signal

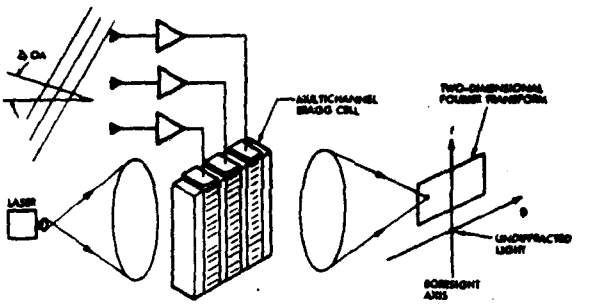


Fig. 2. Functional diagram of a multichannel acousto-optic system to provide a 2-D direction of arrival versus frequency output display of multiple input signal data [3].

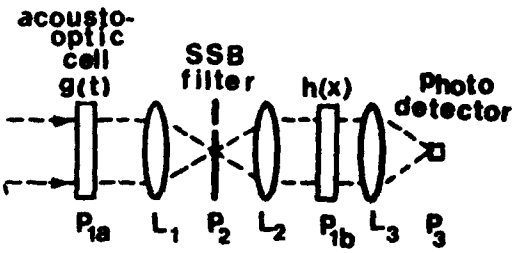


Fig. 3. Schematic diagram of a space integrating acousto-optic correlator.

originated (i.e., the direction-of-arrival of the data). Several such systems with diverse performance specifications have been fabricated and used with quite good success.

The major concern of this review paper will be advanced AO processors, specifically, correlators for signal demodulation and target range and Doppler determination rather than for spectrum analysis and the various 2-D displays noted above. The two fundamental AO correlator architectures use space integration and time integration. The space integrating system is shown [4] schematically in Fig. 3. In this system, the transmittance of the AO cell at P_{1a}, fed with a signal $g(t)$, is described as $g(x\text{-}v_s t)$. This light distribution is imaged onto plane P_{1b} (by lenses L_1 and L_2) where a fixed mask with transmittane $h(x)$ is placed. The light pattern leaving P_{1b} is then the product $h(x)\, g(x\text{-}v_s t)$. Lens L_3 forms the Fourier transform of the this product of two signals. This Fourier transform is evaluated by a single on-axis photodetector. The output plane pattern is then

$$R(t) = \int h(x)\, g(x\text{-}v_s t)dx = h \circledast g \tag{2}$$

or the correlation of the input and reference signals g and h. From (2), we note that the integration is performed over space (x) and the correlation appears as a function of the time output from the photodetector. Hence, this system is referred to as a space integrating AO correlator.

In practice, the input signal to the AO cell and the reference pattern h on the mask are carrier modulated and biased to yield real and positive waveforms. The Fourier transform of

the input data is formed at P_2 where the dc component (bias level) and one sideband are removed by spatial filtering. This single sideband (SSB) filtering in conjunction with quadrature input modulation yields a complex wavefront incident on plane P_{1b}. For simplicity, such issues together with other detailed system aspects such as illumination of the AO cells at the Bragg angle will be suppressed in our AO signal processing system analysis. Instead we shall assume that these techniques (or equivalent optical and electronic heterodyne detection) are used to enable complex correlations to be realized on such systems when necessary, even though real and positive inputs are used.

The second general class of AO correlator is the time integrating system [5] of Fig. 4. In this system, the signal $g(t)$ is used to modulate the output light from a light-emitting diode (LED) or laser diode source in time proportional to the input signal. This light distribution is collimated and used to uniformly illuminate an AO cell at P_1 fed with the signal $h(t)$. In this system, we describe the light distribution incident on P_1 by $g(t)$ and the transmittance of the AO cell as $h(t\text{-}x/v_s)$. Leaving P_1, we find the product $g(t)\,h(t\text{-}x/v_s)$. This pattern is imaged onto a one-dimensional (1-D) output detector at P_3 (with SSB modulation at P_2 as needed). The P_3 detector provides time integration of the incident signal, yielding an output pattern

$$R(x) = \int g(t)\, h(t\text{-}x/v_s)dt = g \circledast h \qquad (3)$$

that is the correlation of the received and reference signals g and h. In this system, the correlation is performed by integration in time on the detector and the correlation is displayed as a function of space. Thus, this system is referred to as a time integrating correlator.

At this point, it is worthwhile to pause to reflect upon the advantages and disadvantages of these two AO correlators. The space integrating system of Fig. 3 can search a large range delay or time delay in the arrival of the input signal. This is obvious since the photodetector has no output until the received signal enters the AO cell. Conversely, in this system, the signal must be in the AO cell for the correlation to occur. Thus, this space integrating system can only handle signals with a time-bandwidth product equal to that of the AO cell or the P_{1b} mask (the mask resolution is normally not the limiting factor) or approximately 10^3 for most devices. Conversely, in the time integrating system of Fig. 4, the received and reference signals must be present within a much smaller time delay equal to the aperture time of the AO cell. However, this system can provide integration over a signal of longer time duration and time-bandwidth product than that of the AO cell. The practical limitation is the SNR or dynamic range of the output detector used, although digital storage and addition of detector outputs in a postprocessor can increase the possible integration time even further. Thus, we find the space integrating system to provide large range or time delay searches, whereas the time integrating system allows processing of large time-bandwidth product signals. In Section III, we describe a hybrid time and space integrating AO system that realizes the advantages of both processors.

The AO systems of Figs. 3 and 4 achieve high bandwidth

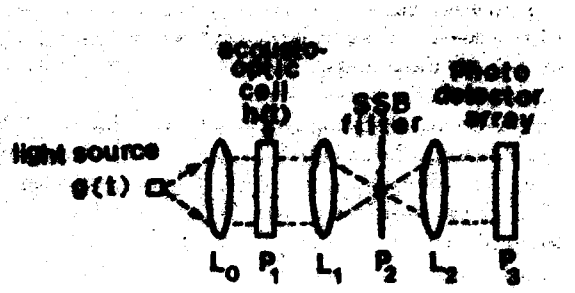


Fig. 4. Schematic diagram of a time integrating acousto-optic correlator.

(up to 1 GHz) and long integration times; however, both systems use 1-D transducers and provide only 1-D output patterns. Some of the most demanding operations requiring advanced signal processors necessitate 2-D output data displays such as the simultaneous range and Doppler of multiple targets. AO signal processing architectures have responded to this challenge with new algorithms and architectures. The basis for these techniques is the optical realization of the chirp-Z transform. To understand this algorithm, we consider how to achieve a Fourier transform on a AO correlator system. A 1-D analysis is used for simplicity. We first write the Fourier transform of a 1-D signal $f(t)$ as

$$F(\omega) = \int f(t) \exp(-j\omega t)dt. \qquad (4a)$$

We then substitute the identity

$$-\omega t = (t-\omega)^2/2 - t^2/2 - \omega^2/2$$

into the exponent in (4a). This yields

$$F(\omega) = e^{-j\omega^2/2} \int f(t) e^{-jt^2/2} e^{j(t-\omega)^2/2} dt, \qquad (4b)$$

which can be rewritten as a correlation

$$F(\omega) = e^{-j\omega^2/2} \left[f(t) e^{-jt^2/2} \right] \circledast \left[e^{jt^2/2} \right]. \qquad (4c)$$

In the form shown (4c), we see that the Fourier transform of $f(t)$ can be realized [6] by premultiplying $f(t)$ by a chirp $\exp(-jt^2/2)$ and correlating the signal with another chirp $\exp(+jt^2/2)$. Postmultiplication by another chirp $\exp(-j\omega^2/2)$ yields the exact Fourier transform. Such an exact Fourier transform is rarely necessary as the magnitude is usually sufficient for most applications. This algorithm can be used on the time integrating system of Fig. 4 by feeding $f(t)$ $\exp(-jt^2/2)$ to the LED and $\exp(+jt^2/2)$ to the AO cell. A similar technique can be used to convert the space integrating correlator of Fig. 3 to a Fourier transform system. An interesting multichannel Fourier transform version of Fig. 4 can be realized [7] if a linear array of N input LED's and a 2-D output time integrating detector are used. If each LED is fed with a separate signal f_1-f_N multiplied by a chirp, the 2-D outputs are the 1-D transforms F_1-F_N of the N input signals, each present on a separate horizontal line in the output plane. Such a system should be of use in advanced Fourier spectroscopy imaging signal processors.

AD-A130 097 OPTICAL PATTERN RECOGNITION FOR MISSILE GUIDANCE(U) CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF ELECTRICAL ENGINEERING D CASASENT 15 NOV 82 AFOSR-TR-83-0556

2/2

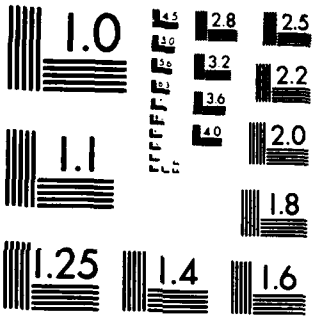
UNCLASSIFIED AFOSR-79-0091 F/G 17/7 NL

END

DATE FILMED

8-83

DTIC



Let us now consider how the above techniques can be used [8] to achieve 2-D signal processing operations using only two 1-D AO transducers. The basic architecture [9] used is the triple product processor shown in simplified form in Fig. 5. In this 2-D time integrating system, the input LED and AO-1 produce the correlation of s_0 and s_1 vertically in x at the ouput plane. The correlation of s_0 and s_2 is formed vertically by AO-2 at the ouput plane. The 2-D system output can be described in terms of three input signals and the delay factors τ_1 and τ_2, associated with AO-1 and AO-2 by

$$f(x,y) = f(\tau_2, \tau_1) = \int s_0(t)\, s_1(t-\tau_1)\, s_2(t-\tau_2)dt. \quad (5)$$

In Sections II and IV, two specific applications of this system using the chirp-Z algorithm are described to demonstrate the flexibility and processing power of this architecture.

FOLDED-SPECTRUM WIDE-BAND OPTICAL SIGNAL PROCESSING

In many wide-band applications, greater than a 10^3-point Fourier transform (10^3 resolvable frequencies) is necessary. In such cases, a 1-D signal can be rastered-recorded on V lines with H cycles per line in a 2-D format. The 2-D optical Fourier transform of this signal record results in a folded-spectrum output with coarse and fine frequency axes [10]. With $T = VT_h$ (where $f_h = 1/T_h$ is the horizontal line scan rate) seconds of data recorded covering a bandwidth W, the 2-D folded-spectrum output contains H coarse frequency loci with a coarse frequency resolution of f_h and with V fine frequencies resolvable on each locus. The output thus has a fine frequency resolution $f_f = 1/T$ over the full signal bandwidth W and represents a VH point (or up to a 10^6-point) Fourier transform.

This technique has been known and used for over 14 years with film recorded data and with real-time 2-D spatial light modulators such as the General Electric light value. These systems have proven most useful and promising; however, they have limited real-time bandwidth capability. Recent device architecture and optical engineering advances now make much larger bandwidth systems of this type possible.

One version of such a system, shown [11] in Fig. 6, uses a new 2-D addressing technique to provide wider bandwidth data recording. In this system, the 1-D input data are fed to an AO cell of length T_c μs. Every T_c μs, a cavity-dumped laser is pulsed on, imaging the full contents of the AO cell in parallel onto one line of a 2-D optically addressed liquid crystal light valve (LCLV) [12]. Successive T_c μs portions of the input signal are recorded on successive vertical lines on the liquid crystal light valve by the AO vertical deflector shown. The raster-recorded signal pattern on the liquid crystal light valve is then read out in reflection with the continuous laser that passes through the beam splitter (BS) shown, reflected from the right-hand side of the liquid crystal light valve, passed back to the

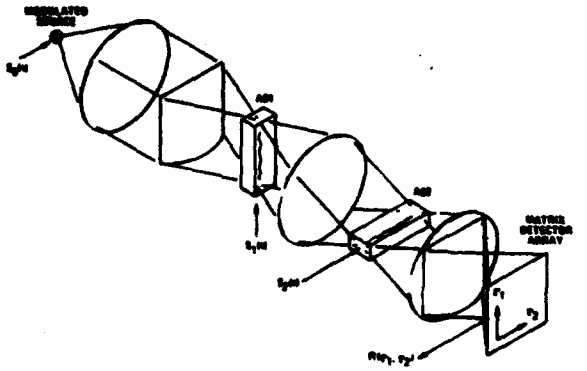


Fig. 5. Schematic diagram of an acousto-optic 2-D time integrating triple product processor [9].

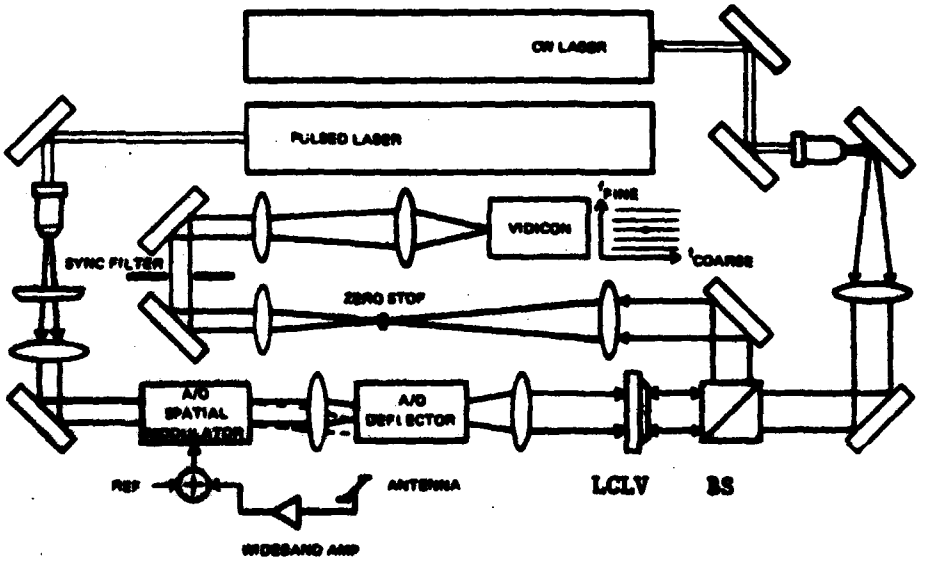


Fig. 6. Schematic diagram of a snapshot-addressed real-time folded-spectrum optical signal processing system [11].

beam splitter, and then up to a 2-D Fourier transform lens system whose output is sensed by a vidicon or other output detectors and fed to various displays and/or digital analysis systems.

This system allows writing of one line of information of $T_c = 6$ μs duration with greater than 10^3 points of data written in parallel every 6 μs. Vertical deflection with 6 μs times and over 500-point resolution are readily available. Such a system can thus provide a spectrum analyzer with over 5×10^5 resolvable frequencies in a signal bandwidth of 100 MHz. The well-engineered real time system of Fig. 6 was fabricated and achieved a bandwidth of 150 MHz and a time-bandwidth product of 200 000. Other larger bandwidth and resolution systems can be fabricated with present optical components using this snapshot line-at-a-time addressing technique.

The triple product processor system of Fig. 5 has also been used to realize a 2-D folded-spectrum optical signal processing system [9]. To achieve this, periodic chirps of period T and chirp rates α and $\alpha(N\text{-}1)/N$ are fed to the LED source and AO-2, respectively. The input signal is multiplied by the chirp fed to the LED. A slow chirp of period NT and chirp rate α/N is fed to AO-1. The chirp rates are chosen so that the triple product in (5) generates a constant difference frequency. A 2-D folded-spectrum output pattern then results, with τ_1 and τ_2 being the fine and coarse frequency axes, respectively. A total integration time of NT is used and a frequency resolution proportional to $1/NT$ results. Several fully automated and on-line versions of such a system exist in various facilities. We achieve large bandwidths and frequency resolutions, both of which can be electrically adjusted by appropriately varying the chirp signals fed to the different system inputs.

OPTICAL SIGNAL PROCESSING FOR SPREAD SPECTRUM COMMUNICATIONS

The long coded waveforms being used in spread spectrum systems pose many unique problems for the associated signal processors. Long coded signals are used to achieve large processing gains for increased noise immunity, better ranging accuracy, and to enable transmission below the noise level to reduce signal detectability by unwanted persons [13]. To achieve the full processing gain possible with such waveforms, the reference and received signal patterns must be time or space aligned, i.e., synchronized. Once this signal alignment is achieved, a correlation on a quite long signal (often much greater than 10^3 bits in length) must be accomplished. Several optical signal processing techniques to address these spread spectrum signal processing issues follow to demonstrate quite new and different algorithms and approaches to signal processing.

Let us first consider how to determine the starting location of a long coded waveform that is repeated for several cycles. If the signal is 10^4 bits in length and the full processing gain is desired, we could record all 10^4 possible sets of the signal in each of its 10^4 possible starting locations and correlate the input signal with all of these possible signal replicas. The size of the associated filter and the required system time-bandwidth product are quite excessive. A more attractive technique was suggested by Spann [14] and later applied to an optical system [15]. In this Spann correlator, the received signal is raster-recorded in the input plane and a unique Fourier transform mask is placed in the frequency plane. The Fourier transform of the product of the transform of the input signal and the mask is formed in an optical frequency plane correlator [16]. The system output is then the correlation of the input and the mask functions. To record the mask for an N-bit signal, we write N as $N = N_1N_2$. We then raster-record the N-bit input signal repetitively with $2N_1$ points per line for $2N_2$ lines, remove the first line in the first column from this pattern, and record an optical matched spatial filter of the remaining $(2N_1-1) \times (2N_2-1)$ bit pattern. Spann [14] has shown that all possible $N_1 \times N_2$ raster-recorded versions of the N-bit repetitive input signal in all N possible starting locations exist in this mask. Thus, the location of the output peak in such a folded-spectrum optical correlator provides information on the starting location of the signal, and the amplitude of the output correlation at this location is the decoded signal data.

AO correlators are also useful for various spread spectrum signal processing functions. If a large range delay search is desired, a space integrating correlator is preferable; conversely, if a correlation of a long coded waveform is desired, a time integrating correlation is the preferred choice. In cases when a large time delay (for synchronization) and correlation of a long signal (for spread spectrum applications) are both desired, a new hybrid time and space integrating AO correlator architecture with the best advantages of both the space integrating and time integrating systems can be used. A schematic diagram of such a system is shown [17] in Fig. 7. Consider its use in the synchronization and decoding of a long frequency-hopped signal in which the transmitted frequency changes every T μs according to a direct sequence or similar code (i.e., a new frequency, or chip, is transmitted every T μs). In Fig. 7, we assume that the AO cell has an aperture time of T μs and that the output time integrating CCD shift register detector has a clock rate of $1/T$ μs. The mask at P_2 contains an aperture on each of its N vertical lines, with the horizontal location of the aperture on line N proportional to the frequency transmitted at time NT. In operation, as each new signal frequency chip enters the AO cell, light passes through the corresponding aperture at P_2 and falls on the corresponding photodetector. With the shifting of the output detector data, successive chips add by time integration on the detector. Thus, the time of occurrence of the output from the shift register detector provides time synchronization information and the output correlation peak value has the SNR of the full code. This system provides both a large time delay search (by virtue of the space integrating front end) and a long signal integration time (achieved by time integrating on the detector with over 10^3 detector elements easily possible with present technology). This system is typical of a new class of AO processors that combine the best features of time and space integrating systems. Many versions of such hybrid time

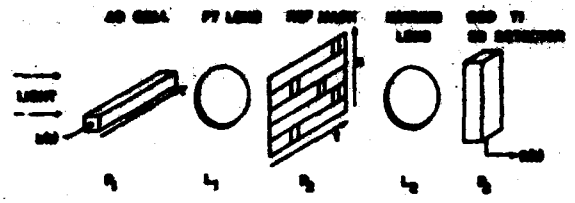
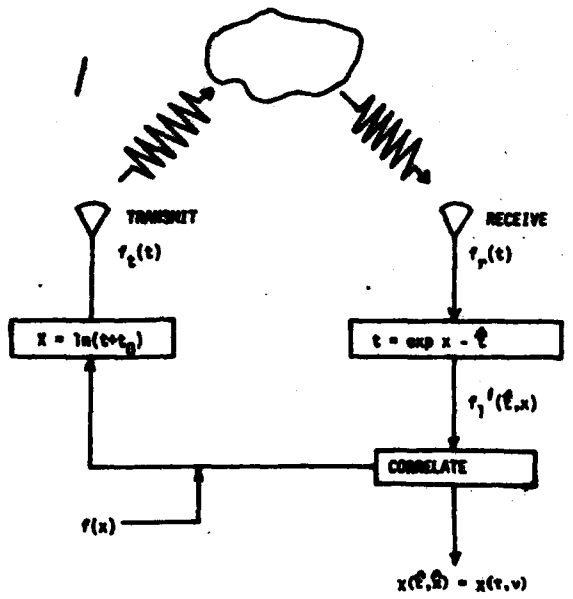


Fig. 7. Schematic diagram of a hybrid time and space integrating correlator for spread spectrum signal synchronization and decoding [17].



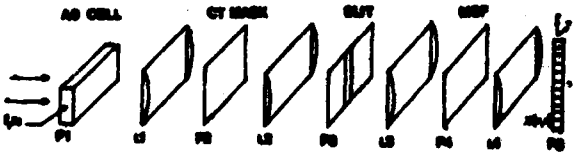


Fig. 8. Functional block diagram (a) and schematic (b) of a space-variant spread spectrum system and processor [18].

and space integrating systems for diverse signal processing applications are expected to emerge in future years.

A quite different spread spectrum technique is shown [18] in functional block diagram form in Fig. 8(a). In this system, a log coordinate transformation (with $t_0 = 0$ for simplicity) is applied to the signal $f(x) = \cos[\omega_0 x + \phi(x)]$ to produce the transmitted signal $f_t(t) = \cos[\ln\omega_0 t + \phi(\ln t)]$. The signal is transmitted, and upon reception, the inverse exp coordinate transformation is applied to yield $f_r(e^x - t)$ where f_r will, in general, be a delayed and Doppler shifted version of f_t. The correlation of f_r and f will have a peak in time $\hat{t}$ proportional to the range of the target and in space x at a location proportional to the target's Doppler velocity. The latter property results because the Fourier transform of a log coordinate transformed signal is the Mellin transform, which is scale invariant [19]. The entire receiver processing can be performed optically on the system [20] of Fig. 8(b). The received signal is fed to the AO cell at P_1 and imaged onto a vertical slit at P_3. A frequency plane mask at P_2 achieves the required exp coordinate transformation for all possible time delays t. The rest of the system is a conventional 1-D matched spatial filter correlator. The moving window input to this system produces 1-D outputs on the detector at P_5. At each instant of time, the P_5 output is a slice in space x (or Doppler) for a given time delay search in the starting location of the coordinate transformation.

Such a system thus provides synchronization (or ranging) and data decoding with Doppler invariance to platform motion. But more so, the transmitted signal is a nonlinear spread spectrum waveform with nonlinearity produced by the coordinate transformation applied prior to transmission. This coordinate-transformation operation nonlinearly spreads the carrier ω_0 to prevent its detection. It also spreads the coded waveform itself. Upon reception, the inverse coordinate transformation compresses the signal so that the correlation need be performed over a shorter length coded signal. Should noise enter the receiver, the exponential coordinate transformation spreads it and shifts its frequency out of the signal passband, providing increased noise immunity. Such a space-variant optical system represents a quite different approach to spread spectrum, radar, and signal processing.

OPTICAL SIGNAL PROCESSING FOR AMBIGUITY FUNCTION COMPUTATION

Determining the range and Doppler of a target simultaneously is one of the most demanding processing functions required in radar. The resultant ambiguity function that must be computed to achieve this operation on two signals s_1 and s_2 is

$$\chi(\tau,f) = \int\int s_1(t)\, s_2^*(t-\tau)\, e^{-j2\pi ft} dt. \qquad (6)$$

Many OSP systems have been proposed and several laboratory systems to realize this operation have been fabricated. Perhaps the most well-engineered system is the one shown [21] in Fig. 9. This system was built and demonstrated. It was required to be a fully real-time system operating at over 150 MHz bandwidth on pulse burst waveforms with time-bandwidth product $> 2 \times 10^5$. The transmitted reference signal for this processor was to be adaptive under computer control at a rate of 30 times/second. All necessary specifications were met by the system, and thus it represents a major achievement in optical engineering and signal processing system performance achieved with state of the art components.

The pulse compressor system of Fig. 9(a) is a space integrating correlator modified with optical heterodyne detection (not shown for simplicity) and a real-time reference mask. The sequence of received pulsed burst waveforms to be

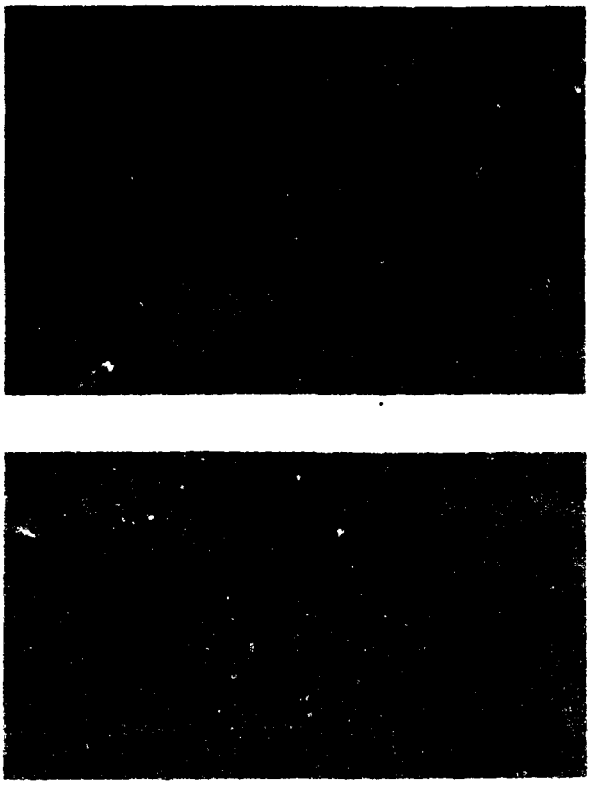


Fig. 9. Schematic diagram of a radar ambiguity function processor (a) pulse compressor, (b) Doppler processor [21].

processed are fed to the AO cell, whose output is SSB filtered and imaged onto the left side of a liquid crystal light valve on which the reference signal is written by imaging from a CRT under computer control. The product of the transmittance of the AO cell and the reference signal is then reflected from the liquid crystal, Fourier transformed, and heterodyned-detected by a photodiode. The time output from the photodiode thus represents the N complex correlations of the reference with the N received signals in the pulse burst. The location of the correlation peak with respect to the signal's time reference provides the necessary range information. Doppler information is contained in the phase across the sequence of correlations.

To obtain the necessary Doppler data, the N correlations are fed to the system of Fig. 9(b). Here they are snapshot written (as in the system of Fig. 6) one line at a time onto N lines of the second liquid crystal light valve. These data are then read out in reflection from the right-hand side, and the 1-D vertical Fourier transform of the pattern on the light valve is formed on the output plane. This provides the necessary range and Doppler output ambiguity function display.

A more complex form of ambiguity function arises when two signals from the same source are passively received at different detectors. To determine the target's differential range and Doppler, the ambiguity function of the received passive signals must be formed. To classify the target, this operation must be performed for many band-limited received signals in different threat bands. The space integrating AO system of Fig. 3 can achieve the necessary operations if a linear output photodetector array is used [22]. In this case, one signal is recorded (with quadrature modulation) on the mask and the second signal is fed to the AO cell. As a function of time, the output from one detector is the correlation of the two signals. If the two signals have different frequencies, imaging one onto the other produces a beat frequency equal to the difference between the frequencies of the two signals. This signal is deflected to different elements of the output detector. Thus, the detector output at each instant of time is a slice of the ambiguity function parallel to the Doppler axis for one τ value. Note that this modified AO system yields the 2-D ambiguity function output using only 1-D transducers and a 1-D output detector.

A more advanced version of an optical signal processing system for this application that is presently being fabricated uses the triple product processor system of Fig. 5. If the inputs to this system are chosen as

$$s_0(t) = f_1(t)\, e^{-j\pi t^2} \tag{7a}$$

$$s_1(t) = f_2(t) \tag{7b}$$

$$s_2(t) = e^{+j\pi t^2}, \tag{7c}$$

it is easy to show that the output in (5) is the desired ambiguity function of (6). In practice, real and positive input signals are necessary. Thus, quadrature modulation, SSB filtering, and other system modifications are necessary to produce the desired result. The triple product processor system thus represents a quite general purpose optical signal processing architecture with the ability to perform 2-D folded-spectrum wide-band signal analysis or ambiguity function computation merely by changing the electronic input signals appropriately.

OPTICAL SIGNAL PROCESSING FOR ADAPTIVE PHASED ARRAY RADAR

The final signal processing application that we consider is the computation of the weights w_n necessary to apply to the received signals s_n of an N element phased array to steer the output antenna in a given direction defined by a steering vector $\bar{S}$ and to adaptively null the entire far field antenna noise distribution in angle and frequency or time. Many candidate techniques [23] exist to solve this problem; the processing necessary for each is usually the problem. In one technique, the cross correlation of all N received signal pairs is computed, and these peak correlation values are entered in a covariance matrix $\underline{M}$. Describing the antenna weights as a vector $\bar{W}$, we can describe the relationship between the known matrix $\underline{M}$ and steering vector $\bar{S}$ and the unknown $\bar{W}$ by

$$\underline{M}\bar{W} = \bar{S}. \tag{8}$$

The solution to (8) is

$$\bar{W} = \underline{M}^{-1}\bar{S}. \tag{9}$$

However, for arrays with many adaptive elements, inverting the matrix $\underline{M}$ in (9) is quite time consuming.

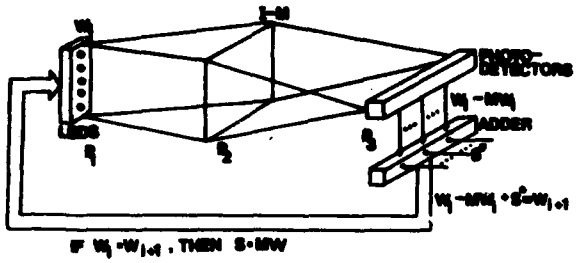


Fig. 10. Schematic diagram of an iterative optical processor for adaptive phased array radar signal processing [24].

An optical system to solve (8) for $\overline{W}$ has been devised and fabricated [24]. It solves (8) using the iterative algorithm

$$\overline{W}_{i+1} = (\underline{I} - \underline{M})\ \overline{W}_i + \overline{S} \qquad (10)$$

where $\underline{I}$ is the identity matrix and $\overline{W}_i$ is the ith iterative. When $\overline{W}_i = \overline{W}_{i+1}$, (10) reduces to (8). The system's output is thus $\overline{W}$ and it has been obtained without having to invert $\underline{M}$. The iterative optical processor used is shown schematically in Fig. 10. The input at P_1 is a linear array of LED's or laser diodes whose outputs are the elements of $\overline{W}_i$. Each LED is imaged vertically and expanded horizontally to uniformly illuminate the respective row of P_2 where a mask with transmittance $(\underline{I} - \underline{M})$ is placed. The light distribution leading each column of this mask is then summed on separate output detectors at P_3. The P_3 output is thus the matrix-vector product $(\underline{I} - \underline{M})\ \overline{W}_i$. To this we add $\overline{S}$ and return the sum to P_1 as the new iterative input, thus realizing (10).

In practice, the parallel addition of $\overline{S}$ can be achieved by adding a row to the P_2 matrix with the summation performed directly on the output detector. Other system modifications allow it to perform complex-valued matrix-vector operations. To decrease crosstalk between LED outputs at P_2 and to provide uniform illumination of each row of P_2, fiber optic connections are used between P_1 and P_2 in the system fabricated. The height of the detector elements used was chosen to equal the vertical size of the P_2 mask. This enabled the photodetector to be mounted directly behind the mask, thus eliminating all optics from the system. The electronic feedback loop includes a microprocessor, arithmetic logic unit, controller, analog-to-digital converters, and the necessary drive and demultiplexing circuitry. Random access memories are used to correct for LED and photodetector nonuniformities and pulse-width modulation of the LED's was employed to provide increased system accuracy. The final 10×10 element fiber optic/microprocessor-based iterative optical processor system assembled performs quite accurately and satisfactorily. More so, it represents yet another OSP architecture that promises to be of quite general use in many diverse data processing problems.

SUMMARY/CONCLUSION

In this review paper, I have attempted to convey the wealth of optical signal processing algorithms, architectures, and systems that exist. By example and specific signal processing case studies, I have described many optical signal processing systems and their use in different applications. In all cases considered, many fabricated systems exist and many other quite attractive new advanced optical signal processing techniques are available with the high bandwidth and throughput necessary for advanced signal processing. It thus appears that optical signal processing systems and the use of bulk acousto-optic transducers offer a quite flexible and powerful approach to signal processing needs of the future.

ACKNOWLEDGMENT

The assistance of many companies in providing data and photographs for this paper is acknowledged, as is the support of the author's research by the Air Force Office of Scientific Research, Hanscom Air Force Base, NASA, and in part by the National Science Foundation.

REFERENCES

[1] *Proc. IEEE, Special Issue on Acousto-Optics*, Jan. 1981.
[2] *Proc. Soc. Photo-Opt. Instr. Eng.*, Vol. 214, 1979.
[3] R. Coppock and R. Croce, *Proc. Soc. Photo-Opt. Instr. Eng.*, vol. 214, p. 124, 1979.
[4] L. Cutrona *et al.*, *IRE Trans. Inform. Theory*, vol. IT-6, p. 386, 1960, and D. Hecht, *Proc. Soc. Photo-Opt. Instr. Eng.*, vol. 90, p. 148, 1976.
[5] R. Sprague and C. Koliopoulos, *Appl. Opt.*, vol. 15, p. 89, 1976.
[6] L. Rabiner *et al.*, *Bell Syst. Tech. J.*, p. 1249, May 1969.
[7] D. Casasent and D. Psaltis, *Appl. Opt.*, vol. 19, p. 2034, 1980.
[8] H. J. Whitehouse *et al.*, "Signal processing architectures using transversal filter technology," in *Proc. IEEE Symp. Circuits Syst.*, Boston, MA, Apr. 1975.
[9] P. Kellman, *Opt. Eng.*, vol. 19, p. 370, 1980.
[10] C. Thomas, *Appl. Opt.*, vol. 5, p. 1782, 1966.
[11] J. Anderson *et al.*, *Proc. Soc. Photo-Opt. Instr. Eng.*, vol. 180, p. 128, 1979.
[12] W. Bleha *et al.*, *Opt. Eng.*, vol. 17, p. 371, 1978.
[13] R. C. Dixon, *Spread Spectrum Systems*. New York: Wiley, 1976.
[14] R. Spann, *Proc. IEEE*, vol. 53, p. 2137, 1965.
[15] D. Casasent and R. Kessler, *Opt. Commun.*, vol. 17, p. 242, 1976.
[16] A. Van der Lugt, *IEEE Trans. Inform. Theory*, vol. IT-10, p. 139, 1964.
[17] D. Psaltis and D. Casasent, *Appl. Opt.*, vol. 19, p. 1546, 1980.
[18] D. Casasent and D. Psaltis, *Opt. Lett.*, vol. 4, p. 18, 1979.
[19] ——, *Proc. IEEE*, vol. 65, p. 77, 1977.
[20] ——, *Proc. Electro-Opt. Syst. Design Conf.*, Oct. 1979, p. 333.
[21] H. Brown and B. Markevich, *Proc. Soc. Photo-Opt. Instr. Eng.*, vol. 128, p. 204, 1977.
[22] B. Kumar and D. Casasent, *EASCON 79 Rec.*, Oct. 1979, p. 595.
[23] *IEE*, vol. 127, part F, Aug. 1980.
[24] D. Psaltis *et al.*, Opt. Lett., vol. 4, p. 348, 1979.

David Casasent is a Full Professor of Electrical Engineering at Carnegie-Mellon University, Pittsburgh, PA, where he holds the George Westinghouse Chair of Electrical Engineering. He is Head of the Hybrid Image and Signal Processing Laboratory. He is the author and Editor of several books and eight journal special issues, chapters in eleven books, and over 180 technical journal papers.

He has received various honors and awards including Fellow of SPIE, OSA, and IEEE for his research in optical data processing. He has also received other awards such as the Ryan Prize for outstanding teaching and research and Best Paper Awards for IEEE-AES and AIAA for his work on radar and missile guidance. He is quite active in conference organizations, various technical societies, and as a consultant to many companies and several government agencies. ■

15. PUBLICATIONS AND TALKS

15.1 PUBLICATIONS (AFOSR SUPPORTED, 1979-DATE)

Publications from 30 September 1979 - 30 September 1980 on work performed under AFOSR-79-0091 are listed in Section 15.1.1. Publications during 30 September 1980 - 30 September 1981 follow in Section 15.1.2 and our new publications from September 1981 - September 1982 follow in Section 15.1.3. A list of submitted and pending papers follows in Section 15.1.4.

15.1.1 PUBLISHED PAPERS UNDER AFOSR SUPPORT (30 SEPTEMBER 1979 - 30 SEPTEMBER 1980)

1. "Photo-DKDP Light Valve in Optical Data Processing", Applied Optics, 18, 3307-3314, October 1979 (Casasent, Luu).
2. "Coherent Optical Pattern Recognition", Nikkei Electronics, 150-181, October 1979 (in Japanese) (Casasent).
3. "Optical Data Processing for Advanced Missile Guidance Needs", AIAA, October 1979 (Casasent).
4. "Spread Spectrum Optical Signal Processors", Proc. EOSD, 333-342, October 1979 (Casasent, Psaltis).
5. "Space Blur Bandwidth Product in Correlator Performance Evaluation", JOSA, 70, 103-110, January 1980 (Kumar, Casasent).
6. "Optical Image Processing", EOSD, Tokyo, January 1980 (in Japanese) (Casasent).
7. "Optical Signal Processing", EOSD, Tokyo, January 1980 (in Japanese) (Casasent).
8. "Beyond Matched Filtering", Opt. Engr., 19, 152-156, March 1980 (Caulfield et al).
9. "Multivariant Technique for Multi-Class Pattern Recognition", Applied Optics, 19, 1758-1761, June 1980 (Psaltis, Casasent).
10. "Optical Fourier Transform Techniques for Advanced Fourier Spectroscopy", Applied Optics, 19, 2034-2037, June 1980 (Casasent, Psaltis).
11. "Nonlinear t-E Curve Effects in an Optical Correlator", Opt. Commun., 34, 4-6, July 1980 (Kumar, Casasent).

12. "Correlation of Images with Random Contrast Reversals", SPIE, 238, 156-165, July 1980 (Barniv, Mostafavi, Casasent).

13. "A Laser Diode Lensless MSF-HOE Correlator", Applied Optics, 19, 2653-2654, August 1980 (Caimi et al).

15.1.2 PUBLISHED PAPERS UNDER AFOSR SUPPORT (30 SEPTEMBER 1980-30 SEPTEMBER 1981)

14. "Hybrid Processor to Compute Invariant Moments for Pattern Recognition", Opt. Lett., 5, 395-397, September 1980 (Casasent, Psaltis).

15. "Optical Word Recognition, Case Study in Coherent Optical Pattern Recognition", Opt. Engr., 19, 716-721, September 1980 (Casasent et al).

16. "Lensless Matched Spatial Filter Correlator Experiments", Opt. Commun., 34, 311-315, September 1980 (M. Shen et al).

17. "HOE/Lensless Matched Spatial Filter Correlator Experiments", Opt. Commun., 34, 316-320, September 1980 (M. Shen et al).

18. "A Laser Diode/Lensless MSF Optical Pattern Recognition System", EOSD, 46-52, November 1980 (Casasent et al).

19. "Optical Pattern Recognition: Matched Spatial Filter Processors", EOSD, 33-39, November 1980 (Casasent).

20. "Optical Pattern Recognition: Beyond Matched Spatial Filtering", EOSD, 39-47, March 1981 (Casasent).

21. "Pattern Recognition: A Review", IEEE Spectrum, 28-33, March 1981 (Casasent).

22. "Processing Flexibility by Hybrid Optical/Digital Techniques", Proc. Workshop of Future Directions in Optical Data Processing, Texax Tech. Rept., 1 March 1981, 17-23 (Casasent, Kumar).

23. "Beyond Holographic Matched Filtering", Israel Journal of Technology, 18, 255-260, March 1981 (Casasent).

24. "Binarization Effects in a Correlator with Noisy Input Data", Applied Optics, 20, 1433-1438, April 1981 (Kumar, Casasent).

25. "Correlation of Images with Random Contrast Reversals", SPIE, 238, 156-165, July 1980 (Barniv, Mostafavi, Casasent).

26. "Image Quality Effects in Optical Correlators", SPIE, 310, 183-192, August 1981 (Casasent, Eiva, Kumar).

27. "Multisensor Image Registration: Experimental Verification", SPIE, 292, 160-171, August 1981 (Barniv, Casasent).

28. "Intra-Class IR Tank Pattern Recognition Using SDFs", SPIE, 292, 25-33, August 1981 (Hester, Casasent).

29. "Inter-Class Discrimination Using SDFs", SPIE, 302, 108-116, August 1981 (Hester, Casasent).

15.1.3 PUBLISHED PAPERS UNDER AFOSR SUPPORT (30 SEPTEMBER 1981 - 30 SEPTEMBER 1982)

30. "An Iterative Optical Processor: Selective Survey of Operations Achievable", Proceedings NASA Langley Conference on Optical Information Processing, Publication 2207, August 1981, 105-118 (Casasent, Neuman).

31. "A Review of Optical Signal Processing", IEEE Commun., 40-48, September 1981 (Casasent).

32. "Optical Signal Processing II: Applications, Systems and New Techniques", EOSD, 41-47, September 1981 (Casasent).

33. "The Soviet Priz Spatial Light Modulator", Applied Optics, 20, 3090-3092, September 1981 (Casasent, Caimi, Khomenko).

34. "A Laser Diode/HOE Pattern Recognition System", Acta Optica Sinica, 1, 401-410, September 1981 (Casasent et al).

35. "Eigenvector Determination by Iterative Optical Methods", Applied Optics, 20, 3707-3710, November 1981 (Kumar, Casasent).

36. "A New Soviet BSO Light Modulator for Optical Data Processing", Proc. EOSD, 297-303, November 1981 (Casasent, Caimi).

37. "A Correlator for Optimum Two-Class Discrimination", Proc. EOSD, 321-330, November 1981 (Casasent et al).

38. "Test and Evaluation of the Soviet Prom and Priz Spatial Light Modulators", Applied Optics, 20, 4215-4220, December 1981 (Casasent, Caimi, Khomenko).

39. "A Microprocessor-Based Fiber-Optic Iterative Optical Processor", Applied Optics, 21, 147-152, January 1982 (Carlotto, Casasent).

40. "Principal Component Imagery for Statistical Pattern Recognition Correlators", Opt. Engr., 21, 43-47, January/February 1982 (Kumar, Casasent).

41. "Adaptive Phased Array Radar Processing Using an Optical Matrix-Vector Processor", SPIE, 341, May 1982 (Casasent, Carlotto).

42. "Advanced Acousto-Optic Signal Processors", Proc. SPIE, 352, August 1982 (Casasent).

43. "New Research in Holographic Pattern Recognition", Proc. SPIE, 353, August 1982 (Casasent).

44. "Synthetic Discriminant Functions for 3-D Object Recognition", Proc. SPIE, 360, August 1982 (Casasent, Kumar, Sharma).

45. "Multidimensional Adaptive Radar Array Processing Using an Iterative Optical Matrix-Vector Processor", Opt. Engr., Vol. 21, No. 5, 814-821, September 1982 (Casasent, Carlotto).

46. "Applications of the Priz Light Modulator", Applied Optics, 21, No. 21, 3846-3854, November 1982 (Casasent, Caimi, Petrov, Khomenko).

15.1.4 PENDING PAPERS

47. "Realization of a Sobel Operator by Coherent Optical Techniques", Published in the Proceedings of the First International Congress on Applications of Lasers and Electro-Optics, Laser Institute of America, Boston, Massachusetts, September 1982 (Chen, Casasent).

48. "A Fisher Discriminant Approach to Distortion-Invariant Pattern Recognition Using Autocorrelations", Published in the Proceedings of the First International Congress on Applications of Lasers and Electro-Optics, Laser Institute of America, Boston, Massachusetts, September 1982 (Casasent, Chang).

49. "Nonlinear Local Image Preprocessing Using Coherent Optical Techniques", Applied Optics, Submitted September 1982 (Casasent, Chen).

50. "Frequency-Multiplexed and Pipelined Iterative Optical Systolic Array Processors", Applied Optics, Submitted September 1982 (Casasent, Jackson, Neuman).

51. "An Iterative Optical Vector-Matrix Processor", SPIE Institute Series, Accepted (Carlotto, Casasent).

52. "Hybrid Time and Space Integrating Processors for Spread Spectrum Applications", Submitted IEEE AES, (Silbershatz, Casasent).

"Generalized Chord Transform for Pattern Recognition", To be Submitted (Casasent, Chang).

15.2 SEMINAR AND PRESENTATIONS OF AFOSR RESEARCH (1 September 1981 - 1 September 1982)

September 1981:

1. NSA - Ft. Meade, Maryland, "Optical Pattern Recognition and Robotics".
2. Carnegie-Mellon University - Pittsburgh, Pennsylvania, "Optical Information Processing at CMU in Electrical Engineering".
3. Westinghouse Corporation - Pittsburgh, Pennsylvania, "Robotics".

November 1981:

4. La Habra, California, "Optical Data Processing".
5. EOSD Conference - Anaheim, California, "A New Soviet BSO Light Modulator for Optical Data Processing".
6. EOSD Conference - Anaheim, California, "A Correlator for Optimum Two-Class Discrimination".
7. Robotics Institute - Pittsburgh, Pennsylvania, "Optical Data Processing".

December 1981:

8. Northrop Corporation - Anaheim, California, "Optical Image Processing".
9. Itek Conference - Palo Alto, California, "Optical Pattern Recognition Using Moments".
10. Itek Conference - Palo Alto, California, "Optical Pattern Recognition Using Synthetic Discriminant Functions".

January 1982:

11. BDM Corporation - Virginia, "Optical Image Processing".

February 1982:

12. NASA Langley - Hampton, Virginia, "Optical Pattern Recognition for Industrial Inspection".
13. Iron and Steel Institute - Pittsburgh, Pennsylvania, "Optical Pattern Recognition for Industrial Automation".

March 1982:

14. Carnegie-Mellon University - Pittsburgh, Pennsylvania (Sophomore Seminar), "Optical Data Processing".

15. Optical Society of America - Pittsburgh, Pennsylvania, "Optical Data Processing".

16. AFIT - Dayton, Ohio, "Iterative Optical Processor for Missile Guidance".

May 1982:

17. SPIE Conference - Washington, D.C., "Adaptive Phased Array Radar Processing Using an Optical Matrix-Vector Processor".

18. Lockheed Corporation - Palo Alto, California, "Sub-Pixel Image Registration".

19. General Dynamics, Pomona Division - Pomona, California, "Moment-Based Optical Pattern Recognition".

20. General Dynamics, Pomona Division - Pomona, California, "Soviet Priz Light Modulator".

June 1982:

21. Eglin Air Force Base - Ft. Walton Beach, Florida, "Optical Patt Recognition Using Synthetic Discriminant Functions for Target Identifi n".

22. Night Vision and Electro Optics Laboratory - Ft. Belvoir, Virginia, "Optical Pattern Recognition: Recent Advances".

23. Kodak - Rochester, New York, "Recent Pattern Recognition and Optical Computing Advances".

July 1982:

24. DARPA Materials Research Council - La Jolla, California, "The U.S. Prom and USSR Priz Spatial Light Modulators".

August 1982:

25. SPIE Conference - San Diego, California, "New Research in Holographic Pattern Recognition".

26. SPIE Conference - San Diego, California, "Synthetic Discriminant Functions for 3-D Robotic Ojbect Recognition".

15.3 THESES SUPPORTED BY AFOSR SUPPORT (SEPTEMBER 1980 - SEPTEMBER 1982)

1. Hiroyasu Murakami, M.S. Dissertation, "Matched Filter Statistical Correlator" (February 1981).

2. Saulius Eiva, M.S. Dissertation, "Image Quality Effects in Optical Correlators" (May 1981).

3. Charles Hester, PhD Dissertation, "Synthetic Filters for Multi-Class Pattern Recognition" (May 1981).

4. Yair Barniv, PhD Dissertation, "Multi-Sensor Image Registration" (May 1981).

5. Mark Carlotto, PhD Dissertation, "Iterative Electro-Optic Matrix Processor" (May 1981).

6. Andrew Sexton, M.S. Dissertation, "Digital Analysis of Space-Variant Optical Processors" (July 1981).

7. Vinod Sharma, "Synthetic Discriminant Functions" (PhD expected in early 1984).

8. John Lycas, "Iterative Optical Processor for Missile Guidance" (Master's expected in 1983).

9. Warren Allmond, "New Holographic Optical Elements and Architectures" (Master's expected in 1983).

10. Eugene Pochapsky, "Digital Preprocessing and Simulation for Optical Pattern Recognition" (Master's expected in 1983).

11. Wen-Thong Chang, "Optical Pattern Recognition" (PhD expected in 1984).

15.4 PATENT DISCLOSURES (SEPTEMBER 1980 - SEPTEMBER 1982)

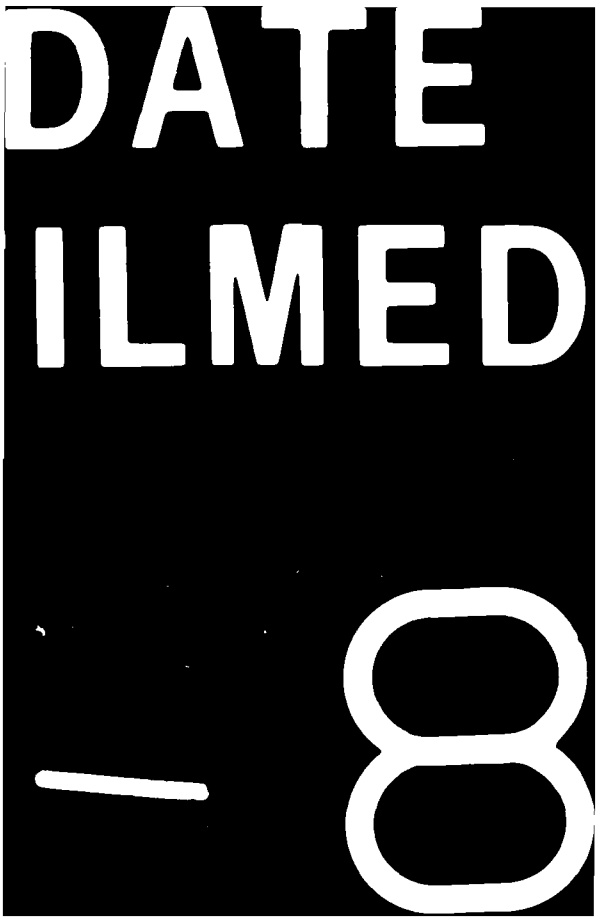
1. Multiple-Invariant Space-Variant Pattern Recognition System.

2. Pattern Recognition by Invariant Moments

3. Synthetic-Discriminant Functions for Multi-Class Pattern Recognition.

4. Equalization and Coherence Measure Correlator.

5. Multi-Variant Technique for Multi-Class Pattern Recognition.


DATE
ILMED
8