科技资料

Intelligent Robots and Computer Vision X: Algorithms and Techniques

PROCEEDINGS

SPIE—The International Society for Optical Engineering

Intelligent Robots and Computer Vision X: Algorithms and Techniques

David P. Casasent
Chair/Editor

11–13 November 1991
Boston, Massachusetts

Sponsored and Published by
SPIE—The International Society for Optical Engineering

Volume 1607

SPIE (Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to the advancement of optical and optoelectronic applied science and technology.

The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:
Author(s), "Title of paper," *Intelligent Robots and Computer Vision X: Algorithms and Techniques,* David P. Casasent, Editor, Proc. SPIE 1607, page numbers (1992).

Library of Congress Catalog Card No. 91-67604
ISBN 0-8194-0744-5

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Fax 206/647-1445

Printed in the United States of America.

CONFERENCE COMMITTEE

Conference Chair
David P. Casasent, Carnegie Mellon University

Program Committee
Rolf-Jürgen Ahlers, Rauschenberger Group (FRG)
Bruce G. Batchelor, University of Wales College of Cardiff (UK)
Stanley M. Dunn, Rutgers University
Ernest L. Hall, University of Cincinnati
James M. Keller, University of Missouri/Columbia
Richard A. Messner, University of New Hampshire
Prasanna G. Mulgaonkar, SRI International
Rosalind W. Picard, Media Laboratory/MIT
Daniel Raviv, Florida Atlantic University
Scott A. Starks, University of Texas/El Paso
Matthew A. Turk, Media Laboratory/MIT

Session Chairs
Session 1—Computer Vision and Pattern Recognition I
David P. Casasent, Carnegie Mellon University

Session 2—Computer Vision and Pattern Recognition II
Matthew A. Turk, Media Laboratory/MIT

Session 3—Computer Vision and Image Processing I
Rosalind W. Picard, Media Laboratory/MIT

Session 4—Computer Vision and Image Processing II
Rosalind W. Picard, Media Laboratory/MIT

Session 5—Image Segmentation I
Stanley M. Dunn, Rutgers University

Session 6—Image Segmentation II
Stanley M. Dunn, Rutgers University

Session 7—Fuzzy Image Processing for Robotics I
James M. Keller, University of Missouri/Columbia

(continued)

INTELLIGENT ROBOTS AND COMPUTER VISION X:
ALGORITHMS AND TECHNIQUES

Volume 1607

Session 8—Fuzzy Image Processing for Robotics II
James M. Keller, University of Missouri/Columbia

Session 9—Software-Oriented Techniques for Autonomous and Robotic Systems
Scott A. Starks, University of Texas/El Paso

Session 10—Image Understanding for Intelligent Robotics
Scott A. Starks, University of Texas/El Paso

Session 11—Intelligent Materials Handling and Vision Systems I
Ernest L. Hall, University of Cincinnati

Session 12—Intelligent Materials Handling and Vision Systems II
Ernest L. Hall, University of Cincinnati

Conference 1607, *Intelligent Robots and Computer Vision X: Algorithms and Techniques,* was part of a four-conference program on Computer Vision held at SPIE's Symposium on Advances in Intelligent Robotic Systems, 10–15 November 1991, in Boston, Massachusetts. The other conferences were:

Conference 1608, *Intelligent Robots and Computer Vision X: Neural, Biological, and 3-D Methods*
Conference 1609, *Model-Based Vision Development and Tools*
Conference 1610, *Curves and Surfaces in Computer Vision and Graphics II*

Program Chair: **Ramon P. DePaula,** NASA Headquarters

INTRODUCTION

The aim of this conference was to reflect the newest research results, trends, and developments in intelligent robots and computer vision. The emphasis in this portion of the tenth conference in this series is algorithms and techniques. See Volume 1608 for papers on other aspects of intelligent robots and computer vision.

This proceedings includes nearly seventy papers from fifteen countries, representing a truly international attendance. The twelve sessions provide groupings of separate topics in computer vision. Session 1 provides a glimpse into various aspects and techniques using pattern recognition. Session 2 emphasizes face recognition algorithms and systems. Image processing is considered in Sessions 3 and 4; this includes feature spaces, edge and contour methods, optical character recognition, and morphology. Session 4 includes papers on the state of the art in packaging, materials handling, software, and manipulation. Segmentation (Sessions 5 and 6) includes fifteen papers, indicating its vital importance and the significant difficulty of this problem. Fuzzy image processing (Sessions 7 and 8) continues to receive much attention and interest, especially with its increasing use in new equipment. Software for robotic systems includes a full session of papers. Image understanding techniques remain a vital methodology for intelligent robotics (Session 10). The last two sessions conclude with a variety of advanced materials handling and vision systems hardware and algorithms.

I thank my administrative assistant, Marlene Layton, and my program committee, plus all session chairs and authors who made this conference the success it was and my job more enjoyable.

David Casasent
Carnegie Mellon University

INTELLIGENT ROBOTS AND COMPUTER VISION X:
ALGORITHMS AND TECHNIQUES

Volume 1607

CONTENTS

(continued)

INTELLIGENT ROBOTS AND COMPUTER VISION X:
ALGORITHMS AND TECHNIQUES

Volume 1607

(continued)

INTELLIGENT ROBOTS AND COMPUTER VISION X:
ALGORITHMS AND TECHNIQUES

Volume 1607

INTELLIGENT ROBOTS AND COMPUTER VISION X: ALGORITHMS AND TECHNIQUES

Volume 1607

SESSION 1

Computer Vision and Pattern Recognition I

Chair
David P. Casasent
Carnegie Mellon University

"DISTORTION-INVARIANT OPTICAL FILTER SETS FOR LARGE CLASS PROBLEMS: OCR CASE STUDY"

DAVID CASASENT

Carnegie Mellon University
Center for Excellence in Optical Data Processing
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213

ABSTRACT

An OCR machine printed problem is selected as an example of a large class pattern recognition problem. We consider discrimination of alpha and numeric fields, recognition of all numbers, recognition of key words (street suffixes, personal titles), state/city/street names etc. These operations are performed on destination address blocks (DABs) in the face of numerous variations in the type face (laser writer, dot matrix, typewriter, etc.), font, data drop out (due to printing errors), point size, ±5° rotations, etc. An optical correlator with banks of distortion invariant hierarchical/inference filters appears to be an ideal adjunct to other OCR techniques (AI, parsing, context, use of lexicons, etc.).

1. INTRODUCTION

We use new distortion-invariant filters [1,2] in an optical correlator (Figure 1). With an entire envelope or a destination address block (DAB) at P_1 and distortion-invariant filters at P_2, we obtain the correlation at P_3. The correlation P_3 contains peaks at the locations of all numbers or key words in the P_1 input. With different laser diodes (LDs) activated at P_0, the system selects different filters at different spatial locations in P_2 (space multiplexing). At different spatial locations in P_2, we store multiple filters (e.g. a set of filters, using frequency-multiplexing) and for each pass through the system we perform four correlations at P_3 in parallel. Analysis of the P_3 outputs determines the next P_2 filters to be accessed (this selection is made by the P_0 LD activated). Alternatively, adaptive filters can be input to P_2 using a spatial light modulator (SLM) as in Figure 2.

Before discussing the role for optics in OCR, we show (Figure 3) one version of an OCR system using our filters. As seen, it uses a hierarchical/inference approach. For simplicity, feedback paths are omitted. DABs are first located (this can be done using our alphanumeric, number, state filters and key word plus context etc. methods). Segmentation is then performed in which the DAB is broken into lines and separate "images" (of words etc.) per line. Parsing is then used to assign "candidate" group types to each digit set (personal or street or city name, zip or street number or secondary (apartment, suite number, etc.)). This is based upon a large database of frequency of words and types at different locations in a DAB. Point size and number of digits per word information is also provided as is the rotation of the input (this is more reliable and useful

 0-8194-0744-5/92/$4.00

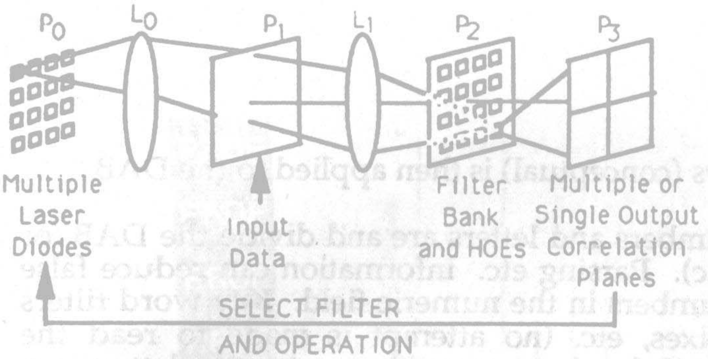


FIGURE 1: Optical correlator with a laser diode accessed fixed-filter bank.

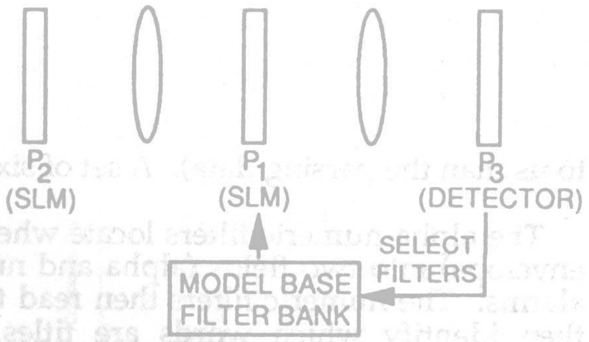


FIGURE 2: Optical correlator with an adaptive P_2 filter.

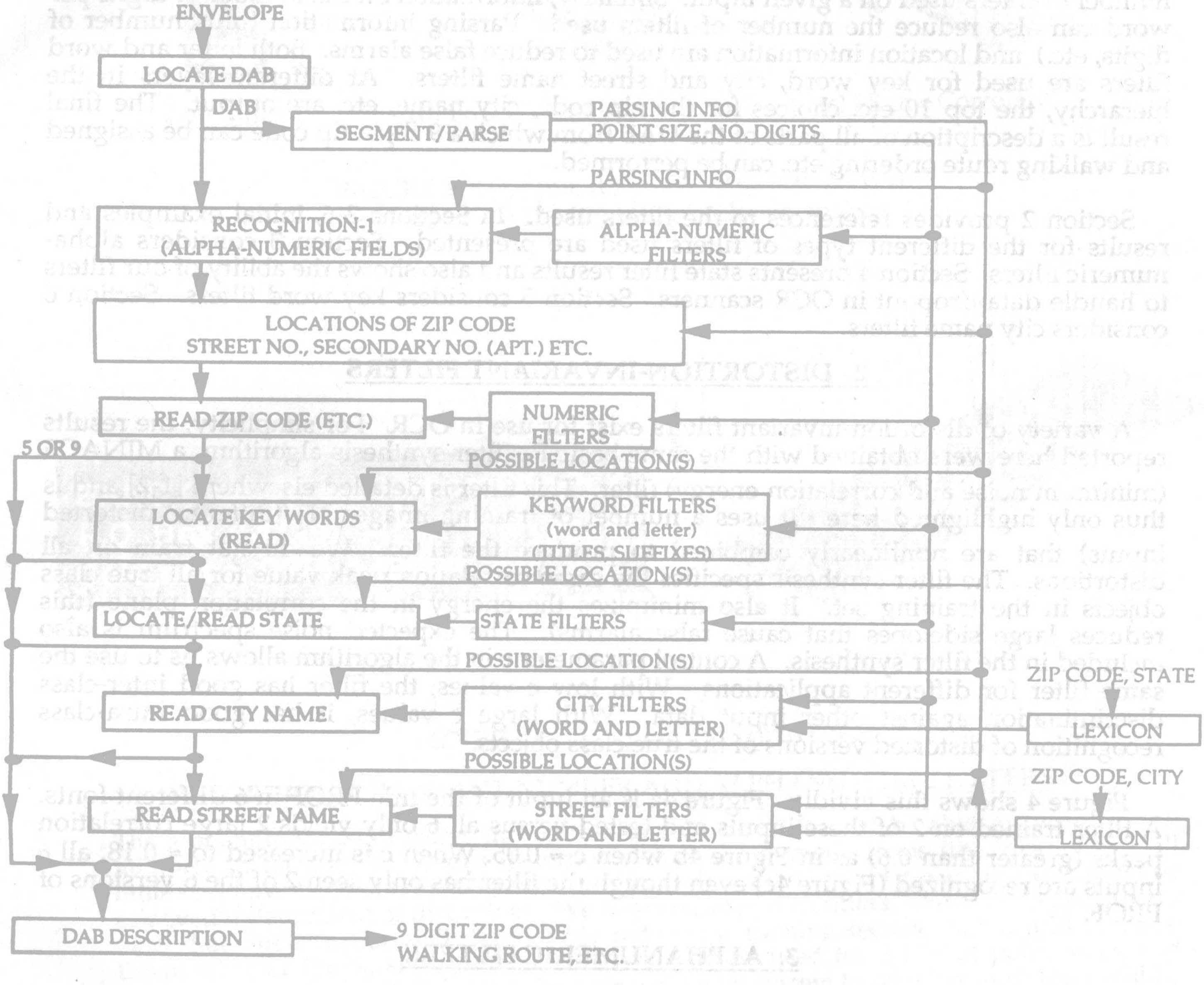


FIGURE 3: Elements of a DAB analysis system.

to us than the parsing data). A set of six filters (conceptual) is then applied to the DAB.

The alpha-numeric filters locate where numbers and letters are and divide the DAB (or envelope) into two fields (alpha and numeric). Parsing etc. information can reduce false alarms. The numeric filters then read the numbers in the numeric field. Key word filters then identify which words are titles, suffixes, etc. (no attempt is made to read the titles/suffixes) using parsing information (142 legal street suffixes with 2-4 letters are possible). The state, city and then the street are identified and read (the zip code etc. information and a lexicon are employed). Information on point size is used to reduce the number of filters used on a given input. Similarly, information on the number of digits per word can also reduce the number of filters used. Parsing information (plus number of digits, etc.) and location information are used to reduce false alarms. Both letter and word filters are used for key word, city and street name filters. At different layers in the hierarchy, the top 10 etc. choices for the zip code, city name, etc. are output. The final result is a description of all parts of the DAB from which a 9 digit zip code can be assigned and walking route ordering etc. can be performed.

Section 2 provides references to the filters used. In Sections 3-6, initial examples and results for the different types of filters used are presented. Section 3 considers alpha-numeric filters. Section 4 presents state filter results and also shows the ability of our filters to handle data dropout in OCR scanners. Section 5 considers key word filters. Section 6 considers city name filters.

2. DISTORTION-INVARIANT FILTERS

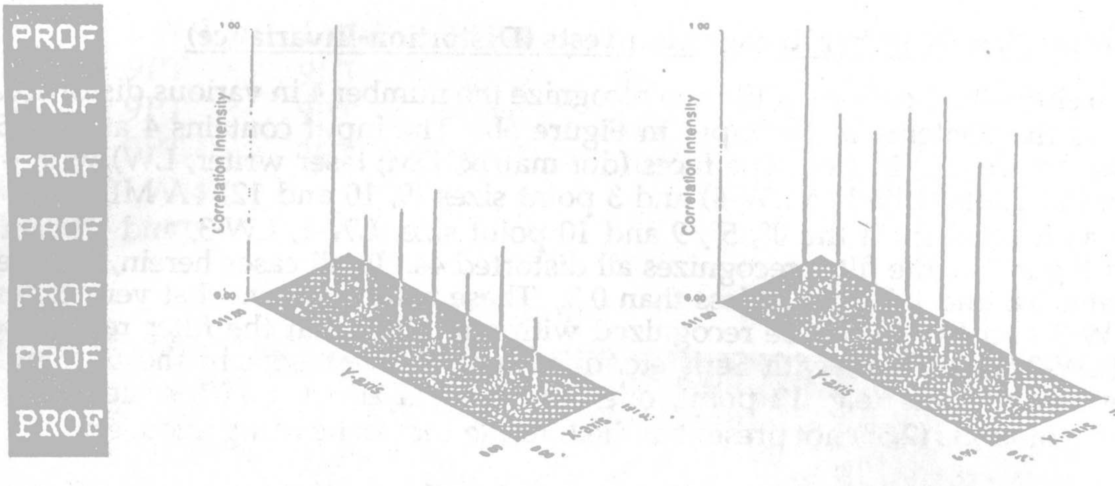
A variety of distortion-invariant filters exist for use in OCR. For simplicity, the results reported here were obtained with the same basic P_2 filter synthesis algorithm, a MINACE (minimum noise and correlation energy) filter. This filter is detailed elsewhere [1-2] and is thus only highlighted here. It uses a number of training images N_T (different distorted inputs) that are nonlinearly combined to produce the filter. We do not train on all distortions. The filter synthesis specifies the same correlation peak value for all true class objects in the training set. It also minimizes the energy in the correlation plane (this reduces large sidelobes that cause false alarms). The expected noise spectrum is also included in the filter synthesis. A control parameter c in the algorithm allows us to use the same filter for different applications. With low c values, the filter has good inter-class discrimination against other input data. With large c values, it has good intra-class recognition of distorted versions of the true class objects.

Figure 4 shows this vividly. Figure 4a is an input of the title PROF in 6 different fonts. A filter trained on 2 of these inputs and tested versus all 6 only yields 2 large correlation peaks (greater than 0.6) as in Figure 4b when $c \approx 0.05$. When c is increased to ≈ 0.18, all 6 inputs are recognized (Figure 4c) even though the filter has only seen 2 of the 6 versions of PROF.

3. ALPHANUMERIC FILTERS

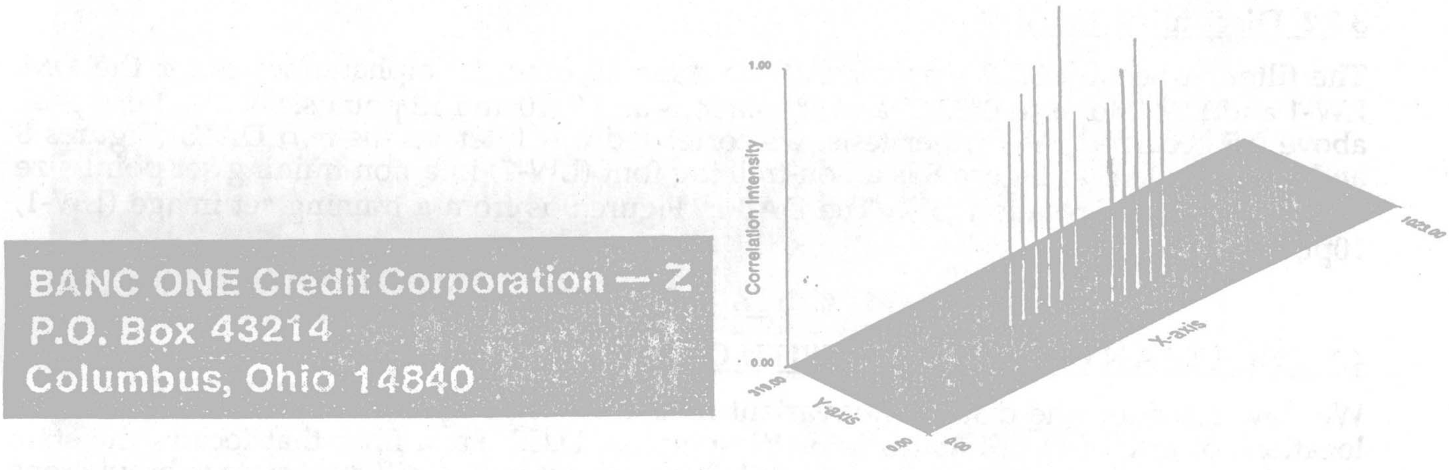
3.1 ALPHANUMERIC FIELD FILTERS

These filters are intended to locate numeric fields and to output a peak of "1" at the location of all numbers in the input. Figure 5 shows a live DAB (altered) and its correlation plane with peaks at the locations of all numbers. Only one false alarm (C in CORPORATION) occurred. This indicates the type of data expected from such filters.



(a) Input (b) Low c Correlation (c) High c Correlation

FIGURE 4: (a) Input, (b) low c ≈ 0.05 MINACE and (c) high c ≈ 0.18 MINACE correlation outputs.



(a) Input (b) Correlation

FIGURE 5: Alphanumeric field results.

3.2 DISTORTION-INVARIANT NUMBER AND ALPHANUMERIC FILTERS

The alphanumeric field filters in Figure 5 and Section 3.1 must be able to recognize all numbers in the face of various distortions. Once the numeric fields (the locations of all numbers) have been obtained (as in Figure 5), each number must be read and this must be achieved independent of distortions. We now consider such filters. We assume no parsing data. Such information will significantly reduce the training set size, but would increase the number of filters, since different filters would be used for different point sizes and rotations. In addition, the time required for the digital system to calculate such data will be much longer than the optical correlation time.

3.2.1 Intra-Class Number Recognition Tests (Distortion-Invariance)

Figure 6 shows test data for a filter to recognize the number 4 in various distortions. Figure 6a shows the contents of the input in Figure 6b. The input contains 4 at 0°, 2.5°, and 5° rotations (rows 1-3) in two type faces (dot matrix, DM; laser writer, LW) with 4 different laser writer fonts (LW-1 to LW-4) and 3 point sizes (9, 10 and 12). A MINACE (c = 0.05) filter was formed from the 0°, 5°, 9 and 10 point size, LW-1, LW-3, and DM images. As seen in Figure 6c, the filter recognizes all distorted 4s. In all cases herein, true peaks must be greater 0.6 and false peaks less than 0.5. These results show: that very different fonts (e.g. LW-3 and LW-1) can be recognized with one filter, that the filter recognizes similar fonts (LW-2 and LW-4) with Serif etc. differences (not present in the training set), that different point sizes (e.g. 12 point, one point size is about a 10% scale difference) and different rotations (2.5°) not present in the training set can be recognized, etc.

In general, one DM and several LW fonts (when the numbers are very different), 2 point sizes, and 0° and 5° rotations are needed in training.

Figure 7 shows the DM and LW-1 font images at 0° to larger size (64×64) to more clearly show differences not apparent in Figure 6.

3.2.2 Discrimination Tests

The filter in Section 3.1.2 was correlated versus all other 35 alphanumerics for the DM, LW-1 and LW-2 fonts at 0°, 2.5° and 5° rotations and 9, 10 and 12 point sizes. No false peak above 0.5 occurred. As further tests, we correlated this filter versus two DABs (Figures 8 and 9). The DAB in Figure 8 is a non-training font (LW-2) in a non-training set point size (12) at the largest rotation (5°). The DAB in Figure 9 is from a training set image (LW-1, 10pt, 0° rotation).

4. STATE FILTERS

4.1 LOCATE ANY STATE NAME WITH ONE FILTER

We now consider one distortion-invariant filter that will output a correlation peak at the locations of any of the 51 states (plus Washington, D.C.), i.e. a filter that locates the state name in an envelope or DAB. Figure 10 shows an input with 6 different state abbreviations and its correlation plane with peaks for all states. Figure 11 shows a DAB and the correlation with the above filter with a peak at the location of the state. The second crosscorrelation peak "MI" in MISS is expected as it is the state abbreviation for Michigan. It is easily ignored using context.

4.2 SCANNER SAMPLING AND DATA DROPOUT VARIATIONS

The results in Figures 10 and 11 (Section 4.1) are more impressive when the sampling etc. differences in the 7 different fonts of Wyoming (WY) that this filter can recognize are seen (Figure 12).

5. KEY WORD (TITLE) FILTERS

This is one filter that will recognize and output a correlation peak at the location of any of a number of key words. We specifically consider different titles (6 different ones) and other variations/distortions).