PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# Design of distortion-invariant correlation filters using supervised learning

Samuel Peter Kozaitis
Rufus H. Cofer
Wesley E. Foor

# Design of distortion-invariant correlation filters using supervised learning

S. P. Kozaitis and R. H. Cofer
Department of Electrical and Computer Engineering
Florida Institute of Technology
Melbourne, FL 32901

W. E. Foor
Photonics Center
Rome Laboratory
Griffiss AFB, NY 113441

## ABSTRACT

We designed binary phase-only filters from a training set of images using a statistical approach. We forced images into clusters and designed filters to recognize objects from that cluster. We report on results obtained by computer simulation comparing the performance of filters to recognize objects from clusters of one and two classes.

## 1.0 INTRODUCTION

Previous attempts at creating distortion-invariant binary phase-only filters (BPOFs) generally do not directly address the need of recognizing objects in the same class as, but outside, the training set. For example, one approach minimizes the mean-squared-error (MSE) between a training set and filter image subject to the constraint that the total MSE be a minimum.[1] However, by examining only maxima of cross-correlation responses, little can be said about the behavior occurring at individual spatial frequencies. Therefore, it may be difficult to identify individual spatial frequencies that may be useful for recognizing objects outside of the training set.

Due to thresholding, a binary imaging system may present problems in the variability of an obtained image. The thresholding process is highly vulnerable to noise and variations of the object and background. The reliability of the thresholding process is critical for object recognition because the binary image contains shape features of the object. The global shape of an object is frequently too perturbed to generate a reliable or specific version of the object. A thresholded version of an image often results in an object that had changed in an unknown or nonrepeatable way. These changes in an object's appearance is enough to significantly degrade the performance of a binary optical correlator.[2]

In contrast to most previous attempts at distortion-invariant BPOF formulation, we developed a filter whose values were determined by features of the training set. We attempted to find a filter that represented the critical characteristics of a cluster of images so that objects outside the training set but in the same class could be identified. Because correlation filters are derived from Fourier transforms of an object, we examined feature extraction in the Fourier domain. We considered the BPOFs of a cluster of images as a set of features to recognize objects and used a statistical approach to examine features that

 0-8194-1195-7/93/$6.00

were invariant with respect to the training set. We retained those Fourier features that were invariant among a training set, and set to zero those that varied using a technique similar to factor analysis[3] to design a ternary filter. Flannery *et al.* showed that a binary spatial light modulator (SLM) could offer an improvement in signal-to-noise ratios of BPOFs by setting the filter to zero at particular spatial frequencies.[4]

The principle components method, which is related to factor analysis, has been used to design correlation filters.[5] Kumar *et al.* maximized the signal-to-noise ratio of an average of an ensemble of images to design a correlation filter. In contrast, we examined an ensemble of BPOFs to select spatial frequencies to recognize images outside of our ensemble. In addition, we formed a cluster from two different classes of images and compared correlation results when a cluster of one class was used.

In the next section we describe our approach for the design of distortion-invariant BPOFs based on clustering. Section 3 describes computer simulations and associated results from comparing two different experiments. Finally, we discuss our conclusions.

## 2.0 STATISTICAL BINARY PHASE-ONLY FILTER

Because BPOFs of objects have been sufficiently descriptive in many binary optical pattern recognition experiments, we considered the BPOFs of training images as a collection of characteristic features from which to perform feature extraction. The Fourier coefficients of the training images were examined over the training set on a pixel-by-pixel basis in terms of their similarity. We considered the difference between Fourier coefficients as a measure of their similarity. The smaller the distance the more similar the coefficients.

Only two possible values exist for an individual pixel of a BPOF. If values of two BPOFs corresponding to the same pixel were the same, then the distance between them was -1; if the values were different then the distance was +1. The distances between coefficients of two BPOFs were described as

$$d_{i,j}[n,m] = H_i[n,m] \oplus H_j[n,m], \tag{1}$$

where $H_i[n, m]$, and $H_j[n, m]$ represent two BPOFs, $n$ and $m$ are discrete spatial frequencies, and $\oplus$ is the exclusive-OR function. If the coefficients at each pixel of the BPOFs were +1 or -1, and we used these values in Eq. (1) to replace the usual logical values 1 and 0 respectively, then $d_{i,j}[n,m]$ was represented as an array of $\pm 1$'s.

Given a training set, we measured the distances between coefficients of BPOFs of each training image and every other training image. We used a criterion function[6] to measure the quality of the cluster of distances as

$$J[n,m] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{i,j}[n,m], \tag{2}$$

where $N$ is the number of images in the training set and a more compact cluster was indicated by a more negative value. Equivalently, $J[n,m]$ could be represented by the function $\bar{J}[n,m]$ as

$$\bar{J}[n,m] = \sum_{i=1}^{N} H_i[n,m], \tag{3}$$

where $H_i[n,m]$ is the BPOF of the $i$th training image and a more compact cluster was indicated by a larger magnitude.

Features may often offer conflicting information as to class association. For example, small magnitudes of $\bar{J}[n,m]$ show that some coefficients are not consistent over a training set. In factor analysis, the goal is to account for a process by a small number of variates.[3] Using this methodology, we retained spatial frequencies that were more compactly clustered than others. Therefore, we retained spatial frequencies that were invariant over the training set. In other words, we eliminated the use of spatial frequencies whose coefficients had large distances over a training set. The resulting filter was described as

$$G_p[n,m] = \begin{Bmatrix} 1 \mathit{if}\ \bar{J}[n,m] \geq p \\ -1 \mathit{if}\ \bar{J}[n,m] \leq -p \\ 0\ \mathit{otherwise} \end{Bmatrix}, \tag{4}$$

where $p$ is a threshold that determines the order of the filter.

The number of possible different filters depends on the number of images in the training set. The variable $p$ could take values from $0 \leq p \leq N$. To describe different filters with each value of $p$, and to avoid choosing a value for $G_p[n,m]$ when $\bar{J}[n,m]$ is 0, then $p$ should be odd if $N$ is odd, and $p$ should be even if $N$ is even. The number of different possible filters for a given $N$ was written as

$$N_{filter} = \begin{cases} \frac{N}{2} & N\ even \\ \frac{N}{2} + \frac{1}{2} & N\ odd \end{cases}. \tag{5}$$

We primarily considered three measures of performance, the SNR, peak-to-correlation energy (PCE), and signal-to-clutter ratio (S/C).[7] We wrote our expression for the SNR in a discretized form using our filter as

$$SNR = \frac{\left|\sum S[n,m]\, G_p[n,m]\right|^2}{\sum P_n[n,m]\left|G_p[n,m]\right|^2}, \tag{6}$$

where $S[n,m]$ is the DFT of an input signal, and $P_n[n,m]$ denotes the power spectral density of the noise and is a constant for white noise. The PCE can be used as a measure of the sharpness of the correlation peak and can be expressed as

$$PCE = \beta \frac{\left|\sum S[n,m]\, G_p[n,m]\right|^2}{\sum |S[n,m]|^2 |G_p[n,m]|^2}. \qquad (7)$$

Usually, the correlation height is the factor that determines whether an object produces a match with a known filter. Therefore, the ratio of lowest in-class correlation peak to the highest out-of-class correlation peak is important and is referred to as the S/C.

## 3.0 EXPERIMENTS AND RESULTS

We initially created three filters from fifteen thresholded images of the same class. We determined the cross-correlation response between our filters and the training set, ten additional images not in the training set but in the same class, and five out-of-class set of images. In addition, we created three filters made from ten training set images and the five out-of-class images. In this way we forced two different classes of images into clusters. To examine the effects of unknown variations of the inputs, we used objects of the same scale and rotation that did not show appreciable cross-correlation with each other. The imagery used was infrared (IR) (8-14 μm) of ground scenes from actual sensors. Images were digitized with 128 x 128 pixels with 8 bits/pixel and thresholding was performed by choosing a single threshold value for the entire image. Fig. 1 shows five images that were representative of our first training set and Fig. 2 shows the five images that were initially out-of-class.

The results of our simulations are shown in Tables 1 and 2 for three different values of $p$. The number of transmitting pixels for the same value of $p$ for the two experiments were different. Training set 1 consisted of only one class and had a higher percentage of transmitting pixels than training set 2, the training set consisting of two classes. This shows that training set 1 had more common features than training set 2. The normalized correlation heights of the training sets were also considered. The results were normalized by setting the maximum correlation height for each filter equal to 100. Although the normalized correlation heights increased with increasing $p$, the correlation heights were lower for training set 2. The SNR generally decreased for filters of the same value of $p$ when training set 2 was used when compared to training set 1. Other performance measures are also shown in Tables 1 and 2.

## 4.0 DISCUSSION AND CONCLUSIONS

We could arbitrarily form clusters from training images using our statistical approach for the design of correlation filters. Although spatial frequencies that are invariant to a cluster may be identified, their correlation response may not have a large SNR. Therefore, we would expect an SNR optimization approach to improve performance.

## 5.0 ACKNOWLEDGMENTS

Most of this work was performed in the Photonics Center of Rome Laboratory at Griffiss AFB, NY.

## 6.0 REFERENCES

[1] D. A. Jared and D. J. Ennis, "Inclusion of filter modulation in synthetic discriminant function construction," *Applied Optics* 28(2), 232-239 (1989)

[2] S. P. Kozaitis and W. E. Foor, "Performance of synthetic discriminant functions for binary phase-only filtering of thresholded imagery," *Optical Engineering* 31(4) 830-837 (1992)

[3] D. N. Lawley, and M. A. Maxwell, Factor Analysis as a Statistical Method, Butterworths: London (1963)

[4] D. L. Flannery, J. S. Loomis, and M. E. Milkovich, "Transform-ratio ternary phase-amplitude formulation for improved correlation discrimination," *Applied Optics* 27(19), 4079-4083 (1988)

[5] B. V. K. Vijaya Kumar, D. Casasent, and A. Murakami, "Principle-component imagery for statistical pattern recognition correlators," *Optical Engineering* 21(1), 43-47 (1982)

[6] R. O. Duda, and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley:New York (1973)

[7] B. V. K. Vijaya Kumar, and L. Hasselbrook, "Performance measures for correlation filters," *Applied Optics* 29(20), 2997-3006 (1990)

**TABLE 1. Results of experiments for filters based on training set 1**

| | $p = 1$ | $p = 9$ | $p = 15$ |
|---|---|---|---|
| Avg. normalized correlation height of training set | 75.1 | 83.9 | 83.6 |
| Avg. normalized correlation height of in-class set | 45.5 | 71.6 | 76.6 |
| Avg. normalized correlation height of out-of-class set | 11.0 | 15.1 | 21.4 |
| Percentage of transmitting pixels | 100 | 17.4 | 5.6 |
| Avg. SNR of training set | 52.5 | 157 | 228 |
| Avg. SNR of in-class set | 31.8 | 134 | 210 |
| Avg. S/C | 7.9db | 7.3db | 5.8db |
| Avg. PCE of training set | 130 | 94.7 | 67.4 |

**TABLE 2. Results of experiments for filters based on training set 2**

| | $p = 1$ | $p = 9$ | $p = 15$ |
|---|---|---|---|
| Avg. normalized correlation height of training set | 49.5 | 56.3 | 57.4 |
| Avg. normalized correlation height of in-class set | 43.6 | 72.0 | 73.6 |
| Percentage of transmitting pixels | 100 | 11.1 | 0.65 |
| Avg. SNR of training set | 56.1 | 68.3 | 16.3 |
| Avg. SNR of in-class set | 32.5 | 77.6 | 16.5 |
| Avg. PCE of training set | 95.3 | 51.2 | 5.4 |

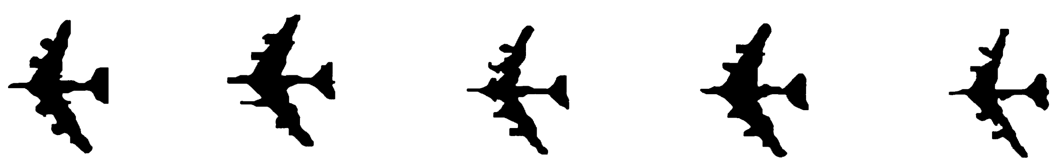

**FIGURE 1. Five images used in training set from one class**

**FIGURE 2. Images used initially for out-of-class set**