Unconstrained correlation filters

Abhijit Mahalanobis, B. V. K. Vijaya Kumar, Sewoong Song, S. R. F. Sims, and J. F. Epperson

A mathematical analysis of the distortion tolerance in correlation filters is presented. A good measure for distortion performance is shown to be a generalization of the minimum average correlation energy criterion. To optimize the filter's performance, we remove the usual hard constraints on the outputs in the synthetic discriminant function formulation. The resulting filters exhibit superior distortion tolerance while retaining the attractive features of their predecessors such as the minimum average correlation energy filter and the minimum variance synthetic discriminant function filter. The proposed theory also unifies several existing approaches and examines the relationship between different formulations. The proposed filter design algorithm requires only simple statistical parameters and the inversion of diagonal matrices, which makes it attractive from a computational standpoint. Several properties of these filters are discussed with illustrative examples.

1. Introduction

Ever since the first use of the optical correlator for implementing matched spatial filters by VanderLugt,[1] researchers have been trying to develop better filters for the recognition of shapes and objects. Such filters are popularly referred to as correlation filters since they are designed for implementation in correlators. A notable contribution to this field was the formulation of the synthetic discriminant function[2] (SDF), which permitted several references or training patterns to be included in the same filter. SDF's are not the only type of correlation filters to be developed in recent years. Among the filters proposed in the past decade are circular-harmonic filters[3] which offer in-plane rotation invariance, and the phase-only filter[4] and its variations, which are well suited for implementation on existing spatial light modulators. However, our concern in this paper is limited to SDF-type filters,[5] which are useful for the recognition of objects in the presence of arbitrary distortions caused by changes in viewing angle, scale, and orientation. Since no restrictions are placed on the magnitude or the phase function of the fully complex SDF's in the frequency domain, all degrees of freedom are available to optimize the performance of these filters. Consequently SDF's may serve as benchmarks for other types of correlation filters that are designed to accommodate restrictions of the optical systems.

The three primary issues to deal with in the design of correlation filters are (i) the ability to suppress clutter and noise, (ii) easy detection of the correlation peak, and (iii) distortion tolerance. Earlier SDF's[5,6] were unable to offer optimum performance with regard to any of these criteria. The first breakthrough in the rigorous development of SDF theory came with the advancement of the minimum variance synthetic discriminant function[7] (MVSDF) by Kumar. The MVSDF was shown to be the optimum filter for minimizing the effects of additive noise. Although the MVSDF was impractical because of the need for inverting a large covariance matrix, the optimization approach opened the doorway for the advancement of SDF theory. Kumar and Bahri[8] described a general framework that encompassed the entire space of possible solutions for SDF's. Following this work, Mahalanobis *et al.* proposed the minimum average correlation energy[9] (MACE) filter, which produced

A. Mahalanobis is with the Hughes Missile Systems Company, P.O. Box 11337, Building 802, Mail Stop 2, Tucson, Arizona 85734; B. V. K. Vijaya Kumar is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213; S. Song is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, California 93107; S. R. F. Sims is with the Research, Development and Engineering Center, U.S. Army Missile Command, Redstone Arsenal, Alabama 35898; J. F. Epperson is with the Department of Mathematical Sciences, University of Alabama in Huntsville, Huntsville, Alabama 35898.

Received 29 March 1993; revised manuscript received 15 November 1993.
0003-6935/94/173751-09$06.00/0.
© 1994 Optical Society of America.

sharp peaks for easy detection of the filter output. The MACE filter is generally known to be sensitive to distortions but readily able to suppress clutter.[10] Another interesting idea was combining the properties of the MACE and the MVSDF filters by deriving a joint optimization criterion.[11] Recently, optimal trade-off filters have been proposed by Réfrégier[12] to combine the properties of various SDF's. The minimum noise and average correlation energy filter proposed by Ravichandran and Casasent[13] is also a promising method for combining noise tolerance and peak detection abilities in a filter.

The SDF-type filters noted above offer either sharp correlation peaks or superior noise tolerance or a combination of both. However, none of the filters are explicitly optimized to offer good distortion tolerance. Indeed, the selection of a particular training set determines the spectral characteristics of the filter, which influences its ability to tolerate distortions. In general, it was observed that filters that produce broader correlation peaks (such as the early SDFs[14,15]) offer better distortion tolerance. However, they may also provide poorer discrimination between classes since these filters tend to correlate broadly with low-frequency information in which the classes may be difficult to separate. Distortion tolerance may also improve when the number of images in the training set is increased. This represents one method for reducing the sensitivity of the MACE filter. However, filter synthesis becomes computationally difficult, and a sufficiently large number of training images may not be available.

In this paper we describe a different approach to filter design that removes hard constraints on the training images and explicitly optimizes a criterion for distortion tolerance. Most of the previous research in SDF design has automatically assumed that the correlation values at the origin are prespecified. There is no need for such a constraining assumption. In fact, once we realize that these prespecified values are designated only for training images and not for test images, the justification for using this assumption decreases even more. Thus by removing the hard constraints, we increase the number of possible solutions, thus improving the chances of finding a filter with better performance. In addition, the filters can be designed to offer good performance in the presence of noise and background clutter while maintaining relatively sharp correlation peaks for easy detection of the output. Thus the three important criteria noted above for the synthesis of correlation filters can be addressed simultaneously. Section 2 describes the theoretical development of the proposed new filters. It may be desirable in some applications to restrict the variation in the filter output, and a method for achieving this is described in Section 3. A procedure for optimally combining distortion tolerance with noise and clutter suppression for the proposed filter is discussed in Section 4. In Section 5 we discuss the effects of training the filter using the phase-only versions of the training images. As will be shown, this approach offers a number of advantages that may be useful for some applications. Specific examples in Section 6 illustrate the performance of the filters. Finally, concluding remarks are given in Section 7.

2. Derivation of the Filter Equation

The primary objective of correlation filters is distortion-tolerant recognition of objects in clutter. This problem is easier to solve for in-plane rotations and scale changes, and several solutions have been proposed.[3,16] However, the prevalent method for handling out-of-plane distortions is to use a training set of representative views of the object. While this makes good intuitive sense, it is unclear how the training set should impart useful information to the filter. Traditionally in the design of SDF-type correlation filters, linear constraints are imposed on the training images to yield a known value at specific locations in the correlation plane. However, placing such constraints in the correlation plane satisfies conditions only at isolated points in the image space but does not explicitly control the filter's ability to generalize over the entire domain of the training images. Various filters exhibit different levels of distortion tolerance even with the same training set and constraints.

In this paper we adopt a statistical approach to filter design. We show that, in addition to yielding sharp peaks and being computationally simple, the proposed filters offer improved distortion tolerance. The reason lies in the fact that we do not treat training images as deterministic representations of the object but as samples of a class whose characteristic parameters should be used in encoding the filter. We assume that the training set consists of N images, each of the true and the false class, and that each image of size $d_1 \times d_2$ contains $d = d_1 d_2$ pixels. In general, the number of images for the true and the false classes can be different. The ith training image for the true class is denoted by $x_i(m, n)$ and is represented in the frequency domain by a $d \times 1$ vector $\mathbf{x}_i$, obtained by lexicographically reordering its two-dimensional discrete Fourier transform, $X_i(k, l)$. Similarly, the false-class training images, $y_i(m, n)$, are represented in vector notation in the frequency domain as $\mathbf{y}_i$. We denote the filter by the $d \times 1$ vector $\mathbf{h}$. We then obtain the two-dimensional filter $H(k, l)$ by rearranging $\mathbf{h}$ into a two-dimensional image. Matrices are denoted by upper-case and vectors by lower-case characters.

The correlation of the ith training image and the filter can be expressed in the frequency domain as

$$\mathbf{g}_i = \mathbf{X}_i \mathbf{h}, \tag{1}$$

where $\mathbf{X}_i$ is a $d \times d$ diagonal matrix containing the elements of $\mathbf{x}_i$. Here, $\mathbf{g}_i$ denotes the discrete Fourier transform of the ith correlation output. The deviation in the shape of the correlation plane with respect to some ideal shape vector $\mathbf{f}$ is quantified by the average squared error (ASE), defined as

$$\mathrm{ASE} = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{g}_i - \mathbf{f})^+(\mathbf{g}_i - \mathbf{f}). \tag{2}$$

Thus ASE is a measure of distortion with respect to reference shape $\mathbf{f}$, which is free to be chosen as desired.

In fact, shape vector $\mathbf{f}$ can be treated as a free parameter in the distortion minimization problem. In the design of the minimum-squared-error (MSE) SDF,[17] $\mathbf{f}$ was specified as Gaussian or ringlike shapes in order to sculpt the correlation surface into these forms. Here, we are interested in the choice of $\mathbf{f}$ that causes least variation among the correlation planes and offers minimum ASE. To find the optimum shape $\mathbf{f}_{\mathrm{opt}}$, we set the gradient of ASE with respect to $\mathbf{f}$ to zero and obtain

$$\nabla_{\mathbf{f}}(\mathrm{ASE}) = \frac{2}{N}\sum_{i=1}^{N} (\mathbf{g}_i - \mathbf{f}) = \mathbf{0} \tag{3}$$

or

$$\mathbf{f}_{\mathrm{opt}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i = \bar{\mathbf{g}}, \tag{4}$$

where

$$\bar{\mathbf{g}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{g}_i = \frac{1}{N}\sum_{i=1}^{N} \mathbf{X}_i\mathbf{h} = \bar{\mathbf{X}}\mathbf{h} \tag{5}$$

is the average correlation plane and $\bar{\mathbf{X}} = (1/N)\sum_{i=1}^{N} \mathbf{X}_i$ is the average training image expressed as a diagonal matrix. We thus see that of all possible reference shapes, using the average correlation plane $\bar{\mathbf{g}}$ offers the smallest possible ASE and hence the least distortion (in the squared-error sense) among the correlation planes.

Accordingly, we substitute $\mathbf{f} = \bar{\mathbf{g}}$ in the ASE expression and obtain the average similarity measure (ASM), defined below as

$$\begin{aligned}\mathrm{ASM} &= \frac{1}{N}\sum_{i=1}^{N} (\mathbf{g}_i - \bar{\mathbf{g}})^+(\mathbf{g}_i - \bar{\mathbf{g}}) \\ &= \frac{1}{N}\sum_{i=1}^{N} (\mathbf{X}_i\mathbf{h} - \bar{\mathbf{X}}\mathbf{h})^+(\mathbf{X}_i\mathbf{h} - \bar{\mathbf{X}}\mathbf{h}) \\ &= \mathbf{h}^+\left[\frac{1}{N}\sum_{i=1}^{N} (\mathbf{X}_i - \bar{\mathbf{X}})^*(\mathbf{X}_i - \bar{\mathbf{X}})\right]\mathbf{h} \\ &= \mathbf{h}^+\mathbf{S}_x\mathbf{h},\end{aligned} \tag{6}$$

where

$$\mathbf{S}_x = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{X}_i - \bar{\mathbf{X}})^*(\mathbf{X}_i - \bar{\mathbf{X}}) \tag{7}$$

is a diagonal matrix measuring the similarity of the training images to the class mean in the frequency domain. For example, if all training images are identical, then $\mathbf{S}_x$ would an all-zero matrix. From Parseval's theorem it is easy to show that the average squared distance from the correlation planes to their mean is the same as that defined by Eq. (6) in the frequency domain.

ASM is one possible metric for distortion since it represents the average deviation of the correlation planes from the mean correlation shape, $\bar{\mathbf{g}}$. ASM is also a measure of the compactness of the class. If filter $\mathbf{h}$ is viewed as a linear transform, then ASM measures the distances of the training images from the class center under this transform. Minimizing ASM therefore leads to a compact set of correlation planes that resemble each other and exhibit the least possible variations. The distortions of the object in the input plane are represented by the training images, $\mathbf{x}_i$. These distortions are reflected in the output as variations in the structure and shape of the corresponding correlation planes, $\mathbf{g}_i$, and are quantified by ASM. If the filter successfully reduces the distortions, then distorted input images should yield similar output planes, leading to a small value of ASM. Conversely, if ASM is minimum and $\bar{\mathbf{g}}$ is well shaped by design, then all true-class correlation planes are expected to resemble $\bar{\mathbf{g}}$ and to exhibit well-shaped structures.

We now formulate our filter design criteria. We do not constrain the peak of the average correlation plane to a prespecified value. Instead, we try to make it as large as possible. Specifically, the intensity of the peak of the average correlation plane may be written as

$$|\bar{g}(0, 0)|^2 = (\mathbf{h}^+\bar{\mathbf{x}})^2 = \mathbf{h}^+\bar{\mathbf{x}}\bar{\mathbf{x}}^+\mathbf{h}. \tag{8}$$

Here, we assume without loss of generality that the peak occurs at the origin of the correlation plane. We explicitly optimize the behavior (e.g., maximize the peak value) of the average correlation plane while minimizing ASM. If $\bar{\mathbf{g}}$ exhibits a sharp well-defined peak with low sidelobes, then minimizing ASM should cause the rest of the class to also follow this behavior and to exhibit similarly shaped correlation planes. Therefore the criteria we optimize to improve distortion tolerance is

$$J(\mathbf{h}) = \frac{\mathbf{h}^+\bar{\mathbf{x}}\bar{\mathbf{x}}^+\mathbf{h}}{\mathbf{h}^+\mathbf{S}_x\mathbf{h}} \tag{9}$$

and is referred to as the average correlation height criterion. The filter of interest maximizes this criterion and thus is called a maximum average correlation height (MACH) filter. The MACH filter maximizes the relative height of the average correlation peak with respect to the expected distortions. The MACH filter yields a high correlation peak in response to $\bar{\mathbf{x}}$ and ensures that other samples of the true class exhibit similar behavior.

Before deriving the filter that maximizes $J(\mathbf{h})$, we state the considerations for the second class. We want to reject the false class by suppressing its average correlation energy (ACE), given by

$$\mathrm{ACE} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}^{+}\mathbf{Y}_i\mathbf{Y}_i^{*}\mathbf{h} = \mathbf{h}^{+}\left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{Y}_i\mathbf{Y}_i^{*}\right)\mathbf{h} = \mathbf{h}^{+}\mathbf{D}_y\mathbf{h}, \tag{10}$$

where $\mathbf{Y}_i$ is a diagonal matrix containing the elements of $\mathbf{y}_i$ and $\mathbf{D}_y = (1/N)\sum_{i=1}^{N} \mathbf{Y}_i\mathbf{Y}_i^{*}$ is the diagonal matrix containing the average power spectrum of the false class.

It should be noted that the defintion of correlation energy is not restricted to the false class only. In fact, the correlation energy for the true class can be defined as $\mathbf{h}^{+}\mathbf{D}_x\mathbf{h}$, where $\mathbf{D}_x = (1/N)\sum_{i=1}^{N} \mathbf{X}_i\mathbf{X}_i^{*}$. The original formulation of the MACE filter uses $\mathbf{D} = \mathbf{D}_x + \mathbf{D}_y$ (i.e., the average correlation energy over the entire training set), but trivial variations are possible by use of either $\mathbf{D}_x$ or $\mathbf{D}_y$ separately. The choice of terms for the correlation energy in MACE and similar filters depends on the nature of the application.

Here, we require the filter to reject the $\mathbf{Y}_i$ images by reducing the false-class ACE. This is done by maximizing the modified performance criterion

$$J'(\mathbf{h}) = \frac{\mathbf{h}^{+}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h}}{\mathbf{h}^{+}\mathbf{S}_x\mathbf{h} + \mathbf{h}^{+}\mathbf{D}_y\mathbf{h}} = \frac{\mathbf{h}^{+}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h}}{\mathbf{h}^{+}(\mathbf{S}_x + \mathbf{D}_y)\mathbf{h}}. \tag{11}$$

At first glance, the expression in Eq. (11) may appear similar to the Fisher ratio. However, the Fisher ratio applies to values at only one point in the correlation plane (presumably the peak) but does not control the rest of the correlation plane. The denominator of Eq. (11) measures similarity and energy over the entire correlation plane and has significantly different physical interpretation from the terms in the Fisher ratio. The latter uses full covariance matrices, whereas $\mathbf{D}_y$ and $\mathbf{S}_x$ are both diagonal. The proposed approach does not suffer from the drawbacks that render the Fisher criterion impractical for image-processing applications. The Fisher ratio and $J'(\mathbf{h})$ are merely similar in mathematical form, better known as the Rayleigh quotient, whose optimization leads to a very different solution in each case.

Since maximizing $J'(\mathbf{h})$ results in a small denominator, the filter will reduce both ASM and ACE, as desired. The optimum filter is found by setting the gradient of $J'(\mathbf{h})$ with respect to $\mathbf{h}$ to zero, or

$$\nabla_{\mathbf{h}}[J'(\mathbf{h})] = 2\frac{\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h}}{\mathbf{h}^{+}(\mathbf{S}_x + \mathbf{D}_y)\mathbf{h}} - 2\frac{(\mathbf{h}^{+}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h})(\mathbf{S}_x + \mathbf{D}_y)\mathbf{h}}{[\mathbf{h}^{+}(\mathbf{S}_x + \mathbf{D}_y)\mathbf{h}]^2} = \mathbf{0}. \tag{12}$$

This can be simplified to

$$\frac{1}{\mathbf{h}^{+}(\mathbf{D}_y + \mathbf{S}_x)\mathbf{h}}[\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h} - \lambda(\mathbf{D}_y + \mathbf{S}_x)\mathbf{h}] = \mathbf{0}, \tag{13}$$

where

$$\lambda = \frac{\mathbf{h}^{+}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h}}{\mathbf{h}^{+}(\mathbf{S}_x + \mathbf{D}_y)\mathbf{h}} \tag{14}$$

is a scalar identical to $J'(\mathbf{h})$. Equating the term inside the brackets in Eq. (13) to zero, we get

$$\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h} - \lambda(\mathbf{D}_y + \mathbf{S}_x)\mathbf{h} = \mathbf{0}, \tag{15}$$

or

$$(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h} = \lambda\mathbf{h}. \tag{16}$$

In going from Eq. (15) to Eq. (16), we have assumed that $(\mathbf{D}_y + \mathbf{S}_x)$ is invertible. Otherwise, Eq. (15) represents a generalized eigenvalue problem. If $(\mathbf{D}_y + \mathbf{S}_x)$ is invertible, $\mathbf{h}$ must be an eigenvector of $(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}$ with corresponding eigenvalue λ. Since λ is identical to $J(\mathbf{h})$, as shown in Eq. (14), we select the eigenvector corresponding to the largest value of λ to maximize $J(\mathbf{h})$. In general, a full-rank $d \times d$ matrix has d nonzero eigenvalues. However, since $\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}$ is the outer product of a vector, it is of unit rank, and $(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}$ has only one nonzero eigenvalue. The corresponding eigenvector is then the obvious choice for the optimum filter and can be found by substituting $\overline{\mathbf{x}}^{+}\mathbf{h} = \alpha$ (a scalar) in Eq. (16) so that

$$\alpha(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}} = \lambda\mathbf{h}, \tag{17}$$

or

$$\mathbf{h} = c(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}}, \tag{18}$$

where $c = \alpha/\lambda$ is a scale factor. It is easy to see from Eq. (11) that the scale factor c has no effect on the performance criterion $J'(\mathbf{h})$. Thus the MACH filter in Eq. (18) is clearly proportional to $(\mathbf{D}_y + \mathbf{S}_x)^{-1}\overline{\mathbf{x}}$, the transformed true-class mean image.

Interestingly the MACH filter depends only on simple class statistics such as the mean image, the average power spectrum matrix, and the ASM matrix. Matrix inversions are simple since both $\mathbf{D}_y$ and $\mathbf{S}_x$ are diagonal. In this sense the MACH filter is easier to compute than the MACE filter since the latter requires the inversion of a full $N \times N$ matrix. Further, only the mean and second-order statistics of each class need to be stored for on-line filter synthesis, which may be useful when dealing with large data bases of training images. The necessary statistics can be estimated off-line with an adequate number of class samples.

An interesting property of the MACH filter is that sharp peaks are obtained for true-class images even though their correlation energy is not explicitly minimized. The reason can be understood by expanding the ASM expression as

$$\begin{aligned}\text{ASM} &= \mathbf{h}^+\left[\frac{1}{N}\sum_{i=1}^{N}(\mathbf{X}_i - \overline{\mathbf{X}})^*(\mathbf{X}_i - \overline{\mathbf{X}})\right]\mathbf{h} \\ &= \mathbf{h}^+\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i^*\right)\mathbf{h} - \mathbf{h}^+\overline{\mathbf{X}}\overline{\mathbf{X}}^*\mathbf{h} \\ &= \mathbf{h}^+\mathbf{D}_x\mathbf{h} - \mathbf{h}^+\overline{\mathbf{X}}\overline{\mathbf{X}}^*\mathbf{h} = \text{ACE} - \mathbf{h}^+\overline{\mathbf{X}}\overline{\mathbf{X}}^*\mathbf{h}. \end{aligned} \quad (19)$$

Clearly, ASM includes the ACE term $\mathbf{h}^+\mathbf{D}_x\mathbf{h}$, and therefore its minimization influences the correlation energies of the true-class images. In fact, the minimization of ASM can be viewed as a generalization of the MACE criterion. If $\mathbf{h}^+ \overline{\mathbf{X}}\overline{\mathbf{X}}^*\mathbf{h}$ is small, then ASM $\approx$ ACE, and the performances of filters based on the two criteria are comparable.

3. Relation Between Maximum Average Correlation Height, Minimum-Squared-Error Synthetic Discriminant Function, and Minimum Average Correlation Energy Filters

We now illustrate the direct relation of the MACH filter to the MSE SDF and the MACE filter. If the correlation energy term $\mathbf{D}_y$ is deleted from the expression of the MACH filter, we obtain

$$\mathbf{h}_{\text{mach}} = \mathbf{S}_x^{-1}\overline{\mathbf{x}}. \quad (20)$$

Let $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \mathbf{x}_3 \ldots \mathbf{x}_N]$ be the $d \times N$ data matrix containing the N training vectors as its columns. It is easy to show that the expression for the MSE SDF[17] using the optimum shape in Eq. (4) for sculpting the correlation surface is

$$\begin{aligned}\mathbf{h}_{\text{mse}} &= \mathbf{S}_x^{-1}\mathbf{X}(\mathbf{X}^+\mathbf{S}_x^{-1}\mathbf{X})^{-1}\mathbf{u} \\ &= \mathbf{S}_x^{-1}\mathbf{X}\mathbf{a} \\ &= \mathbf{S}_x^{-1}\hat{\mathbf{x}}, \end{aligned} \quad (21)$$

where $\hat{\mathbf{x}}$ is a weighted average of the training images and $\mathbf{a} = (\mathbf{X}^+\mathbf{S}_x^{-1}\mathbf{X})^{-1}\mathbf{u}$ is the weight vector necessary to satisfy the hard constraints on the training images. From Eqs. (20) and (21) we see that the expression for the MACH filter closely resembles that for the MSE SDF when the correlation energy term is dropped. The latter is a constrained version of the former and is obtained by minimizing the same criterion, namely, ASM. Therefore the special MACH filter in Eq. (20) without the correlation energy term is referred to as the unconstrained MSE SDF. As shown in Eq. (19), the minimization of the ASM criterion indirectly influences the ACE criterion. Consequently the unconstrained MSE SDF is also expected to yield well-defined peaks even though the filter is based on the minimum ASM criterion, and, not explicitly, on MACE.

The conventional MACE filter is also related to the MACH filter in a similar way. Assuming the definition of correlation energy in terms of the $\mathbf{X}$ data (see Section 2), the filter expression after dropping the ASM term $\mathbf{S}_x$ becomes

$$\mathbf{h}_{\text{mach}} = \mathbf{D}_x^{-1}\overline{\mathbf{x}}. \quad (22)$$

The MACE filter[9] expression is

$$\begin{aligned}\mathbf{h}_{\text{mace}} &= \mathbf{D}_x^{-1}\mathbf{X}(\mathbf{X}^+\mathbf{D}_x^{-1}\mathbf{X})^{-1}\mathbf{u} = \mathbf{D}_x^{-1}\mathbf{X}\mathbf{b} \\ &= \mathbf{D}_x^{-1}\hat{\mathbf{x}}, \end{aligned} \quad (23)$$

where $\hat{\mathbf{x}}$ is again the weighted average image with weights $\mathbf{b} = (\mathbf{X}^+\mathbf{D}_x^{-1}\mathbf{X})^{-1}\mathbf{u}$ chosen to satisfy hard constraints on the training images. As evident in Eqs. (22) and (23), the MACH filter is of the same form as the MACE filter when the ASM term is dropped. Both filters minimize the same criterion (namely, ACE) although the latter does so under constraints. The special MACH filter in Eq. (22) obtained by dropping the ASM term is therefore referred to as the unconstrained MACE filter or the UMACE filter.

4. Controlling Variance of Correlation Peaks

The MACH filter described in Section 2 is designed to permit variations in the correlation-peak values. The ASM term is defined over the entire correlation plane and controls the overall similarity of their shapes but is not a good measure for tightly controlling the variations at one point (e.g., the origin). In some applications it may be necessary to reduce the variations in the correlation peaks significantly. Linear constraints make the peak values identical (resulting in zero variation) but unnecessarily restrict the filters. The approach proposed here is a compromise between the usual method of hard constraints and the flexible optimization described in Section 2.

The correlation output at the origin that is due to the ith training image is

$$c_i = \mathbf{h}^+\mathbf{x}_i, \qquad i = 1, 2, \ldots, N. \quad (24)$$

The average correlation output at the origin is

$$\bar{c} = \frac{1}{N}\sum_{i=1}^{N} c_i = \mathbf{h}^+\overline{\mathbf{x}}, \quad (25)$$

where $\overline{\mathbf{x}}$ is the average training image. The variance in the correlation outputs is then given by

$$\begin{aligned}\sigma^2 &= \frac{1}{N}\sum_{i=1}^{N} |c_i - c|^2 \\ &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{h}^+(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^+\mathbf{h} \\ &= \mathbf{h}^+\, \Sigma_x\, \mathbf{h}, \end{aligned} \quad (26)$$

where

$$\Sigma_x = \frac{1}{N}\sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^+ \quad (27)$$

is an estimate of the class covariance matrix. Here, Σ_x is a $d \times d$ matrix and not necessarily of diagonal form. Since Σ_x is computed from N outer products and since the average vector $\bar{\mathbf{x}}$ is subtracted from all $\mathbf{x}_i$ values, the rank of Σ_x will be at most $N - 1$. Since $d \gg N$ in general, this matrix is singular. However, the MACH performance criterion can be successfully modified with σ^2, as we now show.

The modified criterion that includes the variations in the peak values is defined as

$$J''(\mathbf{h}) = \frac{|\bar{c}|^2}{\sigma^2 + \mathrm{ACE}} = \frac{\mathbf{h}^+\overline{\mathbf{x}}\overline{\mathbf{x}}^+\mathbf{h}}{\mathbf{h}^+\,\Sigma_x\,h + h^+\mathbf{D}_y\mathbf{h}}. \tag{28}$$

The filter that maximizes $J''(\mathbf{h})$ in Eq. (28) yields a large peak in response to average image $\bar{\mathbf{x}}$ while reducing the variations in the output at the origins of the correlation planes produced by the N training images. Sharp peaks are expected as a result of this optimization since the average correlation energy is also reduced. As shown in Section 2, the optimum filter is found by setting the gradient of $J''(\mathbf{h})$ with respect to $\mathbf{h}$ to zero, which leads to

$$\overline{\mathbf{x}}\overline{\mathbf{x}}^+\mathbf{h} = (\Sigma_x + \mathbf{D}_y)\mathbf{h}. \tag{29}$$

Thus $J''(\mathbf{h})$ is maximized by choosing $\mathbf{h}$ to be the dominant eigenvector of $(\Sigma_x + \mathbf{D}_y)^{-1}\overline{\mathbf{x}}\overline{\mathbf{x}}^+$ [assuming that $(\Sigma_x + \mathbf{D}_y)$ is invertible], which is

$$\mathbf{h} = c(\Sigma_x + \mathbf{D}_y)^{-1}\bar{\mathbf{x}}, \tag{30}$$

where the scale factor c does not affect $J''(\mathbf{h})$. The filter expression in Eq. (30) is referred to as the generalized MACH (GMACH) filter. Since Σ_x is a full matrix, computing the inverse in Eq. (30) can be cumbersome. Fortunately, since Σ_x is computed from N outer products, the required inverse can be computed in N steps.[18]

Here, we have described a process using σ^2 instead of ASM for controlling the variations in the correlation-peak values. ASM characterizes how similar the correlation surfaces are, whereas σ^2 characterizes how similar correlation-peak values are. Although the resulting matrices are not diagonal, the filter can be computed with an efficient algorithm.

5. Optimal Trade-Off Design of Minimum-Average Correlation Height Filters

An important consideration in the design of correlation filters is the trade-off between various performance criteria.[10,11] Réfrégier has suggested an optimal approach[12] to trade off one performance measure against another. Here, we describe the procedure for trading off between the ASM, ACE, and σ^2 criteria in the design of the MACH filter.

Let us suppose we wish to design $\mathbf{h}$ to maximize $|c|^2$ while holding ASM and ACE at fixed values. The statement of the problem would be

$$\underset{\mathbf{h}}{\mathrm{Maximize}}\ \mathbf{h}^+\overline{\mathbf{x}}\overline{\mathbf{x}}^+\mathbf{h}$$

subject to $\mathbf{h}^+\mathbf{S}_x\mathbf{h} = \delta$ and $\mathbf{h}^+\mathbf{D}_y\mathbf{h} = E_0$, where δ and E_0 are scalars. We form the Lagrange functional

$$\phi(\mathbf{h}) = \mathbf{h}^+\overline{\mathbf{x}}\overline{\mathbf{x}}^+\mathbf{h} - \alpha(\mathbf{h}^+\mathbf{S}_x\mathbf{h} - \delta) - \beta(\mathbf{h}^+\mathbf{D}_y\mathbf{h} - E_0). \tag{31}$$

Setting the gradient of $\phi(\mathbf{h})$ with respect to $\mathbf{h}$ to zero yields

$$(\overline{\mathbf{x}}\overline{\mathbf{x}}^+ - \alpha\mathbf{S}_x - \beta\mathbf{D}_y)\mathbf{h} = \mathbf{0}, \tag{32}$$

or

$$(\alpha\mathbf{S}_x + \beta\mathbf{D}_y)\mathbf{h} = \bar{\mathbf{x}}(\bar{\mathbf{x}}^+\mathbf{h}), \tag{33}$$

or

$$\mathbf{h} = k(\alpha\mathbf{S}_x + \beta\mathbf{D}_y)^{-1}\bar{\mathbf{x}}. \tag{34}$$

Absorbing constants together and setting $\gamma = \beta/\alpha$, we see that the optimal filter for fixing the performance criteria at specific values is of the form

$$\mathbf{h} = c(\mathbf{S}_x + \gamma\mathbf{D}_y)^{-1}\bar{\mathbf{x}}. \tag{35}$$

Here, γ is a user-selected parameter. Small γ leads to more similar correlation planes but with larger background intensities, whereas larger values of γ may yield more fluctuations in the correlation plane shape but low correlation energies. Essentially, setting $\alpha = 0$ in Eq. (33) [or γ large in Eq. (34)] yields the UMACE filter described in Section 3. Similarly choosing $\gamma = \beta = 0$ leads to the unconstrained MSE SDF, also described in Section 3. The MACH filter can be made to trade off between these two filters by variation of the value of γ.

6. Amplitude-Normalized Filter Designs

A number of interesting simplifications occur if the Fourier transforms of the training images are amplitude normalized. This is the same as using phase-only training images. In general the Fourier transform of a training image is complex and can be represented as

$$X_i(k, l) = A(k, l)\exp[j\phi(k, l)], \tag{36}$$

where $A(k, l) = |X_i(k, l)|$ is the magnitude of the Fourier transform and $\phi(k, l)$ is its phase. We normalize each training image such that $A(k, l) = 1$ for all k and l, which yields the amplitude-normalized training images, given by

$$\hat{X}_i(k, l) = \exp[j\phi(k, l)]. \tag{37}$$

The $\hat{X}_i(k, l)$ images essentially retain only phase information. The amplitude-normalized images are lexicographically reordered as usual to construct the vectors $\mathbf{x}_i$, and the filter is designed as outlined in

Sections 2–5. We now discuss the simplifications as a result of this process.

First, we note that the correlation matrix $\mathbf{D}_y$ reduces to the identity matrix, $\mathbf{I}$. Therefore the correlation matrix becomes class independent, and the correlation energy expression is always the same for all classes, including those not included in the training set. The simplified expression for the average correlation energy is

$$\mathrm{ACE} = \mathbf{h}^{+}\mathbf{I}\mathbf{h} = \mathbf{h}^{+}\mathbf{h}. \tag{38}$$

The expression in Eq. (38) holds true for all images, including background clutter and noise. Thus identity matrix $\mathbf{I}$ serves as a generic model for the power spectrum for all classes when magnitude normalization is used. Therefore it becomes unnecessary to include images of a specific false class when optimizing the simplified MACH criterion

$$J' = \frac{h^{+}\overline{\mathbf{x}}\overline{\mathbf{x}}^{+}\mathbf{h}}{\mathbf{h}^{+}(\mathbf{S}_x + \delta\mathbf{I})\mathbf{h}}. \tag{39}$$

The MACH filter is then given by

$$\mathbf{h} = (\mathbf{S}_x + \delta\mathbf{I})^{-1}\overline{\mathbf{x}}. \tag{40}$$

This filter can be expected to reject all images and patterns that do not belong to the true class since their correlation energies are all described by the same expression when magnitude normalization is used. The true-class information is contained in the ASM expression, which unlike the ACE remains class dependent. Magnitude normalization can therefore be important in applications in which the false classes are too many to specifically describe or are unknown or when every occurrence of background clutter cannot be modeled. In all such cases the correlation energy information is captured by the generic model in Eq. (38) and minimized by the filter expression in Eq. (40).

Another interesting simplification results in the expression of the MACE and UMACE filters. Since $\mathbf{D}_y = \mathbf{I}$, the MACE filter becomes the straightforward projection SDF based on the magnitude-normalized images, i.e.,

$$\mathbf{h}_{\mathrm{mace}} = \mathbf{X}(\mathbf{X}^{+}\mathbf{X})^{-1}\mathbf{u}. \tag{41}$$

The conventional MACE filter whitens the training images on the average, which would be the same as magnitude normalization for the special case of a single image in the training set. The processing described here is an extension that whitens each training image individually before synthesizing the filter. Therefore although the expression in Eq. (41) is simply an ordinary projection SDF, it is expected to yield sharp correlation peaks. Further, when the same processing is applied to the UMACE filter, its expression reduces to

$$\mathbf{h}_{\mathrm{umace}} = \overline{\mathbf{x}} \tag{42}$$

since $\mathbf{D}_y = \mathbf{I}$. Thus the remarkable conclusion that the optimum filter is the average of the training images is inevitable when magnitude normalization is used. The theoretical framework of MACH filters presented in this paper supports the optimality of this simple solution, which may be useful in a number of applications.

7. Simulation Results

We used a data base of 35 tank images (each having 128×128 pixels) for our numerical tests. These images were taken at 10° intervals in aspect. Example broadside and end-on views are shown in Figs. 1(a) and 1(b), respectively. All filters were synthesized by use of segmented training images with background removed. All tests were done on images containing the target in the background. Unless stated otherwise, the odd-numbered images (i.e., image 1, 3, . . . , 35) are used as training images, and all 35 images are used as test images.

The performance evaluation of the correlation filters is done as follows. When the input is the centered target image, we want the output correlation to peak at the origin. As the target moves in the

(a)

(b)

Fig. 1. Target in background at (a) broadside view and (b) end view.

Table 1. Performance[a] of Unconstrained Correlation Filters and Improvements Compared with the Minimum Average Correlation Energy Filter

Filter Type	Hits ($PE \leq 2$)	Near Hits ($2 < PE \leq 10$)	Misses ($PE > 10$)	Hits and Near Hits for Nontraining Images	Average PE of Misses
Minimum average correlation energy	21	5	9	9	64.8
Unconstrained minimum average correlation energy	22	7	6	12	32.1
Generalized maximum average correlation height	22	7	6	12	32.2
Maximum average correlation height	22	8	5	13	25.2

[a]The number of hits, near hits, or misses depends on the peak error (PE).

input scene, the correlation peak moves the same way. However, the correlation peak may not be at the origin because of distortions, noise, and other impairments. The Euclidean distance from the origin to the actual correlation-peak position is called the peak error (PE). Peaks with $PE \leq 2$ are treated as hits, and peaks with $2 < PE \leq 10$ are considered near hits.

The image data base was partitioned into broadside and end-view groups. Thus one training set and the corresponding filter considered only side views, whereas the other training set and its corresponding filter dealt with only front and back views. The test images were correlated with both filters, and the higher of the two correlation peaks was used to determine the PE.

In the first test all 35 images were used for training, and the same 35 were used for testing the filters. As expected, the performance was very good. All four filters tested (i.e., MACE, UMACE, MACH, and GMACH) resulted in 34 hits (i.e., $PE \leq 2$). The one image that was not properly recognized contained a sensor flaw in the data.

The above tests used the same data for training and testing, leading to optimistic results. In the next test we used only half the set (i.e., every other image) for training, and we tested on all 35 images. The results are shown in Table 1. In this table we show the performance results for the four filters of interest. We show for each filter the number of hits (i.e., $PE \leq 2$), the number of near hits (i.e., $2 < PE \leq 10$), and the number of misses ($PE > 10$). In addition, we show the total number of nontraining images leading to hits and near hits. This number is indicative of the distortion tolerance of the filter being considered. The last column lists average PE, obtained by averaging the PE for the misses.

Amplitude normalization was used for the UMACE, MACH, and GMACH filters. Their performance is compared with the standard MACE filter. In Table 1 all new filters show improved performance (i.e., a smaller number of misses, lower average PE, and better distortion tolerance). Among the three new filters the MACH filter leads to the smallest number of misses, the lowest PE, and the highest distortion tolerance (as measured by the number of hits and near hits in the nontraining set). Thus MACH appears to provide the best performance. However, the other new filters (UMACE and GMACH) also offer comparable performances.

6. Conclusion

The original MACE filter was the first SDF-type filter to control explicitly the shape of the correlation plane. This was achieved by minimizing correlation energy subject to linear constraints. The MACE filter demonstrated the use of diagonal matrices in the frequency domain (which are easy to compute and to invert) for practical filter synthesis. In spite of these advances the MACE filter lacked good distortion tolerance in general. To overcome this, we have generalized the concept of the average correlation energy (ACE) to a metric called the average similarity measure (ASM) [see Section 2, Eq. (19)] in this paper.

The ASM criterion is the first explicit mathematical treatment of distortion tolerance in correlation filters. Its optimization leads to the MACH filter, which minimizes distortions in a mean-square-error sense. This results in the performance improvements shown in Table 1. Other variations such as the UMACE and GMACH filters also exhibit better performance than the original MACE design.

We observed that amplitude normalization helps correlation performance, although perhaps marginally in some cases. Partitioning the data by aspect provides useful MACH filter performance. Including widely varying references in the training set degrades filter performance in general. In certain situations the simple UMACE filter can give performance comparable to the MACH filter.

In summary, the filter design techniques presented here relax the constraints on the correlation peak while directly optimizing a criterion for distortion tolerance. The resulting filters are easy to compute and work well. We also outlined the method for optimal trade-off between the different performance criteria in the design of MACH filters, and we introduced the idea of amplitude normalization to combat the effects of unknown clutter.

References

1. A. B. VanderLugt, "Signal detection by complex matched spatial filtering," IEEE Trans. Inf. Theory **IT-10**, 139–145 (1964).
2. C.F. Hester and D. Casasent, "Multivariant technique for multiclass pattern recognition," Appl. Opt. **19**, 1758–1761 (1980).
3. Y. N. Hsu and H. H. Arsenault, "Optical pattern recognition using the circular harmonic expansion," Appl. Opt. **21**, 4016–4019 (1982).
4. J. L. Horner and P. D. Gianino, "Phase-only matched filtering," Appl. Opt. **23**, 2324–2335 (1984).

5. B. V. K. Vijaya Kumar, "Tutorial survey of composite filter designs for optical correlators," Appl. Opt. **31,** 4773–4801 (1992).
6. D. Casasent, "Unified synthetic discriminant function computational formulation," Appl. Opt. **23,** 1620–1627 (1984).
7. B. V. K. Vijaya Kumar, "Minimum variance synthetic discriminant functions," J. Opt. Soc. Am. A **3,** 1579–1584 (1986).
8. Z. Bahri and B. V. K. Vijaya Kumar, "Generalized synthetic discriminant functions," J. Opt. Soc. Am. A **5,** 562–571 (1988).
9. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent, "Minimum average correlation energy filter," Appl. Opt. **26,** 3633–3640 (1986).
10. A. Mahalanobis and D. Casasent, "Performance evaluation of minimum average correlation energy filters," Appl. Opt. **30,** 561–572 (1991).
11. S. I. Sudharsanan, A. Mahalanobis, and M. K. Sundareshan, "A unified framework for the synthesis of synthetic discriminant functions with reduced noise variance and sharp correlation structure," Opt. Eng. **29,** 1021–1028 (1990).
12. Ph. Réfrégier and J. Figue, "Optimal trade-off filter for pattern recognition and their comparison with Weiner approach," Opt. Computer Process. **1,** 3–10 (1991).
13. G. Ravichandran and D. Casasent, "Minimum noise and correlation energy optical correlation filter," Appl. Opt. **31,** 1823–1833 (1992).
14. D. Casasent and W. T. Chang, "Correlation synthetic discriminant functions," Appl. Opt. **25,** 2343–2350 (1986).
15. D. Casasent, G. Ravichandran, and S. Bollapragada, "Gaussian minimum average correlation energy filters," Appl. Opt. **30,** 5176–5181 (1991).
16. A. Mahalanobis, D. Casasent, and B. V. K. Vijaya Kumar, "Spatial-temporal correlation filter for in-plane distortion invariance," Appl. Opt. **25,** 4466–4472 (1986).
17. B. V. K. Vijaya Kumar, A. Mahalanobis, S. Song, S. R. F. Sims, and J. F. Epperson, "Minimum squared error synthetic discriminant functions," Opt. Eng. **31,** 915–922 (1992).
18. G. W. Stewart, *An Introduction to Matrix Computations* (Academic, Orlando, Fla., 1973).