Recent advances in correlation filter theory and applications

B. V. K. Vijaya Kumar[a]*, Joseph A. Fernandez[a], Andres Rodriguez[b], Vishnu Naresh Boddeti[c]
[a]Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA
[b]Sensors Directorate, US Air Force Research Laboratory (AFRL), Dayton, OH
[c]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

Advanced correlation filters (CFs) were introduced over three decades ago to offer distortion-tolerant object recognition and are used in applications such as automatic target recognition (ATR) and biometric recognition. Some of the advances in CF design include minimum average correlation energy (MACE) filters that produce sharp correlations and offer excellent discrimination, optimal tradeoff synthetic discriminant function (OTSDF) filters that allow the filter designer to control the tradeoff between peak sharpness and noise tolerance, maximum average correlation height (MACH) filter that removes correlation peak constraints to reduce filter design complexity and quadratic correlation filters (QCFs) that extend the linear CFs to include second-order nonlinearity. In this paper, we summarize two recent major advances in CF design. First is the introduction of maximum margin correlation filters (MMCFs) that combine the excellent localization properties of CFs with the very good generalization abilities of support vector machines (SVMs). Second is the introduction of zero-aliasing correlation filters (ZACFs) that eliminate the aliasing in CF design due to the circular correlation caused by the use of discrete Fourier transforms (DFTs).

Keywords: correlation filters, support vector machine, aliasing, circular correlation, face recognition, iris recognition, automatic target recognition

1. INTRODUCTION

Correlation filters (CFs) have been successfully applied to a variety of pattern recognition applications[32] including automatic target recognition (ATR) and biometric recognition. ATR is a critical aspect of numerous applications including guided weapons, autonomous surveillance, and unmanned vehicles. Biometrics are characteristics that differ from person to person (or group of persons), and are important in many security applications such as accessing a secure building or system, identifying a person of interest in a scene, and determining whether a person's actions present a threat. The term "correlation filter" is in part due to the history of using optical correlators to obtain cross-correlations. A review of optical correlators can be found elsewhere.[32]

CFs can be designed to yield correlation peaks for each target of interest in the scene while exhibiting low response to clutter and background. Attractive properties such as shift-invariance, noise robustness, graceful degradation, and efficient implementation make CFs well-suited for ATR and biometric applications. In the past decade, CFs have been used in a variety of applications including target recognition,[9,20,21] target tracking,[2,8] face recognition,[26,27,35] face localization,[3] face tracking,[2] biometric encoding,[1] fingerprint recognition,[33] iris recognition,[30] ocular recognition,[22] pedestrian localization,[4] and pedestrian action recognition.[12,24]

In the CF approach, a carefully designed template (loosely called a "filter" and in some literature a "sliding window template") $h(m,n)$ is cross-correlated with a query image $x(m,n)$ to produce the output $g(m,n)$. This operation is efficiently carried out in the frequency domain,

$$G(u,v) = X(u,v)H^*(u,v), \tag{1}$$

where $*$ denotes the conjugate, and $G(u,v)$, $X(u,v)$ and $H(u,v)$ are the 2-D discrete Fourier transforms (DFTs) of the correlation output, the query image, and the template, respectively. When the query image is an authentic match (also called true-class or Class-1) $g(m,n)$ should exhibit a sharp peak at the center of the biometric signature's location, and when the query image is an impostor (also called false-class or Class-2) $g(m,n)$ should

*kumar@ece.cmu.edu; phone 1-412-268-3026

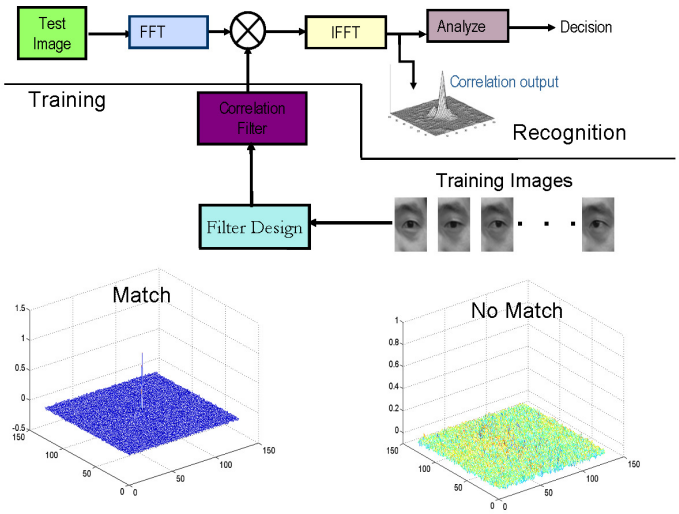


Figure 1. A filter is designed using ocular images of the same subject. A test image is correlated with the filter. If the test image (e.g., a face) is from the same subject, the correlation plane should exhibit a sharp peak at the ocular region, otherwise the correlation plane should not have any significant peak.

not have any significant peak. The higher the correlation peak, the higher the probability that the image is authentic. This peak or some other relevant metric such as the peak-to-correlation energy (PCE) or peak-to-sidelobe ratio (PSR) is used to determine whether the query image is from the authentic class or not. The location of the peak indicates the location of the target in the image. Thus CFs offer the ability to simultaneously localize and classify an object of interest. Figure 1 shows a diagram summarizing the overall training and testing stages. Throughout this paper we used the following terminology: *target* means the object of interest or biometric signature; *classification* means finding the class-label of the target; *localization* means finding the target's location in a given scene (we do not assume that the targets are centered); and *recognition* means both classification and localization.

This paper is organized as follows. In Section 2, we present the background of CFs and highlight important CF designs including minimum average correlation energy (MACE) filters that produce sharp correlations and offer excellent discrimination, optimal tradeoff synthetic discriminant function (OTSDF) filters that allow the filter designer to control the tradeoff between peak sharpness and noise tolerance, maximum average correlation height (MACH) filter that removes correlation peak constraints to reduce filter design complexity and quadratic correlation filters (QCFs) that extend the linear CFs to include second-order nonlinearity. We summarize two recent major advances in CF design: in Section 3 we summarize maximum margin correlation filters (MMCFs) that combine the excellent localization properties of CFs with the very good generalization abilities of support vector machine (SVM), and in Section 4 we summarize zero-aliasing correlation filters (ZACFs) that eliminate the aliasing in the CF design due to the circular correlation caused by the use of discrete Fourier transforms (DFTs). We offer some concluding remarks in Section 5.

Throughout this paper, we use the following notational conventions:

- lower case non bold letters, e.g., $x_i(m,n)$, denote spatial domain 2-D arrays (also called planes or images)
  - we loosely use $x_i(m,n)$ to represent both image i and the (m,n) pixel value of image i
- upper case non bold letters, e.g., $X_i(u,v)$, denote frequency domain 2-D arrays
- lower case bold letters denote vectors (lexicographically arranged) of 2-D frequency domain arrays, e.g, $\mathbf{x}_i$ denotes $X_i(u,v)$
  - we loosely refer to $\mathbf{x}_i$ as the ith image when we mean the vectorized 2-D DFT representation of the ith image

  - we loosely refer to $\mathbf{g}$ as the correlation plane when we mean the vectorized 2-D DFT representation of the correlation plane
  - we loosely refer to $\mathbf{h}$ as the template (or filter) when we mean the vectorized 2-D DFT representation of the template

- lower case bold letters with an inverted hat denote vectors of 2-D spatial domain arrays, e.g, $\check{\mathbf{x}}_i$ denotes $x_i(m,n)$
- upper case bold letters denote matrices, e.g., $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$
  - some matrices are diagonal matrices representations of vectors, e.g., $\mathbf{X}_i = \text{diag}(\mathbf{x}_i)$
- the $\otimes$ symbol denotes the 2-D correlation operator of either two 2-D arrays, e.g., $g_i(m,n) = x_i(m,n) \otimes h(m,n)$, or of the implied 2-D arrays represented by their vector versions, e.g., $\mathbf{g} = \mathbf{x}_i \otimes \mathbf{h}$
- the $\odot$ symbol denotes the Hadamard product, e.g., $\mathbf{x}_i \odot \mathbf{h}$
- the $\oslash$ symbol denotes the Hadamard divide, e.g., $\mathbf{g}_i \oslash \mathbf{x}_i$
- the overbar symbol denotes the mean of a set, e.g., $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$
- the overdot symbol denotes the desired output, e.g., $\dot{\mathbf{g}}_i$ denotes the desired $\mathbf{g}_i$
- the superscript T symbol denotes the transpose, e.g., $\mathbf{h}^T$
- the superscript asterisk symbol denotes the conjugate, e.g., $\mathbf{X}_i^*$
- the superscript dagger symbol denotes the conjugate transpose, e.g., $\mathbf{h}^\dagger$

2. CORRELATION FILTER BACKGROUND

Most CFs are templates carefully designed using a set of training images that captures the expected distortions (in testing). The minimum average correlation energy (MACE) filter[13] is designed to reduce the energy of the correlation output resulting in a sharp peak at the location of the target facilitating target recognition. The optimal-tradeoff synthetic discriminant function (OTSDF) filter[18] is an extension of the MACE filter to achieve robustness to additive noise. MACE and OTSDF are constrained to have a certain value for the inner product between the training image and the filter. This inner or dot product value is referred to as the value at the origin (for centered images) or the correlation peak in the correlation plane (also known as the correlation output). For example the correlation peak can be constrained to be 1 for the authentic images and 0 for impostor images so that in testing the filter produces a high value (near 1) for authentic images and a low value (near 0) for impostor images.

Another advance in CF designs was removing the correlation peak constraints. Removing these constraints increases the solution space and may improve the chances of finding a filter with better recognition performance. The family of unconstrained CFs include the maximum average correlation height (MACH) filter,[15] the unconstrained MACE (UMACE) filter,[15] and the minimum output sum of squared error (MOSSE) filter.[4] Another advance were quadratic correlation filters (QCFs)[9] that determine and use a quadratic nonlinearity to maximize a metric of separation between authentic and impostor classes.

2.1 Minimum average correlation energy (MACE) filter

The MACE filter[13] is the first filter designed to control the shape of the correlation plane $g_i(m,n)$ and not just the peak value $\mathbf{h}^\dagger\mathbf{x}_i$. The correlation output shape is controlled by minimizing the average correlation energy (ACE) of the correlation plane from the training images while simultaneously satisfying the correlation outputs $\mathbf{h}^\dagger\mathbf{x}_i$ to yield a pre-specified value u_i for all $i = \{1, \ldots, N\}$, where N is the number of training images.

The MACE filter facilitates recognition by producing very sharp delta-function-like peaks with minimum sidelobes for authentic class training images and no such sharp peaks for imposter training images. However the

recognition performance decreases for non-training intra-class images.[5,13] In practice many images of interest have strong low frequency components. Since the MACE filter effectively whitens the spectrum (on the average), it enhances high frequency components. It is therefore sensitive to distortions, i.e., to images outside of the training set, as well as to additive (high frequency) noise. In practice, although this filter has been successfully used in pattern recognition applications,[28] variations of this filter are more robust (e.g., the OTSDF[18] discussed below).

Savvides and Kumar[27] showed that the MACE filter can be efficiently trained online without inverting a matrix at each iteration, thus reducing the computational requirement. The design is able to train online and adapt to varying data streams caused by changes in illuminations, backgrounds, and/or different views (e.g., due to rotation, scale, pose, non-rigid deformation, etc).

Boddeti, Su, and Kumar[1] used a modified form of the MACE filter for biometric encryption. The traditional MACE filter produces a sharp peak at the correlation output of centered images. The modified MACE filter produces multiple peaks at different locations by adjusting the phase of the images in the frequency domain. This filter is used for face verification and the peak locations are used to encode a secret key of the authorized user.

2.1.1 Derivation

The hard constraints are given by $\mathbf{h}^\dagger \mathbf{x}_i = u_i$, where $\mathbf{x}_i$ represents the ith image, $\mathbf{h}$ represents the filter, and u_i is the pre-specified peak filter response. The correlation plane in response to image $\mathbf{x}_i$ is represented by $\mathbf{g}_i = \mathbf{X}_i^* \mathbf{h}$, where diagonal matrix $\mathbf{X}_i$ contains $\mathbf{x}_i$ along its diagonal. The energy of the ith correlation output is $E_i = \frac{1}{d}\mathbf{g}_i^\dagger \mathbf{g}_i$, where d is the dimension of $\mathbf{g}_i$ and $\frac{1}{d}$ is a scalar that accounts for the fact that inner products in the space domain are scaled by $\frac{1}{d}$ in the frequency domain. Since all the E_i $\{i = 1, \ldots, N\}$ cannot be simultaneously minimized subject to $\mathbf{X}^\dagger \mathbf{h} = \mathbf{u}$, the ACE is minimized instead. The ACE can be expressed as

$$\begin{aligned} E_{avg} &= \frac{1}{N}\sum_{i=1}^{N} E_i = \frac{1}{Nd}\sum_{i=1}^{N} \mathbf{g}_i^\dagger \mathbf{g}_i = \frac{1}{Nd}\sum_{i=1}^{N} \mathbf{h}^\dagger \mathbf{X}_i \mathbf{X}_i^* \mathbf{h} \\ &= \frac{1}{Nd}\sum_{i=1}^{N} \mathbf{h}^\dagger \mathbf{D}_i \mathbf{h} = \mathbf{h}^\dagger \left(\frac{1}{Nd}\sum_{i=1}^{N} \mathbf{D}_i \right) \mathbf{h} = \mathbf{h}^\dagger \mathbf{D} \mathbf{h}, \end{aligned} \tag{2}$$

where diagonal matrix $\mathbf{D}_i = \mathbf{X}_i \mathbf{X}_i^*$ contains the power spectrum of $\mathbf{x}_i$ along its diagonal, and diagonal matrix $\mathbf{D} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{D}_i$ contains the average power spectral density of the training images along its diagonal.

The quadratic $\mathbf{h}^\dagger \mathbf{D} \mathbf{h}$ is minimized subject to the linear constraints $\mathbf{X}^\dagger \mathbf{h} = \mathbf{u}$ when

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{X}(\mathbf{X}^\dagger \mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{u}. \tag{3}$$

2.2 Optimal tradeoff synthetic discriminant function (OTSDF) filter

The OTSDF filter was introduced in 1990.[18,19] The average correlation energy (ACE) (see Eq. 2) is denoted by $E_1 = \mathbf{h}^\dagger \mathbf{D} \mathbf{h}$ and the output noise variance (ONV) is denoted by $E_2 = \mathbf{h}^\dagger \mathbf{P} \mathbf{h}$, where diagonal matrix $\mathbf{P}$ contains the power spectral density of the input noise along the diagonal (usually approximated by the identity matrix, i.e., white input noise is assumed).[31] Minimizing E_1 typically leads to high-frequency emphasizing filters whereas minimizing E_2 typically leads to low-frequency emphasizing filters. Thus, minimizing one criterion significantly deteriorates the performance from the point of view of the other criterion. An optimal filter is defined such that for a given value of E_1, E_2 is minimized. Typically, slightly increasing the value of E_1 from its minimum, greatly improves E_2, and vice-versa as shown in Fig. 2. In practice this filter is widely used.

2.2.1 Derivation

The goal is to minimize E_1 subject to a specified value of E_2 and to the linear constraints $\mathbf{X}^\dagger \mathbf{h} = \mathbf{u}$ (this method could be generalized for more than two criteria). Lagrange multipliers are used to obtain the following functional

$$\begin{aligned} \mathcal{L}_\beta(\mathbf{h}, \beta, \Delta) &= E_1 + \beta E_2 - 2\Delta^\dagger(\mathbf{X}^\dagger \mathbf{h} - \mathbf{u}) \\ &= \mathbf{h}^\dagger \mathbf{D} \mathbf{h} + \beta \mathbf{h}^\dagger \mathbf{P} \mathbf{h} - 2\Delta^\dagger(\mathbf{X}^\dagger \mathbf{h} - \mathbf{u}) \\ &= \mathbf{h}^\dagger \mathbf{T} \mathbf{h} - 2\Delta^\dagger(\mathbf{X}^\dagger \mathbf{h} - \mathbf{u}), \end{aligned} \tag{4}$$

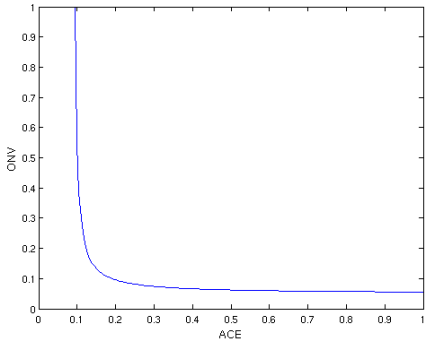


Figure 2. Tradeoff between the output noise variance (ONV) and the average correlation energy (ACE). Note that, although not shown, the values extend beyond 1.

where $0 \leq \beta \leq \infty$ is a scalar Lagrange multiplier, $\Delta \neq 0$ is a vector of nonzero Lagrange multipliers, and $\mathbf{T} = \mathbf{D} + \beta\mathbf{P}$. Note that negative values for β are not considered because $E_2 \geq 0$. The values of β lie between $\beta = 0$ when only E_1 is optimized and $\beta = \infty$ when only E_2 is optimized. The solution that minimizes $\mathbf{h}^\dagger\mathbf{T}\mathbf{h}$ subject to $\mathbf{X}^\dagger\mathbf{h} = \mathbf{u}$ is

$$\mathbf{h} = \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}^\dagger\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{u}. \tag{5}$$

In order to use a bounded scalar, β can be replaced with $\beta = \frac{1}{\lambda}(1-\lambda)$, where $0 \leq \lambda \leq 1$, i.e.,

$$\begin{aligned} \mathbf{T} &= \mathbf{D} + \beta\mathbf{P} = \mathbf{D} + \frac{1}{\lambda}(1-\lambda)\mathbf{P} = \frac{1}{\lambda}(\lambda\mathbf{D} + (1-\lambda)\mathbf{P}) \\ &\propto \lambda\mathbf{D} + (1-\lambda)\mathbf{P}. \end{aligned} \tag{6}$$

For $\lambda = 1$ OTSDF becomes the MACE filter.

2.3 Unconstrained correlation filters: MACH and UMACE filters

The first unconstrained CFs introduced in 1994[15] were the maximum average correlation height (MACH) filter and the unconstrained MACE (UMACE) filter. Previous CFs were constrained to produce an inner product of $\mathbf{h}^\dagger\mathbf{x}_i = u_i$ for the training images, but such hard constraints are not necessarily satisfied by the non-training images. Removing these constraints increases the solution space and may improve the chances of finding a filter with better recognition performance.

The MACH filter has been investigated for many applications.[11,14,16,24,34] The MACH filter is designed to minimize the average (dis-)similarity measure (ASM), i.e., the scatter of the correlation planes, and simultaneously minimize the ACE and maximize the average correlation peak intensity ($|\bar{\mathbf{g}}|^2 = |\mathbf{h}^\dagger\bar{\mathbf{x}}|^2$). The UMACE is a variations of the MACH filter that ignores ASM.

Savvides and Kumar[27] showed that the UMACE filter can be efficiently trained online and adapt to varying data streams caused by changes in illuminations, backgrounds, and/or different views (e.g., due to rotation, scale, pose, non-rigid deformation, etc).

2.3.1 Derivation

The ideal desired correlation output that minimizes the distortion of the correlation outputs with respect to $\dot{\mathbf{g}}$ measured as the mean squared error (MSE)

$$e = \frac{1}{Nd}\sum_{i=1}^{N}|\mathbf{g}_i - \dot{\mathbf{g}}|^2 \tag{7}$$

is found by taking the gradient of e with respect to $\dot{\mathbf{g}}$, setting it equal to zero, and solving for $\dot{\mathbf{g}}$. This gives

$$\dot{\mathbf{g}}_{OPT} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{g}_i = \bar{\mathbf{g}}. \tag{8}$$

Substituting Eq. 8 into Eq. 7 gives the ASM,

$$\begin{aligned} \text{ASM} &= \frac{1}{Nd}\sum_{i=1}^{N}|\mathbf{g}_i-\bar{\mathbf{g}}|^2 = \frac{1}{Nd}\sum_{i=1}^{N}|\mathbf{X}_i^*\mathbf{h}-\bar{\mathbf{X}}^*\mathbf{h}|^2 \\ &= \mathbf{h}^\dagger\left(\frac{1}{Nd}\sum_{i=1}^{N}(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{X}_i-\bar{\mathbf{X}})^*\right)\mathbf{h} = \mathbf{h}^\dagger\mathbf{S}\mathbf{h}, \end{aligned} \tag{9}$$

where diagonal matrix $\mathbf{X}_i$ contains $\mathbf{x}_i$ along its diagonal, and diagonal matrix $\mathbf{S} = \frac{1}{Nd}\sum_{i=1}^{N}(\mathbf{X}_i-\bar{\mathbf{X}})(\mathbf{X}_i-\bar{\mathbf{X}})^*$ represents a measure of the similarity (or more correctly, dissimilarity) of the training images to the true-class mean. The ACE is the previously derived quadratic $\mathbf{h}^\dagger\mathbf{D}\mathbf{h}$ (see Eq. 2). The average peak intensity may be expressed as

$$|\bar{u}|^2 = |\mathbf{h}^\dagger\bar{\mathbf{x}}|^2 = \mathbf{h}^\dagger\bar{\mathbf{x}}\bar{\mathbf{x}}^\dagger\mathbf{h}. \tag{10}$$

The filter $\mathbf{h}$ that simultaneously maximizes the average peak intensity $|\bar{u}|^2$ and minimizes both ASM and ACE is obtained using the following Rayleigh quotient (RQ),

$$J(\mathbf{h}) = \frac{\mathbf{h}^\dagger\bar{\mathbf{x}}\bar{\mathbf{x}}^\dagger\mathbf{h}}{\mathbf{h}^\dagger\mathbf{S}\mathbf{h}+\mathbf{h}^\dagger\mathbf{D}\mathbf{h}} \tag{11}$$

which is maximized when

$$\mathbf{h}_{MACH} = (\mathbf{D}+\mathbf{S})^{-1}\bar{\mathbf{x}}. \tag{12}$$

The UMACE filter is obtained by ignoring $\mathbf{S}$, i.e.,

$$\mathbf{h}_{UMACE} = \mathbf{D}^{-1}\bar{\mathbf{x}}. \tag{13}$$

2.4 Minimum output sum of squared error (MOSSE) filter

The MOSSE filter was introduced in 2010.[2] This filter is designed to minimize the MSE between the desired correlation plane and the actual correlation plane. The simplicity of the MOSSE filter allows the filter to adapt in real time to changes due to rigid-body motion, deformation, and/or lighting. The filter adapts by weighting new images more, with weights for older images decaying exponentially over time. The application investigated was adaptively recognizing faces and other objects-of-interest as the images go through different changes in illuminations and poses. It was reported that the filter processed frames at a rate of 669 frames per second using a standard desktop with 2.4 GHz Core 2 Duo CPU.

2.4.1 Derivation

The MSE between the correlation output $\mathbf{g}$ and the desired correlation output $\dot{\mathbf{g}}_i$ of the training images is given by

$$\begin{aligned} \text{MSE} &= \frac{1}{Nd}\sum_{i=1}^{N}|\mathbf{g}_i-\dot{\mathbf{g}}_i|^2 = \frac{1}{Nd}\sum_{i=1}^{N}\left(\mathbf{g}_i^\dagger\mathbf{g}_i - 2\mathbf{g}_i^\dagger\dot{\mathbf{g}}+\dot{\mathbf{g}}_i^\dagger\dot{\mathbf{g}}_i\right) = \frac{1}{Nd}\sum_{i=1}^{N}\left(\mathbf{h}^\dagger\mathbf{X}_i\mathbf{X}_i^*\mathbf{h} - 2\mathbf{h}^\dagger\mathbf{X}_i\dot{\mathbf{g}}_i+\dot{\mathbf{g}}_i^\dagger\dot{\mathbf{g}}_i\right) \\ &= \mathbf{h}^\dagger\left(\frac{1}{Nd}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i^*\right)\mathbf{h} - 2\mathbf{h}^\dagger\left(\frac{1}{Nd}\sum_{i=1}^{N}\mathbf{X}_i\dot{\mathbf{g}}_i\right)+\frac{1}{Nd}\sum_{i=1}^{N}\dot{\mathbf{g}}_i^\dagger\dot{\mathbf{g}}_i = \mathbf{h}^\dagger\mathbf{D}\mathbf{h} - 2\mathbf{h}^\dagger\mathbf{p}+E_f, \end{aligned} \tag{14}$$

where diagonal matrix $\mathbf{X}_i$ contains the entries of the $\mathbf{x}_i$ along its diagonal, $\mathbf{D} = \frac{1}{Nd}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i^*$, $\mathbf{p} = \frac{1}{Nd}\sum_{i=1}^{N}\mathbf{X}_i\dot{\mathbf{g}}_i$, $E_f = \frac{1}{Nd}\sum_{i=1}^{N}\dot{\mathbf{g}}_i^\dagger\dot{\mathbf{g}}_i$, and $\dot{\mathbf{g}}$ represents the desired correlation output. The $\mathbf{h}$ that minimizes the MSE is found by taking its gradient and setting it equal to zero and solving for $\mathbf{h}$, i.e.,

$$\mathbf{h} = \mathbf{D}^{-1}\mathbf{p}. \tag{15}$$

2.5 Quadratic correlation filter (QCF)

QCFs were introduced in 2004.[10] A disadvantage of linear CFs is that several CFs are required to handle the wide variety of target appearances. When these CFs are applied to a test image, their outputs are compared to select a winner. A QCF requires several linear CFs as well but has the advantage that these CFs are designed to work together to produce a single correlation output. In addition, quadratic classifiers are able to exploit the higher-order statistics of the data, potentially leading to superior recognition performance. In one set of experiments using gray-scale images, QCFs were shown to outperform other CFs.[7]

2.5.1 Derivation

QCF maximizes a metric of separation between the overall outputs for two classes of targets. To express this separation metric, let $\check{\mathbf{x}}^{(c)}$ (all the notation in this derivation is in the spatial domain) be a vectorized image from Class c where $c \in \{1, 2\}$ with d elements and let $\mathbf{Q}$ be a $d \times d$ matrix. The goals is to make $y = \left(\check{\mathbf{x}}^{(c)}\right)^T \mathbf{Q}\check{\mathbf{x}}^{(c)}$ large and positive when $c = 1$ (i.e., when the target is from Class 1) and large and negative when $c = 2$ (i.e., when the target is from Class 2). Let

$$\phi_c = E\left\{\left(\check{\mathbf{x}}^{(c)}\right)^T \mathbf{Q}\check{\mathbf{x}}^{(c)}\right\} \tag{16}$$

indicate the mean QCF output for all training images from Class c. In the most basic QCF design, $\mathbf{Q}$ is computed such that

$$J(\mathbf{Q}) = |\phi_1 - \phi_2| \tag{17}$$

is a large value. In other words, a large between-class separation is desired. To enable a correlation-type QCF implementation architecture, $\mathbf{Q}$ is assumed to be of the form

$$\mathbf{Q} = \sum_{i=1}^{N_1} \check{\mathbf{f}}_i \check{\mathbf{f}}_i^T - \sum_{i=1}^{N_2} \check{\mathbf{b}}_i \check{\mathbf{b}}_i^T. \tag{18}$$

It can be shown[10] that $J(\mathbf{Q})$ is maximized for a given N_1 and N_2 when $\check{\mathbf{f}}_1, \cdots, \check{\mathbf{f}}_{N_1}$ and $\check{\mathbf{b}}_1, \cdots, \check{\mathbf{b}}_{N_2}$ are the eigenvectors corresponding to the N_1 largest positive and the N_2 largest negative eigenvalues, respectively, of

$$\mathbf{R} = \mathbf{R}_1 - \mathbf{R}_2, \tag{19}$$

where

$$\mathbf{R}_c = E\left\{\check{\mathbf{x}}_i^{(c)} \left(\check{\mathbf{x}}_i^{(c)}\right)^T\right\} \tag{20}$$

is the correlation matrix for all the training images in Class c.

To show how to apply the QCF to a test image $x(m, n)$, first let test chip $\check{\mathbf{z}}$ denote a vectorized subregion (equal in size to a training image) of $x(m, n)$ with QCF output

$$\begin{aligned} y &= \check{\mathbf{z}}^T \mathbf{Q}\check{\mathbf{z}} = \check{\mathbf{z}}^T \left(\sum_{i=1}^{N_1} \check{\mathbf{f}}_i \check{\mathbf{f}}_i^T\right) \check{\mathbf{z}} - \check{\mathbf{z}}^T \left(\sum_{i=1}^{N_2} \check{\mathbf{b}}_i \check{\mathbf{b}}_i^T\right) \check{\mathbf{z}} \\ &= \check{\mathbf{z}}^T \mathbf{F}\mathbf{F}^T \check{\mathbf{z}} - \check{\mathbf{z}}^T \mathbf{B}\mathbf{B}^T \check{\mathbf{z}} = \check{\mathbf{v}}^T \check{\mathbf{v}} - \check{\mathbf{w}}^T \check{\mathbf{w}}, \end{aligned} \tag{21}$$

where $\mathbf{F} = [\check{\mathbf{f}}_1, \ldots, \check{\mathbf{f}}_{N_1}]$, $\mathbf{B} = [\check{\mathbf{b}}_1, \ldots, \check{\mathbf{b}}_{N_2}]$, $\check{\mathbf{v}} = \mathbf{F}^T \check{\mathbf{z}}$, and $\check{\mathbf{w}} = \mathbf{B}^T \check{\mathbf{z}}$. The value of v_i (the ith element of $\check{\mathbf{v}}$) can be obtained for all locations within the image $x(m, n)$ by means of 2-D correlation, i.e.,

$$v_i(m, n) = x(m, n) \otimes f_i(m, n), \tag{22}$$

where the eigenfilter $f_i(m, n)$ is the $\check{\mathbf{f}}_i$ eigenvector reshaped as 2-D filter. A similar derivation can be shown for w_i (the ith element of $\check{\mathbf{w}}$). Fig. 3 shows the architecture to apply the QCF to image $x(m, n)$. The QCF output

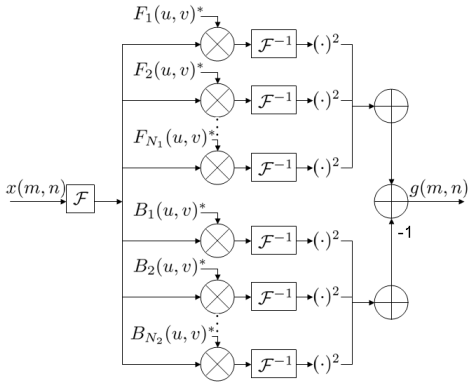


Figure 3. Efficient architecture to apply the QCF to image $x(m,n)$.

$g(m,n)$ is

$$\begin{aligned} g(m,n) &= \sum_{i=1}^{N_1} v_i^2(m,n) - \sum_{i=1}^{N_2} w_i^2(m,n) \\ &= \sum_{i=1}^{N_1} (x(m,n) \otimes f_i(m,n))^2 - \sum_{i=1}^{N_2} (x(m,n) \otimes b_i(m,n))^2, \end{aligned} \tag{23}$$

which can be efficiently implemented in the frequency domain as

$$g(m,n) = \sum_{i=1}^{N_1} \left(\mathcal{F}^{-1}\left\{\mathcal{F}\{x(m,n)\}\mathcal{F}^*\{f_i(m,n)\}\right\}\right)^2 - \sum_{i=1}^{N_2} \left(\mathcal{F}^{-1}\left\{\mathcal{F}\{x(m,n)\}\mathcal{F}^*\{b_i(m,n)\}\right\}\right)^2, \tag{24}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the DFT and IDFT, respectively.

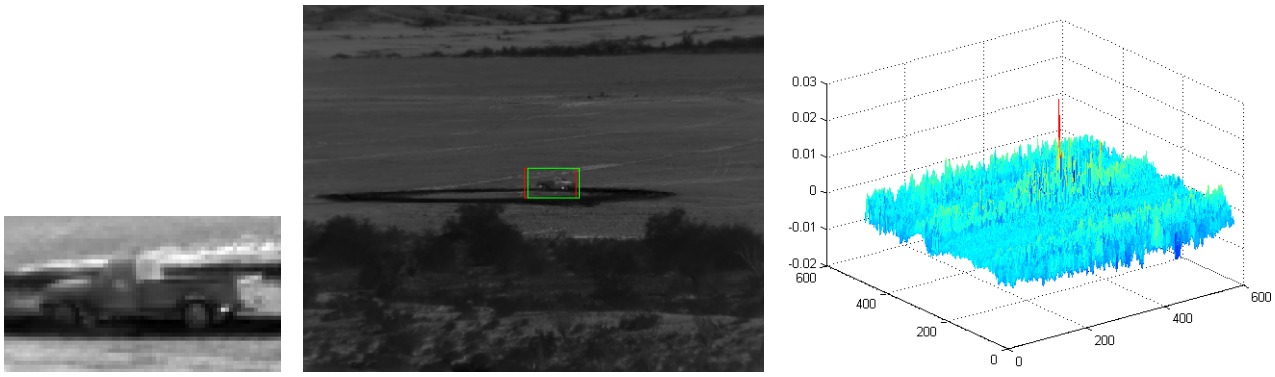
3. MAXIMUM MARGIN CORRELATION FILTERS

A recent breakthrough in CF design is maximum margin correlation filters (MMCFs).[20] MMCFs combine the excellent localization properties of CFs with the very good generalization abilities of support vector machines (SVMs). Traditionally, constrained CF designs are constrained such that the dot product of a training image and the CF template is set to a specific value. In the MMCF formulation, however, this hard constraint is removed and is replaced with an inequality constraint instead; the dot product of a training image and the CF template must be *larger than or equal to* some value.

MMCFs is designed as a tradeoff between maximizing the distance between the hyperplane and data points (i.e., maximizing the margin as in SVM) and minimizing the ACE in order to have a sharp peak in the correlation output. The MMCF multi-objective function can be written as follows,

$$\begin{aligned} \min_{\mathbf{h},b} \quad & \lambda \mathbf{h}^\dagger \mathbf{h} + C \sum_{i=1}^{N} \xi_i + (1-\lambda)\mathbf{h}^\dagger \mathbf{D} \mathbf{h} \\ s.t. \quad & t_i(\mathbf{h}^\dagger \mathbf{x}_i + b) \geq u_i - \xi_i, \end{aligned} \tag{25}$$

where $u_i = 1$ and $t_i = 1$ for true-class images, $u_i = 0$ and $t_i = -1$ for false-class images, $\mathbf{h}^\dagger\mathbf{h} + \frac{C}{\lambda}\sum_{i=1}^{N}\xi_i$ is the margin criterion, $\mathbf{h}^\dagger\mathbf{D}\mathbf{h}$ is the ACE criterion, and $0 < \lambda \leq 1$ is the parameter which trades-off margin maximization and energy minimization (i.e., sharp peaks used for object localization). Setting $\lambda = 1$ will ignore the localization criterion and result in the conventional SVM classifier for centered images. Subsuming one



(a) Pickup truck (b) Target localization (c) Correlation Output

Figure 4. The MMCF response (in (c)) to the test image (in (b)). The MMCF is designed to produce a large value for the pickup truck image (in (a)) and small values for background. The green box shows the ground truth target window, and the red box show the window at the highest correlation value.[20]

quadratic term into the other quadratic term, Eq. 25 can be written as follows,

$$\begin{aligned} \min_{\mathbf{h},b} \quad & \mathbf{h}^\dagger \mathbf{S} \mathbf{h} + C \sum_{i=1}^{N} \xi_i \\ s.t. \quad & t_i(\mathbf{h}^\dagger \mathbf{x}_i + b) \geq u_i - \xi_i, \end{aligned} \tag{26}$$

where $\mathbf{S} = \lambda \mathbf{I} + (1-\lambda)\mathbf{D}$. The MMCF objective function can be written as follows,

$$\begin{aligned} \min_{\tilde{\mathbf{h}},b} \quad & \tilde{\mathbf{h}}^\dagger \tilde{\mathbf{h}} + C \sum_{i=1}^{N} \xi_i \\ s.t. \quad & t_i(\tilde{\mathbf{h}}^\dagger \tilde{\mathbf{x}}_i + b) \geq u_i - \xi_i, \end{aligned} \tag{27}$$

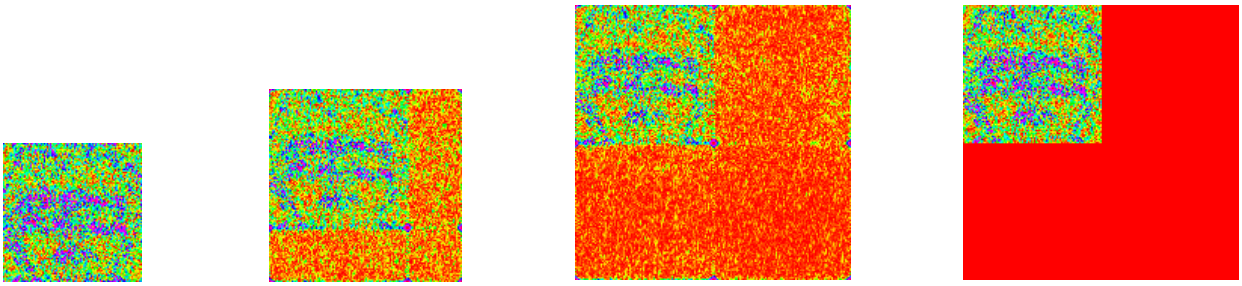
where $\tilde{\mathbf{h}} = \mathbf{S}^{\frac{1}{2}}\mathbf{h}$ and $\tilde{\mathbf{x}}_i = \mathbf{S}^{-\frac{1}{2}}\mathbf{x}_i$. This is the SVM objective function which means that the MMCF design can be implemented using a standard SVM solver by using transformed images to find $\tilde{\mathbf{h}}$.

MMCF has been tested against SVM, OTSDF, MACH, and MOSSE in various computer vision tasks including ATR in a large scene, eye localization in face images, and face classification in centered images. The details of these experiments are found elsewhere.[20] ATR required localizing (in a large scene) and classifying eight targets in 1600 images (see Fig. 4 for one test image and correlation output). The recognition performance of the top three classifiers were MMCF (74.3%), SVM (56.7%), and OTSDF (37.9%). Eye localization required correctly detecting the left and right eyes in 2200 people in the FERET[17] dataset. The top three classifiers were MMCF (96.4%), MOSSE (93.5%), and SVM (88.6%). Face classification required classifying a face into 337 possible classes (the number of subjects) from the Multi-PIE database.[?] The top three classifiers were MMCF (73.5%), MOSSE (64.3%), and OTSDF (55.2%). In all the experiments, MMCF outperformed SVM and state-of-the-art CFs.

4. ZERO-ALIASING CORRELATION FILTERS

In the frequency domain, correlation may be represented as an element-wise multiplication between the DFTs of two signals. However, it is well known that multiplying two DFTs together results in a circular correlation, rather than a linear correlation. In the past, CFs have been formulated with this element-wise multiplication, which implicitly assumes a circular correlation. However, CFs are applied to test data using linear correlation. Therefore, there is an inconsistency between how CFs are designed and how they are applied.

This problem was first explored by Sudharsanan, et al.,[29] in which the well-known MACE filter was reformulated in the space domain. While this CF avoided circular correlation effects, it required a cumbersome



(a) DFT: 92×92 (b) DFT: 128×128 (c) DFT: 183 × 183 (d) DFT: 183 × 183

Figure 5. (best viewed in color) (a-c) The designed templates using different DFT sizes; note that the tail of the templates (for (b) 128×128 and (c)183×183) are non-zero. (d) The designed 183×183 template using the ZAMACE formulation; note that the tail of the template is constrained to be all zero (denoted by the color red).

formulation that was very inefficient to compute. Despite this work, the circular correlations in CF design were largely assumed to be approximately the same as linear correlation, and, as such, CF designs over the years continued to ignore the circular correlation issues in their frequency domain formulations. Recently, Rodriguez and Vijaya Kumar[23] explored several different methods to reduce circular correlation effects during CF design. These methods include various methods of padding training images or windowing training images to reduce edge effects. However, none of the methods presented adequately deal with eliminating circular correlation. In 2013, Fernandez and Vijaya Kumar[6] introduced a solution to the long standing problem of circular correlation in CF designs. This solution, known as zero-aliasing correlation filters (ZACFs) imposes zero-aliasing constraints on the CF design to ensure that the multiplication of DFTs in the frequency domain does in fact correspond to a linear correlation.

In this section, we illustrate the circular correlation problem with existing CF designs. Then, we introduce the ZACF formulation for the MACE filter, and illustrate how it leads to improved recognition performance.

4.1 Circular correlation problem in existing CF designs

When CFs are designed in the frequency domain, the resulting template is of the same size as the DFT used to train the CF. Typically, CFs have been designed either using a DFT the size of the training images or a DFT of size $2d - 1$, where d is size (number of pixels) of the image. The rationale for the latter is that, by zero padding training images, the DFT-based correlation will correspond to a linear correlation, rather than a circular correlation. While this assumption has been used greatly in practice, it is not accurate. To see why, we illustrate three resulting templates in Fig. 5(a-c). Here, we train a MACE filter using face images of size 92×92. We use three different DFT sizes: 92×92, 128×128, and 183×183. In each case, we obtain a template of a different size, equal in size to that of the DFT that was used. Note that the three templates are different from each other. Furthermore, the values in the tail of the templates are all non-zero, regardless of how much zero padding is used on the training images. Therefore, the element-wise multiplication of the DFT of the CF template and the DFT of the training images will always represent a circular correlation, regardless of zero padding.

To fix the circular correlation problem, we introduced hard constraints on the CF design that force the tail of the template to zero. We require that training images are padded and a DFT of size $2d - 1$ is used. By zero-padding training images and imposing constraints on the filter design, we obtain a template that is all zeros in the tail. The element-wise multiplication of the DFT of this template with DFTs of padded training images corresponds to a linear correlation, rather than a circular correlation. An example template obtained from the ZACF approach is shown in Fig. 5(d).

The zero-aliasing MACE (ZAMACE) filter is formulated as follows.[6] Note that this formulation is for 1D signals; the details for 2D signals (images) are found elsewhere.[6] As in the MACE design, the ACE (see Eq. 2) is minimized subject to the peak constraints

$$\mathbf{X}^{\dagger}\mathbf{h} = \mathbf{u}. \tag{28}$$

For the ZAMACE formulation, zero-aliasing constraints are imposed (which force the tail of the template to zero), i.e.,

$$\mathbf{A}^{\dagger}\mathbf{h} = \mathbf{0} \tag{29}$$

where $\mathbf{A}$ is a matrix formed from the IDFT matrix

$$\mathbf{A}^{\dagger} = \begin{bmatrix} 1 & e^{j2\pi(1)(N_x)/N} & \cdots & e^{j2\pi(N-1)(N_x)/N} \\ 1 & e^{j2\pi(1)(N_x+1)/N} & \cdots & e^{j2\pi(N-1)(N_x+1)/N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j2\pi(1)(N-1)/N} & \cdots & e^{j2\pi(N-1)(N-1)/N} \end{bmatrix}. \tag{30}$$

These constraints (i.e., Eq. 28 and Eq. 29) can be jointly represented as

$$\mathbf{B}^{\dagger}\mathbf{h} = \mathbf{k} \tag{31}$$

where

$$\mathbf{B}^{\dagger} = \begin{bmatrix} \mathbf{X}^{\dagger} \\ \mathbf{A}^{\dagger} \end{bmatrix} \text{ and } \mathbf{k} = \begin{bmatrix} \mathbf{u} \\ \mathbf{0} \end{bmatrix}. \tag{32}$$

Minimizing the ACE in Eq. 2 subject to the linear equality constraints in Eq. 31 leads to the new ZAMACE filter[6] given by

$$\bar{\mathbf{h}} = \mathbf{D}^{-1}\mathbf{B}(\mathbf{B}^{\dagger}\mathbf{D}^{-1}\mathbf{B})^{-1}\mathbf{k} \tag{33}$$

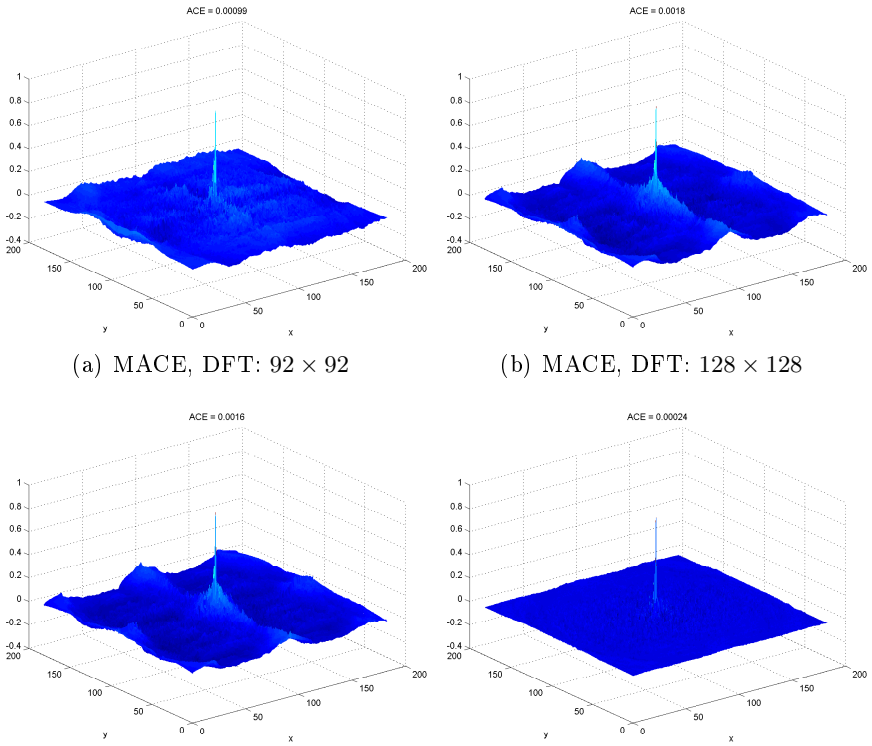
The superior performance of the ZAMACE filter is demostrated by correlating it and the original CF templates with one of the images from the training set. In Fig. 6(a-c), the correlation outputs exhibit higher ACE than the zero-aliasing case shown in Fig. 6(d). This is because the original MACE formulation implicitly assumes circular correlation, and, as such, is not actually minimizing the ACE from a linear correlation. ZAMACE, on the other hand, minimizes the ACE of a linear correlation. The result is a correlation output that features a sharp peak and very low sidelobes, which is the primary objective behind the MACE filter.

We showed that the ZAMACE filter exhibits significantly better performance than the conventional MACE filter. Using the AT&T/ORL Face Dataset,[25] we tested classification performance for human faces. The AT&T/ORL dataset features 10 images of 40 different subjects. Using leave-one-out cross validation, we built CFs using 9 images of each subject, testing on the remaining image of each subject. This was repeated 10 times. We achieved a lower equal error rate for the ZAMACE filter (0.086) compared to a conventional MACE filter using padded training images (0.17). Details can be found elsewhere.[6]

ZACFs offer a fundamental and significant solution to the long-standing issue of circular correlation in CF designs. ZACFs trained using zero-aliasing constraints satisfy the original objective function (minimize the ACE) better than the original MACE filter by ensuring that the objective function accurately represents a linear correlation. The result is significantly improved recognition performance.

5. CONCLUSION

In this paper we reviewed important CFs and cited works where CFs have been successfully used for a variety of ATR and biometric applications, including vehicle recognition, target tracking, face recognition, face localization, face tracking, identification encoding, fingerprint recognition, iris recognition, ocular recognition, pedestrian localization, and human actions. We summarized two recent breakthroughs in CFs: MMCFs and ZACFs and their superior performance in various ATR and biometric experiments.



(a) MACE, DFT: 92×92 (b) MACE, DFT: 128×128

(c) MACE, DFT: 183×183 (d) ZAMACE, DFT: 183×183

Figure 6. Correlation outputs for (a-c) three examples of MACE filters and (d) the ZAMACE filter.

REFERENCES

1. V. N. Boddeti, F. Su, and B. V. K. Vijaya Kumar. A biometric key-binding and template protection framework using correlation filters. In *Lecture Notes on Computer Science*, 2009.
2. D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2010.
3. D. S. Bolme, B. A. Draper, and J. R. Beveridge. Average of synthetic exact filters. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2009.
4. D. S. Bolme, Y. M. Lui, B. A. Draper, and J. R. Beveridge. Simple real-time human detection using a single correlation filter. In *IEEE Int'l Workshop on Performance Evaluation of Tracking and Surveillance*, 2010.
5. D. Casasent, G. Ravichandran, and S. Bollapraggada. Gaussian minimum average correlation energy correlation filters. *Applied Optics*, 30(35):5176–5181, 1991.
6. J. A. Fernandez and B. V. K. Vijaya Kumar. Zero-aliasing correlation filters. *International Symposium on Image and Signal Processing and Analysis*, pages 101–106, 2013.
7. R. A. Kerekes and B. V. K. Vijaya Kumar. Selecting a composite correlation filter design: a survey and comparative study. *Opt. Eng.*, 47(6):1–18, 2008.
8. R. A. Kerekes and B. V. K. Vijaya Kumar. Enhanced video-based target detection using multi-frame correlation filtering. *IEEE Trans. Aerospace and Electronic Systems*, 45(1):289–307, 2009.
9. A. Mahalanobis, R. Muise, and S. R. Stanfill. Quadratic correlation filter design methodology for target detection and surveillance applications. *Applied Optics*, 43(27):5198–5205, 2004.
10. A. Mahalanobis, R. Muise, S. R. Stanfill, and A. Van Nevel. Design and application of quadratic correlation filters for target detection. *Applied Optics*, 40(3):837–850, 2004.
11. A. Mahalanobis and H. Singh. Application of correlation filters for texture recognition. *Applied Optics*, 33(11):2173–2179, 1994.

12. A. Mahalanobis, R. Stanfill, and K. Chen. A bayesian approach to activity detection in video using multi-frame correlation filters. In *Proc. SPIE*, 2011.
13. A. Mahalanobis, B. V. K. Vijaya Kumar, and D. Casasent. Minimum average correlation energy filters. *Applied Optics*, 26(5):3633–3640, 1987.
14. A. Mahalanobis, B. V. K. Vijaya Kumar, and R. Frankot. Intraclass and between-class training-image registration for correlation-filter synthesis. *Applied Optics*, 39(17):2918–2924, 2000.
15. A. Mahalanobis, B. V. K. Vijaya Kumar, S. Song, S. R. F. Sims, and J. F. Epperson. Unconstrained correlation filters. *Applied Optics*, 33(17):3751–3759, 1994.
16. A. Nevel and A. Mahalanobis. Comparative study of maximum average correlation height filter variants using ladar imagery. In *Proc. SPIE*, 2003.
17. P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
18. Ph. Réfrégier. Filter design for optical pattern recognition: multicriteria optimization approach. *Opt. Lett.*, 15(15):854–856, 1990.
19. Ph. Réfrégier and J. Figue. Optimal trade-off filters for pattern recognition and their comparison with the wiener approach. *Optical Computing and Processing*, 1(3):245–266, 1991.
20. A. Rodriguez, V. N. Boddeti, B. V. K. Vijaya Kumar, and A. Mahalanobis. Maximum margin correlation filters: A new approach for localization and classification. *IEEE Trans. Image Processing*, 22(2):631–643, 2013.
21. A. Rodriguez and B. V. K. Vijaya Kumar. Automatic target recognition of multiple targets from two classes with varying velocities using correlation filters. In *IEEE Int'l Conf. of Image Processing*, 2010.
22. A. Rodriguez and B. V. K. Vijaya Kumar. Segmentation-free ocular detection and recognition. In *Proc. SPIE*, 2011.
23. A. Rodriguez and B. V. K. Vijaya Kumar. Dealing with circular correlation effects. In *Proc. SPIE*, 2013.
24. M. D. Rodriguez, J. Ahmed, and M. Shah. Action MACH–a spatio-temporal maximum average correlation height filter for action recognition. In *IEEE Conf. on Computer Vision and Pattern Recognition, 2008*, 2008.
25. F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.
26. M. Savvides, J. Heo, J. Thornton, P. Hennings, C. Xie, K. Venkataramani, R. A. Kerekes, M. Beattie, and B. V. K. Vijaya Kumar. Biometric identification using advanced correlation filter methods. In *Springer-Verlag Lecture Notes in Computer Science, Ambient Intelligence*, 2005.
27. M. Savvides and B. V. K. Vijaya Kumar. Efficient design of advanced correlation filters for robust distortion-tolerant face recognition. In *IEEE Conf. Advanced Video and Signal Based Surveillance*, 2003.
28. M. Savvides, B. V. K. Vijaya Kumar, and P. Khosla. Face verification using correlation filters. In *IEEE Workshop on Automatic Identification Advanced Technologies*, 2002.
29. Subramania I. Sudharsanan, Abhijit Mahalanobis, and Malur K. Sundareshan. Unified framework for the synthesis of synthetic discriminant functions with reduced noise variance and sharp correlation structure. *Optical Engineering*, 29(9):1021–1028, 1990.
30. J. Thornton, M. Savvides, and B. V. K. Vijaya Kumar. A Bayesian approach to deformed pattern matching of iris images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(4):596–606, 2007.
31. B. V. K. Vijaya Kumar. Minimum-variance synthetic discriminant functions. *J. Opt. Soc. Am. A*, 3(10):1579–1584, 1986.
32. B. V. K. Vijaya Kumar, A. Mahalanobis, and R. D. Juday. *Correlation Pattern Recognition*. Cambridge Univ. Press, 2005.
33. B. V. K. Vijaya Kumar, M. Savvides, C. Xie, K. Venkataramani, J. Thornton, and A. Mahalanobis. Biometric verification with correlation filters. *Applied Optics*, 43(2):391–402, 2004.
34. B. Walls and A. Mahalanobis. Performance of the MACH filter and DCCF algorithms in the presence of data compression. In *Proc. SPIE*, 1999.
35. C. Xie, M. Savvides, and B. V. K. Vijaya Kumar. Quaternion correlation filters for face recognition in wavelet domain. In *Proc. Int'l Conf. on Acoustic, Speech and Signal Processing*, 2005.