Enhanced Matched Filter Theory and Applications

KAVEH HEIDARY
Department of Electrical Engineering and Computer Science
Alabama A&M University
kaveh.heidary@aamu.edu

Abstract: Enhanced matched filter (EMF) comprises a distortion tolerant correlation filter and its associated threshold. It is an effective signal detection tool with superior immunity to noise, distortion, and clutter, used for imagery-based detection, authentication, classification, recognition, and tracking of targets of interest. The EMF is synthesized by combining multiple image templates of the target of interest, acquired under prescribed target states and view conditions. In autonomous vision and tracking systems, one EMF can potentially replace copious manifold of exemplar images without adversely affecting the classifier performance. This leads to proportional reduction of operation phase computational load, and concomitant smaller footprint, lighter, faster, and more power efficient smart vision systems. This paper develops the underlying theory of EMF and provides the analytical models of its operation. The EMF and the standard matched filter (MF) performance results based on analytical formulations and empirical studies are presented and are compared to the performance data using virtual and real test images. Results pertaining to the performance comparison of the EMF and the synthetic discriminant function (SDF) filters are also presented.

Keywords: Enhanced Matched Filter, Correlation Filtering, Distortion Tolerant, Image Classification.

1. Introduction

The growing proliferation of imaging and machine vision systems has led to the incessant production of vast volumes of imagery data. Machine vision deployment is wide ranging and persistent, spanning diverse application areas including advanced driver assist systems, biometrics, medical diagnostics, environmental monitoring, public safety, defense and security, robotic navigation, virtual reality, entertainment, manufacturing, education, and others [1-9]. In machine vision systems, including those with human in the loop, due to the massive volumes of sensor data involved, human inspection of all the images is virtually impossible. Effective utilization of the colossal quantities of imagery data, therefore, necessitates a high degree of automation. Autonomous vision systems, depending on the application area, may be required to detect, authenticate, classify, recognize, and track objects of interest in image and video files, based on combinations of spatial and spectral signatures. Pseudo autonomous vision systems, on the other hand, are expected to queue the human in the loop towards the spatiotemporal regions of interest. The system must sift through enormous volumes of data proficiently to present the human operator with timely actionable intelligence or with drastically abridged datasets that make further analysis practicable. This paper explores an innovative tool for detection, classification, and recognition of objects in imagery data, based on the objects' spatial signatures. Specifically, it presents a technique for reducing data storage and processing loads while concurrently improving the detector and classifier robustness.

The objective of the image detector and classifier presented here is different from those of the recently developed and very powerful neural network based deep learning image classifiers [6-16]. Deep learning image classification algorithms using convolutional neural networks are purely data driven and require copious amounts of training data, which may not be available in some applications. The image classifier presented here is intended for niche application areas with dearth of training images, where model transparency and adaptability are desired. The image classifier presented here allows for computationally efficient online real-time retraining of the classifier which affords an evolutionary mechanism for improving the classifier as new episodes of the target of interest are detected during the operation phase.

The universally accepted gold standard for detection of known signatures in noise corrupted sensor signals is the matched filter (MF), which reduces to template matching under white noise condition [10-19]. The standard method for locating an object of interest in a two-dimensional (2D) intensity image is template matching or correlation filtering. The appealing attributes of correlation filters include linearity, shift invariance and graceful degradation, which make them suitable for detection of targets at arbitrary positions in the image frame in the presence of noise and clutter. In template matching, a set of image templates representing the object of interest under assorted view conditions, constitutes the target filter bank. The image at the sensor output

is correlated with the set of templates to establish the presence, or lack thereof, and the locations of the objects of interest in the image under test. The images of a particular object have an immense degree of variability owing to changes in intrinsic and extrinsic conditions. Intrinsic conditions include object articulation, non-rigid deformation, temperature, in-plane, and out-of-plane rotation. Extrinsic conditions that affect the object image include range, view angle, in-plane rotation, lighting, shadowing, background, partial occlusion, other environmental conditions, and sensor noise. Objects associated with a target class of interest can project an innumerably diverse set of signatures in the sensor focal plane array [20-35]. Hence, the filter associated with a particular target class may potentially contain multitude of image templates. In a typical application, the autonomous vision system must store target filters associated with all object classes of interest for that application. It must also process each sensor image by spatially correlating it with respect to potentially many image templates related to various target filters associated with the respective object classes of interest [28]. This results in data storage and processing requirements that may not be available in practical systems especially in mobile applications with power constraints. For example, an autonomous vision system that is required to ascertain presence and locations of several classes of objects of interest in a video stream must store potentially thousands of image templates and correlate each sensor image with many templates in real-time. For practical systems, this presents an insurmountable data storage and computational load.

To circumvent the computational challenges presented by target image variability, various classes of distortion tolerant correlation filters have been developed and successfully implemented [26, 27, 32-55]. These are composite filters obtained by combining multiple training images associated with the respective target class of interest. The computational complexities of processing a sensor image with one target template and the corresponding composite filter are equivalent. Replacing a typical set of target templates with the respective composite filter, therefore, leads to reducing the operational phase processing and data storage loads. The computational effort expended for synthesis of the distortion tolerant composite filter pays dividend in the operation phase. The expenditure of computational resources for filter synthesis is a one-time investment, whereas the operation phase savings are recurring. This paper presents an intuitive and computationally efficient technique for the synthesis of a distortion tolerant composite correlation filter. The enhanced matched filter (EMF) presented here is obtained by straightforward amalgamation of the corresponding set of target-class training images. Training images, representing prescribed target states and view angles, are properly conditioned and added pixel-wise. The resultant synthetic image is subsequently normalized to form the EMF. One of the advantages of EMF, with respect to other composite filters reported in the literature, is the potential for operation phase dynamic improvement of the filter. The EMF training and its application for detection of targets of interest in the sensor images are performed entirely in the image space without the need for transformation to a feature space and feature extraction. The EMF can be adaptively upgraded while detecting new target episodes, without involving the constituent original trainer image templates. A distinguishing feature of EMF is the ascribed threshold, which is a byproduct of the filter computation process and does not rely on non-target class training images.

The paper organization is as follows. Section 2 provides definitions and the detailed procedure for the EMF synthesis. Section 3 presents a theoretical framework for assessment of the EMF filter performance and develops mathematical formulas for determining the filter threshold. Analytical formulas for computation of the expected values of the MF and EMF thresholds are derived. As far as can be ascertained by the author, hitherto closed-form mathematical expressions for the expected values of MF and EMF thresholds have not been reported in the open literature. The analytical comparisons of the MF and EMF thresholds account for their relative performance in relation to detection of untrained-on images. Section 4 develops and implements a simple procedure for comparing EMF and MF thresholds using synthetically generated images. Presented in Section 5 are filter performance test results using actual binary images. In Section 6 performance of the EMF is compared with an image classifier based on the synthetic discriminant function SDF described in [5]. Conclusions and suggested future work are presented in Section 7.

2. Formulation of the MF and EMF Correlation Filters and Thresholds

This Section presents the process of determining the trainer image that best characterizes the target class

training set of images. Here, the characteristic trainer and the associated threshold constitute the matched filter (MF). The procedure for computing the correlation filter and the associated threshold comprising the enhanced matched filter (EMF) is also described. Let us assume that a set of training images representing the target class of interest under prescribed target states and view conditions is

available. The trainer set peak cross-correlation matrix is given below.

$$\Lambda = \left[\lambda_{p,q}\right]_{Q\times Q}; \lambda_{p,p} = 1, \lambda_{p,q} = \lambda_{q,p} < 1 \ \ \forall p \neq q \ (1).$$

Where, $Q, \lambda_{p,q}$ denote, respectively, the number of images in the training set, and the peak mutual cross-correlation between typical trainer pairs as given below.

$$\lambda_{pq} = \max_{m,n}\left\{\sum_{v=1}^{N}\sum_{u=1}^{M} s_p(u,v)s_q\big((u-m+1),(v-n+1)\big)\right\}; 1 \le p,q \le Q \quad (2).$$

Where, $s_p(m,n)$ represents a typical normalized training image, and M, N denote spatial dimensions of the image along, respectively, the vertical and horizontal directions. It is assumed that all trainers have the same spatial dimensions. This is done by appending zero rows and columns to images whose spatial dimensions are smaller than the largest image in the set. Here and henceforth, periodic continuation of the matrices is used for the computation of correlation surfaces. The images in (2) are assumed to be normalized such that for each image the mean of pixel values is zero and the sum of squares of pixel values is one.

$$\forall q: \sum_{n=1}^{N}\sum_{m=1}^{M} s_q(m,n) = 0\,, \sum_{n=1}^{N}\sum_{m=1}^{M} s_q^2(m,n) = 1 \ \ (3).$$

Each training image is normalized as shown below.

$$\bar{g}_q(m,n) = g_q(m,n) - \frac{\left(\sum_{n=1}^{N}\sum_{m=1}^{M} g_q(m,n)\right)}{MN} \quad (4).$$

$$s_q(m,n) = \frac{\bar{g}_q(m,n)}{\sqrt{\sum_{n=1}^{N}\sum_{m=1}^{M}\bar{g}_q^2(m,n)}} \quad (5).$$

Where, g_q, s_qdenote, respectively the raw and the corresponding normalized training image. Among the set of Q training images, the trainer that best represents the set is the one whose minimum peak correlation with respect to all the trainers has the greatest value. This trainer is called the characteristic trainer or the anchor and is identified as follows.

$$\exists k: \min_q \lambda_{kq} \ge \min_q \lambda_{pq} \ \ \forall p, 1 \le p \le Q \quad (6).$$

$$h_{MF}(m,n) = s_k(m,n) \quad (7).$$

Where, k is the index of the characteristic trainer, and $h_{MF}(m,n)$ denotes the characteristic trainer. Coupled with a threshold value, the characteristic trainer can potentially be utilized as a binary classifier as will be shown shortly. In this paper, the characteristic trainer and its threshold are collectively referred to as the matched filter (MF).

$$T_{MF} = \min_{1\le q\le P} \lambda_{kq} \quad (8).$$

Where, k is given in (6), and h_{MF}, T_{MF} denote, respectively, the correlation filter and the threshold associated with the MF. An image whose normalized peak correlation with respect to h_{MF} exceeds T_{MF} is classified as target, otherwise it is classified as non-target.

Next, the process of formulating the enhanced matched filter (EMF) is articulated. Each of the training images, for the purpose of spatially aligning it with respect to the anchor h_{MF}, must be shifted in accordance with the location of the peak of its cross-correlation surface with respect to the anchor.

$$\forall q \ \exists\big(1 \le i_q \le M, 1 \le j_q \le N\big) \ : \ \ldots$$
$$\sum_{n=1}^{N}\sum_{m=1}^{M} h_{MF}(m,n)s_q\Big(\big(m-i_q+1\big),\big(n-j_q+1\big)\Big) \ge$$
$$\ldots \sum_{n=1}^{N}\sum_{m=1}^{M} h_{MF}(m,n)s_q\big((m-u+1),(n-v+1)\big) \quad (9).$$

$$\hat{s}_q(m,n) = s_q\Big(\big(m-i_q+1\big),\big(n-j_q+1\big)\Big) \quad (10).$$

Where,$\hat{s}_q(m,n)$ denotes a typical normalized trainer which has been spatially aligned with respect to the anchor. The enhanced matched filter (EMF) template is obtained by the pixel-wise summation of the spatially aligned trainers given in (10), and subsequent renormalization of the resultant synthetic image, as shown below.

$$\hat{h}_{EMF}(m,n) = \sum_{q=1}^{Q} \hat{s}_q(m,n) \quad (11).$$

$$h_{EMF}(m,n) = \frac{\hat{h}_{EMF}(m,n)}{\sqrt{\sum_{n=1}^{N}\sum_{m=1}^{M} h_{EMF}^2(m,n)}} \quad (12).$$

The nominal value of the EMF threshold is computed as the minimum peak correlation of $h_{EMF}(m,n)$ with respect to all the constituent trainers as shown below.

$$C_q = \max_{m,n}\left\{\sum_{v=1}^{N}\sum_{u=1}^{M} h_{EMF}(u,v)s_q\big((u-m+1),(v-n+1)\big)\right\} (13).$$

$$T_{MF} = \min_{1\le q\le P}\big(C_q\big) \quad (14).$$

Substituting for h_{EMF} from (11,12) in (13,14), utilizing (2,3), after simplification, one arrives at the following expression for the EMF threshold.

$$T_{EMF} = \frac{min_{1\le p\le Q}\left\{\sum_{q=1}^{Q}\lambda_{p,q}\right\}}{\sqrt{Q + 2\sum_{q=p+1}^{Q}\sum_{p=1}^{Q}\lambda_{pq}^2}} \quad (15).$$

Appendix A summarizes the procedural steps for obtaining the enhanced matched filter and the corresponding threshold. One can use the MF or the EMF as a binary classifier as described next. The input image is first normalized in accordance with (4,5), and its peak cross-correlation (PCC) with respect to the template associated with the designated classifier is computed. If PCC exceeds the respective threshold, the input image is labeled as target, otherwise it is labeled as non-target. Two types of classification error can occur, namely, false negative, where a target class image is classified as non-target, and false positive, where a non-target class image is labeled as target. Raising the threshold value results in decreasing the false

positive and increasing the false negative error, whereas, lowering the threshold results in higher false positive and lower false negative classification error.

3. Analytic Formulas for Computing the MF and EMF Thresholds

The steps for obtaining filter thresholds were outlined in the previous section. Equations (8) and (15) are used for computing the exact threshold values of the MF and EMF classifiers, respectively. As explained before, raising the classifier threshold beyond the respective nominal value, while lowering the false-positive error increases the false negative error, to the extent that one or more constituent trainers will not be properly captured by the classifier. A quantitative measure of the relative performance of the two types of classifiers entails comparison of their respective threshold values. All things being equal, the classifier with higher nominal threshold is considered superior, due to its potentially lower false-positive error, for comparable false-negative error rates for the two classifier types. Closed-form mathematical formulas for computing the expected values of the MF and EMF thresholds are given next.

3.1 Expected Value of the Matched Filter Threshold

The peak cross-correlation matrix of the target-class training set of images, given in (1), is used to find the nominal value of the MF threshold as follows.

$$\forall p\ 1 \le p \le Q\ V_p = \min_{1\le q\le Q}(\lambda_{pq}) \quad (16).$$

$$T_{MF} = \max_{1\le p\le Q}(V_p) \quad (17).$$

Where, Q is the number of trainers, $\lambda_{p,q}$ is the peak cross-correlation between a trainer pair, and T_{MF} denotes the MF threshold. The expected value of the nominal threshold is computed in terms of the trainer set maximum disparity factor defined below.

$$\gamma = 1 - \min_{1\le p.q\le Q}(\lambda_{pq}) \quad (18).$$

$$\mathrm{E}\{T_{MF}\} = \mathrm{E}\left\{\max_{1\le p\le Q}(V_p)\right\} \quad (19).$$

Where, γ is the trainer set maximum disparity factor, and $\mathrm{E}\{x\}$ denotes the expectation operator.

$$\mathrm{E}\{V_j\} = \int_{1-\gamma}^{1} v\, f_{V_j}(v)\, dv \quad ;\quad 1 \le j \le Q \quad (20).$$

$$f_{V_j}(v) = \frac{d}{dv} F_{V_j}(v) \quad (21).$$

Where, $f_{V_j}(v), F_{V_j}(v)$ denote, respectively, the probability density function (PDF) and the cumulative distribution function (CDF).

$$F_{V_j}(v) = \Pr\{V_j \le v\} = 1 - \Pr\{V_j > v\} \quad (22).$$

$$\Pr\{V_j > v\} = \prod_{q=1, q\ne j}^{Q} \Pr\{\lambda_{jq} > v\} \quad (23).$$

Where, $\Pr\{x\}$, $\mathbf{\Pi}$ denote, respectively, the probability and product operators. To arrive at (23) one must assume that the mutual peak correlation between an arbitrary pair of trainers is independent of the peak correlations of those trainers with respect to a third trainer. It is further assumed that the peak correlation between two arbitrarily selected trainers is uniformly distributed.

$$\Pr\{\lambda_{jq} > v\} = \frac{1-v}{\gamma} \quad 1-\gamma \le v \le 1 \quad (24).$$

Substituting (24) in (23), and differentiating the resultant CDF of (22), one arrives at the following PDF expression. Substituting the resultant PDF in (20) leads to the expectation value given below.

$$f_{V_p}(v) = \frac{Q-1}{\gamma^{Q-1}}(1-v)^{Q-2}; 1-\gamma \le v \le 1 \quad (25).$$

$$\mathrm{E}\{V_p\} = 1 - \gamma + \frac{\gamma}{Q} \quad ;\quad 1 \le p \le Q \quad (26).$$

Equation (26) gives the expected value of the minimum peak correlation of a typical trainer with respect to all the trainers of a certain class. Therefore, considering (19) the expected value of the MF threshold is expressed as follows.

$$\mathrm{E}\{T_{MF}\} = \mathrm{E}\{W\} \ : W = \max_{1\le p\le Q}(V_p) \quad (27).$$

$$\mathrm{E}\{W\} = \int_{1-\gamma}^{1} w f_W(w) dw : \ f_W(w) = \frac{d}{dw} F_W(w) \quad (28).$$

$$F_W(w) = \Pr\left\{\max_{1\le p\le Q}(V_p) \le w\right\} = \prod_{p=1}^{Q} \Pr\{V_p \le w\} \quad (29).$$

$$\Pr\{V_p \le w\} = \int_{1-\gamma}^{w} f_{v_p}(v) dv \quad (30).$$

Substituting (25) in (30), after integration and subsequent substitution in (29), we arrive at the following CDF expression.

$$\mathrm{F}_W(w) = \left[1 - \frac{(1-w)^{Q-1}}{\gamma^{Q-1}}\right]^Q \quad (31).$$

Differentiating (31) and substituting the result in (28), integrating by parts and simplifying, one arrives at the following.

$$\mathrm{E}\{W\} = 1 - \int_{1-\gamma}^{1} \left[1 - \frac{(1-w)^{Q-1}}{\gamma^{Q-1}}\right]^Q dw \quad (32).$$

The closed-form analytical expression for the integral in (32) is give in Appendix B. Substituting from (32) in (27) leads to the following expression for the expected value of the MF threshold.

$$\mathrm{E}\{T_{MF}\} = 1 - \gamma(Q-1)^Q \prod_{q=1}^{Q}\left(\frac{q}{1+(Q-1)q}\right) \quad (33).$$

It is noted that the expected value of the MF threshold is a linearly decreasing function of the

trainer set disparity factor, and the slope is dependent on the number of trainers, which is an intuitively expected result. In the limit, as the number of trainers approaches infinity, the expected threshold is given below.

$$\lim_{Q\to\infty}\{T_{EMF}\} = 1-\gamma \qquad (34)$$

3.2 Expected Value of the Enhanced Matched Filter Threshold

It is seen from (15) that, in general, the EMF threshold can be expressed as the ratio of two random variables as shown below.

$$\mathrm{T}_{EMF} = \frac{X}{U}, \qquad X = \min_{1\le p\le Q}\{V_p\}, U = \sqrt{W} \qquad (35a).$$

$$V_p = 1 + \sum_{\substack{q=1\\ q\neq p}}^{Q} \lambda_{p,q}, W = Q + 2\sum_{q=p+1}^{Q}\sum_{p=1}^{Q}\lambda_{p,q}^2 \qquad (35b).$$

Invoking the central limit theorem [56] the random variables V_P, W are approximated as normally distributed with means and standard deviations given below.

$$V_p \sim \mathrm{N}\left(m_{V_p}, \sigma_{V_p}\right), \quad W \sim \mathrm{N}(m_W, \sigma_W) \qquad (36).$$

$$m_{V_p} = \frac{\gamma}{2} + \left(1-\frac{\gamma}{2}\right)Q, \sigma_{V_p} = \gamma\sqrt{\frac{(Q-1)}{12}} \qquad (37a).$$

$$m_W = \left(1-\frac{\gamma}{2}\right)Q^2 + \frac{\gamma Q}{2}, \sigma_W = \gamma\sqrt{\frac{Q(Q-1)}{6}} \qquad (37b).$$

Where, as before Q, γ are, respectively, the number of trainers and the trainer set maximum disparity factor, and $\mathrm{N}(m, \sigma)$ denotes the normal distribution function. For typical parameter values associated with practical scenarios, i.e., $Q\sim10, \gamma\sim0.5$, it is seen from (37-b) that $m_W \gg \sigma_W$. Using (35) and considering the foregoing observations one arrives at the following expression for the expected value of the EMF threshold.

$$\mathrm{E}\{T_{EMF}\} \cong \frac{\mathrm{E}\left\{\min_p[V_p]\right\}}{\mathrm{E}\{U\}} \cong \frac{\mathrm{E}\left\{\min_p[V_p]\right\}}{\sqrt{m_W}} \qquad (38).$$

Where, m_W is given in (37-b) and the computation of the numerator term in (38) proceeds as follows.

$$\mathrm{E}\left\{\min_p[V_p]\right\} = \int_{-\infty}^{\infty} x f_X(x)\,dx \qquad (39a).$$

$$X = \min_p[V_p], \quad f_X(x) = \frac{d}{dx}F_X(x) \qquad (39b).$$

$$F_X(x) = 1 - \Pr\{X > x\} = 1 - \prod_{p=1}^{Q}\Pr\{V_p > x\} \qquad (40).$$

Where, the mutual independence of samples of the random variable V_P is implied. Using the normal distribution function of (36) for computing the probability terms in (40), substituting in (39) and simplifying, one arrives at the following.

$$\Pr\{V_p > x\} = \frac{1}{2}\mathrm{erfc}\left(\frac{x-m_{V_p}}{\sqrt{2}\sigma_{V_p}}\right) \qquad (41).$$

$$f_X(x) = \frac{Q}{\sqrt{2\pi}2^{Q-1}\sigma_{V_p}} e^{-\left(\frac{x-m_{V_p}}{\sqrt{2}\sigma_{V_p}}\right)^2}\left[erfc\left(\frac{x-m_{V_p}}{\sqrt{2}\sigma_{V_p}}\right)\right]^{Q-1} \qquad (42).$$

Where, $erfc(x)$ is the complementary error function. Substituting the PDF of (42) in (39), after simplification, one arrives at the following.

$$\mathrm{E}\left\{\min_p[V_p]\right\} = m_{V_p} + \frac{Q\sigma_{V_p}}{2^{(Q-1)}}\sqrt{\frac{2}{\pi}} \times \ldots \ldots \int_{-\infty}^{\infty} y e^{-y^2}[erfc(y)]^{(Q-1)}dy \qquad (43).$$

The integral in (43), for $Q = 1,2$ can be evaluated exactly, and for $Q \geq 3$ is approximated as follows.

$$\int_{-\infty}^{\infty} y e^{-y^2}[erfc(y)]^{(Q-1)}dy = \ldots \ldots \begin{cases} 0 & ; \quad Q=1 \\ -\frac{1}{\sqrt{2}} & ; \quad Q=2 \\ -\prod_{q=3}^{Q}\left(2-\frac{0.45}{\sqrt{q}}\right) & ; \quad Q\geq 3 \end{cases} \qquad (44).$$

Substituting the expression of (44) in (43), and subsequent substitution of the result in (35), one arrives at the following closed-form expression for the expected value of the EMF threshold.

$$E\{T_{EMF}\} \cong \frac{\left(1-\frac{\gamma}{2}\right)Q + \frac{\gamma}{2} - \gamma Q\sqrt{\frac{Q-1}{24\pi}}\prod_{q=3}^{Q}\left(1-\frac{0.225}{\sqrt{Q}}\right)}{\sqrt{\left(1-\frac{\gamma}{2}\right)Q^2 + \frac{\gamma Q}{2}}} \qquad (45).$$

It is seen from (45) that the EMF threshold is related to the trainer set disparity factor non-linearly. This contrasts with the linear relationship between the MF threshold and γ as expressed in (33). In the limit, as the number of trainers approaches infinity, the expected value of the EMF threshold is given below.

$$\lim_{Q\to\infty}\mathrm{E}\{T_{EMF}\} \cong \sqrt{1-\frac{\gamma}{2}} \qquad (46).$$

It is instructive to compare the asymptotic expressions for the expected values of the thresholds of two types of correlation filters, namely, MF and EMF given by (34, 46). It is seen that the rate of descent of the EMF threshold, as the maximum trainer set disparity factor is increased, is much slower than that of the MF threshold. For any value of γ $(0 < \gamma < 1)$, the expected value of the EMF threshold is greater than that of MF. This property is responsible for the EMF classifiers' lower false positive error than MF classifiers, for a prescribed level of false negative error. The numerical examples of the next section affirm this principle.

3.3 Analytical and Empirical Comparisons of Filter Thresholds

In this section the expected values of the thresholds for two types of correlation filters, namely, MF and EMF, are assessed using the analytical formulations of the previous two sections and empirical techniques. The effects of two parameters, namely the number of trainers Q, and the trainer set maximum disparity factor □ □ on the thresholds are examined. Here, it is assumed that the peak correlations between pairs of randomly selected trainers are independent samples of the random variable uniformly distributed in $[(1-\gamma), 1]$. This assumption is made to permit closed-form analysis and direct comparison of the models, and its violation does not in any way negate the conclusions.

The plots of Figure 1 show the effect of trainer set maximum disparity factor on the expected values of filter thresholds for MF and EMF, based on the analytical formulations provided by Equations (33, 45). For each of the filters the threshold values are plotted for two settings of the number of training images. It is seen that for fixed □ and Q, the expected value of T_{EMF} exceeds that of T_{MF} by a great margin.

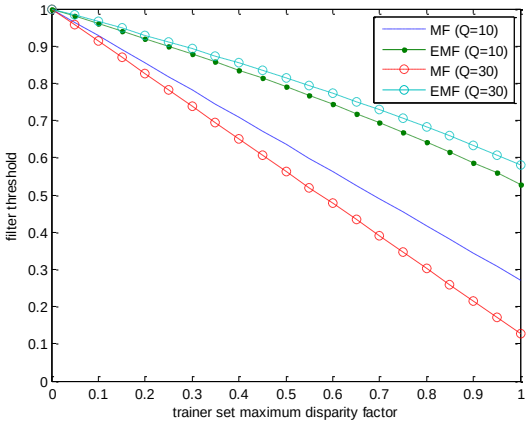


FIGURE 1: Effect of trainer set maximum disparity factor □ on the expected values of filter thresholds for MF and EMF. Numbers in the caption denote number of trainers.

The plots of Figure 2 show the effect of the number of trainers on the expected values of filter thresholds. For each filter type the threshold values are plotted for two settings of the trainer set maximum disparity factor □. As before, for fixed □ □and Q, the expected value of EMF threshold exceeds that of MF by a great margin.

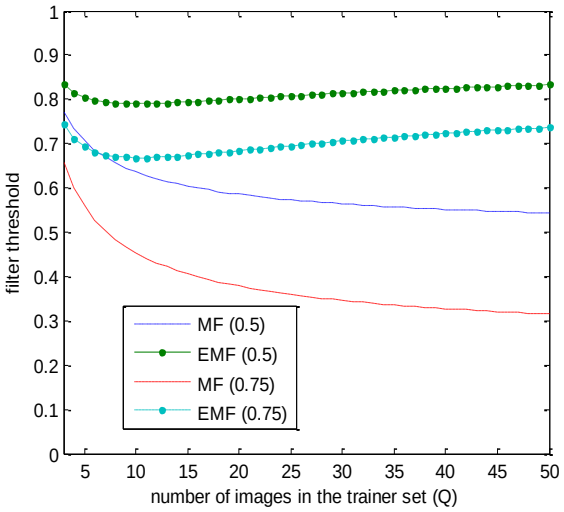


FIGURE 2: Effect of the number of trainers on the expected values of filter thresholds. Numbers in the caption denote trainer set maximum disparity factor □.

The plots of Figure 3 show the Monte Carlo based empirical data related to the effect of □ on filter thresholds, with the number of trainers fixed at Q=50. For each □ setting, one thousand 50×50 correlation matrices were generated, such that each off-diagonal element is an independent sample of the uniform probability distribution in [(1-□), 1]. The true threshold values associated with each correlation matrix are computed using Equations (8,15). The minimum, maximum, and mean thresholds for MF and EMF across one-thousand trials are plotted in Figure 3.

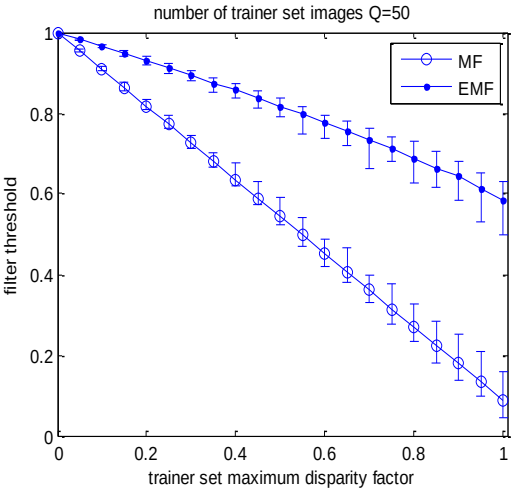
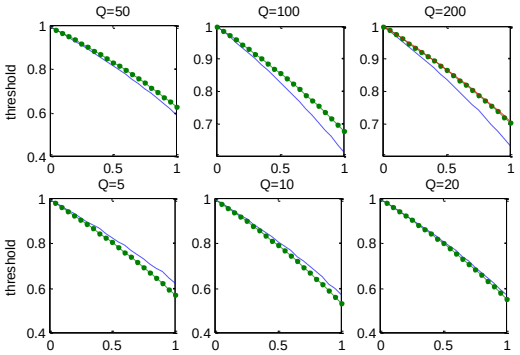


FIGURE 3: Filter thresholds versus trainer set maximum disparity factor □.

The plots of Figure 4 compare the expected values of the EMF threshold based on empirical and analytical formulation of Equations (15) and (45), respectively. It is noted that the analytical formula of Equation (45), which is based on the approximate closed-form formulation of the integral in (44), yields very accurate results. The plots of Figure 4 show that the expression in (45), slightly

overestimates the true threshold for small values of Q, and it slightly underestimates the true threshold for larger values of Q. The first-order approximation used in (38) does not include the effect of the mutual correlation of random variables X, U in (35), which partially accounts for the discrepancy between the analytical and simulated results of Figure 4. However, if one utilizes the third-order expansion of the $\frac{X}{U}$ function in (35), and considers their mutual correlation coefficient and respective variance expressions, it will become clear that the approximation of (45) underestimates the true value of the threshold for large values of Q and overestimates the threshold for smaller values of Q.



FIGUR 4: Expected EMF threshold versus maximum trainer set disparity factor □. The solid line with dots is based on Monte Carlo simulations, and the dashed line is obtained from the analytical formula. Numbers at the top of plots denote trainer set populations. The horizontal axes represent maximum disparity factor.

4. Comparisons of the MF and EMF Thresholds Using Synthetic Images

To demonstrate the superior efficacy of EMF in comparison to MF, several simulations were conducted using computer generated virtual images. In the example of Figure 5, the set of target-class training images comprises nine randomly generated white rectangles of arbitrary dimensions and aspect ratios against black background. For each trainer, the dimensions of the black frame and the center coordinates of the white rectangle are chosen randomly. The trainers are shown on the left where the rightmost trainer in the top-row is the MF correlation filter with threshold T_{MF}=0.4097 as determined by Equations (6-8). The image on the right shows the EMF correlation filter with threshold T_{EMF}=0.6697, obtained in accordance with Equations (12, 15) of Section 2. It is noted that, in this example, the EMF threshold is substantially higher than the MF threshold. This means, for comparable false-negative error rates, the EMF has substantially lower false-positive error in comparison to the MF.

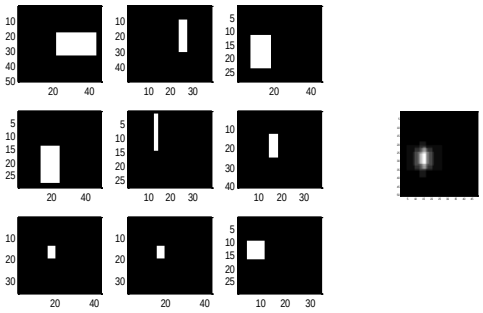


FIGURE 5: Nine images on the left represent the trainer set. The MF template is the rightmost image in the top row. The image on the right is the EMF templet.

Next, a geometric interpretation of the image hyperspace is presented to illustrate the relationship between the trainer set disparity factor and the filter thresholds for MF and EMF. A geometric interpretation of the effect of threshold on filter performance is also presented. Each image is envisaged as a point in the hyperspace of all possible images. The mutual disparity factor involving an image pair is expressed as $\delta = 1 - \lambda$, where δ denotes the mutual disparity factor or distance between the image pair and λ is the peak cross-correlation between the two corresponding normalized images. The mutual disparity factor is akin to the Euclidean distance between two points representing the respective images in the image hyperspace.

The hypersphere centered at the MF with radius (1-T_{MF}) in the image hyperspace has a much larger volume than that centered at the EMF with radius (1-T_{EMF}), since T_{MF} is generally much smaller than the corresponding T_{EMF} as seen from the examples of Figures 1-5. The larger volume surrounding the MF, inevitably, leads to higher probability of capturing non-target images, and therefore higher false-positive error rate. The probabilities of capturing non-trained-on target class images, however, are comparable for two filters since both hyperspheres encompass all the trainers. In the example of Figure 5 for instance, the hyperspheres centered at the MF and EMF, with corresponding radii, both capture all the nine trainers. This leads to comparable probabilities of false-negative errors for the MF and EMF.

In the example of Figure 6 the number of trainers was varied from one to twenty-five. For each setting of the number of trainers, the white rectangles comprising the target-class training set of images were generated as before. The respective MF was identified, and the corresponding threshold was computed. The EMF was also synthesized from the trainers, and the corresponding threshold was computed. The maximum size of each image frames was set at 50×50-pixels. For each setting of the number of trainers, the simulation was repeated

twenty times, and the resulting threshold values, T_{MF} and T_{EMF}, were averages across all twenty trials. The plots of Figure 6 show the average threshold values for the resultant MF and EMF. It is seen that T_{EMF} values are consistently and substantially higher than T_{MF} values. This property, in general, leads to lower false-positive errors for the EMF in comparison to MF, for comparable false-negative errors.

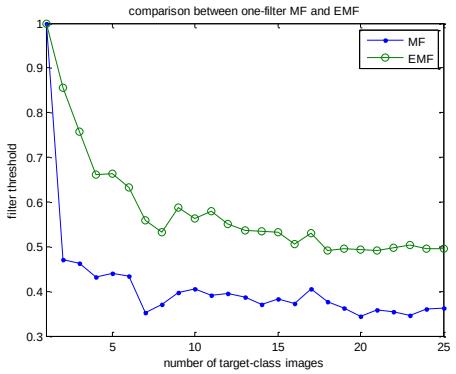


FIGURE 6: Comparison of MF and EMF thresholds.

In the example of Figure 7, for each setting of the nominal number of trainers the number of randomly generated white rectangles is higher by one. The image whose minimum peak correlation with respect to all the generated images in the set is greatest (anchor/prototype image) is eliminated, and the remaining images constitute the training set. As before, the MF and EMF correlation filters and the respective thresholds are computed. The plots of Figure 7 show the comparison between threshold values of the two filters for each setting of the nominal number of trainers. It is seen that the advantage offered by the EMF in this case is even more striking than that of the example of Figure 6. This is a significant observation, since in practical scenarios only a limited number of target-class images are available for training the classifier. It is assumed that the large-scale target-class training set has been properly partitioned into compact clusters each comprised of a small number of images that form a cohesive volume in the image hyperspace. The circumscribing hypersphere in the hyperspace, that contains within its volume all the training images associated with a particular training cluster and has the smallest volume is the best filter for that zone of the hyperspace. Ideally, the trainers in that zone are distributed uniformly; the circumscribing hypersphere is the convex hall of the set; the correlation filter associated with MF constitutes the center of the hypersphere; and the MF threshold is equal to its radius. However, if the hyperspace distribution of the trainers is non-uniform, the circumscribing hypersphere is expected to have a larger volume than that of the convex hull. Formation of the trainer set by removing the prototype from the set of randomly generated images results in an image cluster whose center in the hyperspace does not coincide with a physical image. The MF, which in this case is the prototype image of the reduced set, is one of the remaining images in the cluster. The hypersphere which is centered at the MF and encases all the cluster images has a larger radius (lower threshold) than the previous case and therefore contains more non-target entities. This, in general, will lead to higher false-positive error. Removal of the prototype from the pre-cluster set of images does not, in general, change the overall outline of the constellation in the hyperspace. Constructing the EMF from the reduced set, results in a synthetic image which, in general, is very similar to the EMF constructed from the original set of images in the cluster. In both cases the synthetic image representing the EMF would be at or very close to the center of the convex hull. The hypersphere radius (EMF threshold) is therefore insensitive to the removal of the prototype from the original image set.

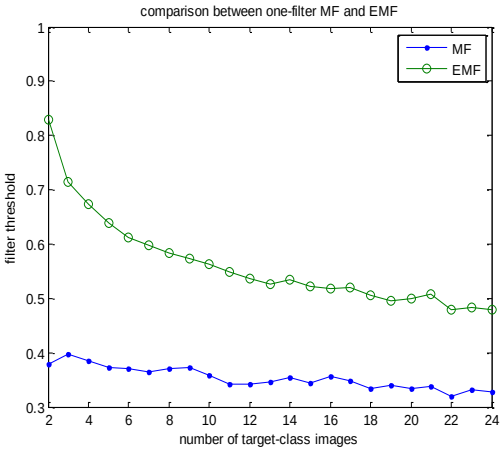


FIGURE 7: Thresholds of MF and EMF versus number of trainers.

The example of Figure 8 further compares the MF and EMF thresholds for different number of trainers under realistic scenarios where the known target-class images (trainers) may be obtained by nonuniform sampling of the target-class image hyperspace. For each setting of the number of target-class images, first the images are generated randomly to form the raw image cluster. Subsequently, one image at a time is peeled off from the raw cluster to arrive at a reduced cluster. Each time, the image with the greatest minimum peak correlation with respect to all the remaining images (cluster prototype) is eliminated from the cluster. In the plots of Figure 8, the abscissa represents the number of eliminated images from the raw cluster to arrive at the actual trainer set, from which the MF and EMF are computed. The number next to each plot denotes the number of original images. The plots of Figure 8 illustrate the deleterious effect of the nonuniform sampling of the target-class image hyperspace for formation of the trainer set on the MF. It is also seen that this nonuniform sampling, which is to be expected in all practical scenarios, has virtually no effect on the EMF.

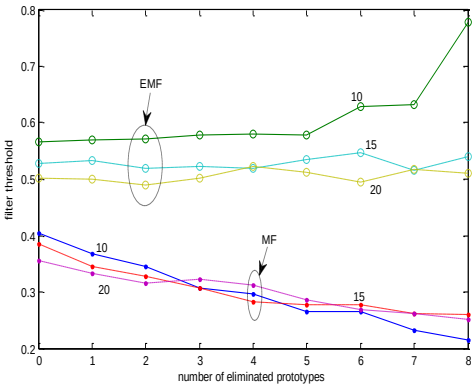
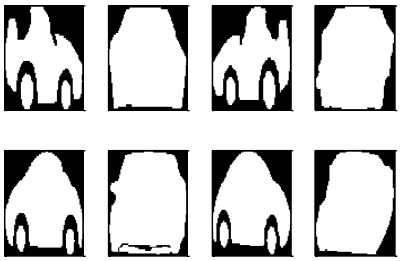


FIGURE 8: Effects of non-uniform sampling of the target class image hyperspace on MF and EMF thresholds.

5. Comparison of the MF and EMF Performance Using Real Images

The target and non-target classes in this section denote two user-specified sets of image masks stored in the Amsterdam Library of Object Images (ALOI), described in [57]. The database from which the training and test images are derived is in [58]. In the experiment of Figures 9-13, image masks associated with objects number-138 *(blue car)* and 136 *(purple car)* in ALOI represent target and non-target classes, respectively. Here, each class comprises seventy-one image masks associated with the respective object, obtained at uniformly spaced view angles five-degrees apart and identical ranges. Figure 9 shows samples associated with the target and non-target classes. Eight images were randomly selected from the target class to form the trainer set. The test set is comprised of 134 images, including 71 non-target and the remaining 63 target class images. The training set was then utilized to select the appropriate MF and to synthesize the EMF in accordance with the procedures of Section 2. Each of the two filters were utilized to classify 134 unlabeled test images. Correct classification (detection) and misclassification rates denote, respectively, the percentage of untrained-on target class test images that are correctly labeled, and the percentage of non-target class images that are erroneously labeled as target.

This experiment was repeated one-hundred times, each time using eight randomly chosen target class images as the training set and the remaining 63 target and 71 non-target class images as the test set. Figure 10 shows the receiver operating characteristic (ROC) plots for the MF and the EMF obtained by averaging across 100 trials of the experiment. The superior performance of the EMF in comparison to the MF is evident from the plots of Figure 10. The EMF achieves 93.65-percent correct classification with zero misclassification, whereas the MF achieves 42.86-percent detection at zero misclassification. At 100-percent detection, the EMF and MF have misclassification rates of, respectively, 0.52 and 7.54-percent. Figure 11 shows the set of eight target-class trainers and the corresponding synthesized EMF associated with four different instantiations of the experiment.



FIGUR 9: Top and bottom rows show four samples of, respectively, object-136 *(purple car)* and object-138 *(blue car)* masks of the ALOI dataset.

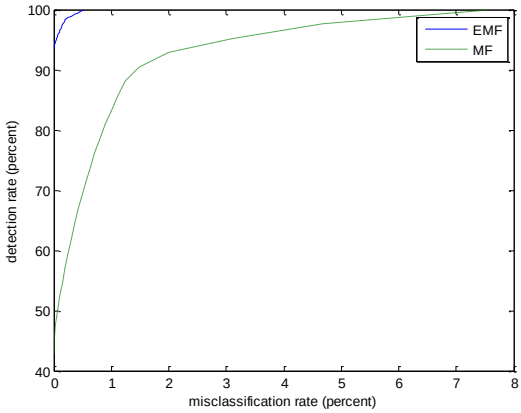


FIGURE 10: Comparison between MF and EMF performance.

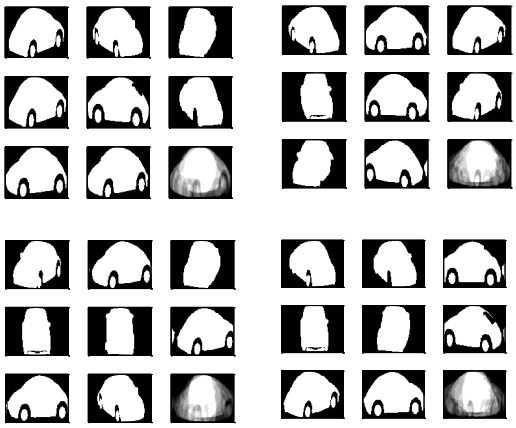


FIGURE 11: Four typical instantiations of the 100 experiments. Each3 × 3-block shows eight trainers and the respective synthesized EMF shown in the lower-right corner of the block.

The experiment was then repeated by reversing the target and non-target class designations. In the new experiment the image masks associated with objects number-136 and 138 represent target and non-target classes, respectively. As before, eight images were randomly selected from the target class to form the

trainer set. The test set is comprised of 134 images including 71 non-target and the remaining 63 target class images. The training set was then utilized to select the appropriate MF and to synthesize the EMF, which were subsequently utilized to classify the images in the test set. Figure 12 shows the ROC plots for two types of filters, obtained by averaging the corresponding filter performance across one-hundred instantiations of the experiment. As expected, the EMF outperforms the MF by a great margin. Figure 13 shows the set of eight target-class trainers and the corresponding synthesized EMF associated with four instantiations of the experiment.

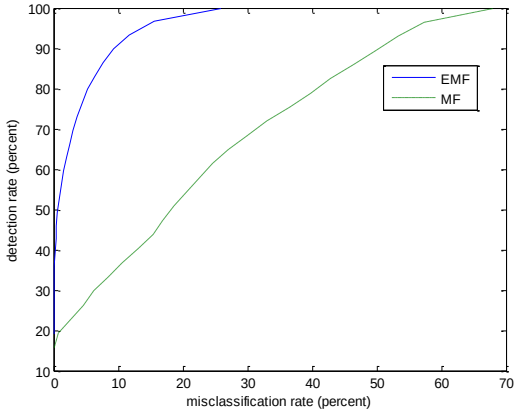


FIGURE 12: Comparison between MF and EMF performance.

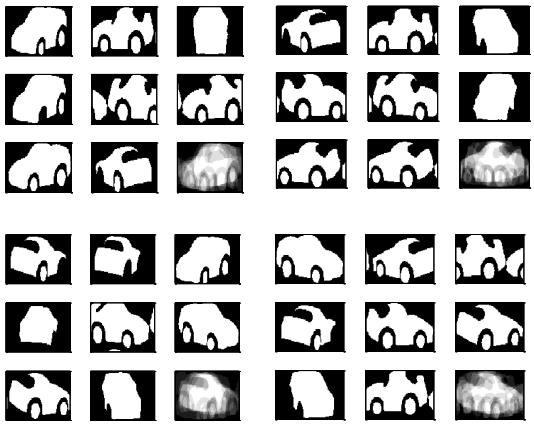


FIGURE 13: Four typical instantiations of the 100 experiments. Each3 × 3-block shows eight trainers and the respective synthesized EMF shown in the lower-right corner of the block.

In the experiment of Figure 14 the classifier zone of effectiveness was restricted by designating the first 35 *purple-car* image masks in ALOI as the target class universe. These are the masks corresponding to view angles between zero and 170-degrees. The training set of images consists of eight randomly selected samples from the 35 *purple-car* masks, and the test set comprises 98 images, including all 71 *blue-car* and the remaining 27 *purple-car* masks. The ROC plots of Figure 14 show the performance of the MF and the EMF for this experiment.

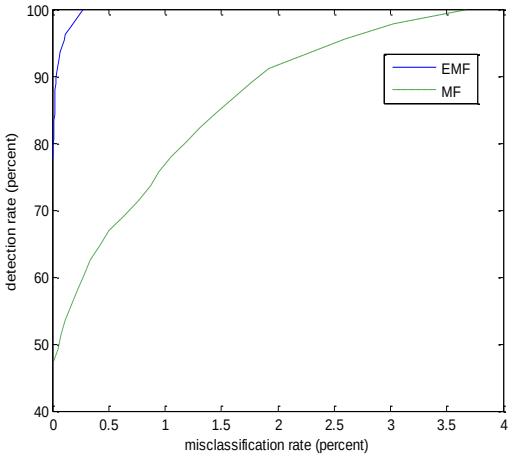


FIGURE 14: Comparison between the MF and the EMF performance.

The plots of Figures 15 and 16 show filter thresholds for the MF and the EMF based on empirical data as well as the theoretical formulas of Section 3. In Figure 15, the objet-136 masks denote the target class, from which a user-prescribed number of images (Q) are randomly chosen to form the trainer set. For each value of Q, the experiment was repeated 200 times, T_{MF} and T_{EMF} were computed and were averaged across all trials. For each value of Q, the theoretical formulas of Section 3, Equations (33) and (45) with maximum disparity factor of □=0.358, were also used to compute the expected values of MF and EMF thresholds. The maximum disparity factor is obtained by subtracting the minimum peak cross-correlation among 71 object-136 images from one. In Figure 16, the object-55 masks denote the target class, for which the maximum disparity factor is □=0.491. In both cases, it is seen that EMF threshold is substantially greater than the MF threshold. This is further confirmation of the analytical results of Section 3 and the empirical results of Section 4, which was based on synthetically generated images. As the number of trainers is increased, the difference between empirical and theoretical values of T_{MF} decreases, as expected. It is also seen that there is a discrepancy between theoretical and empirical T_{EMF} values, which is attributable to the fact that spatially aligning all the trainers with respect to the anchor trainer does not guarantee pair-wise spatial alignment amongst all the trainers, as implied by Equation (15).

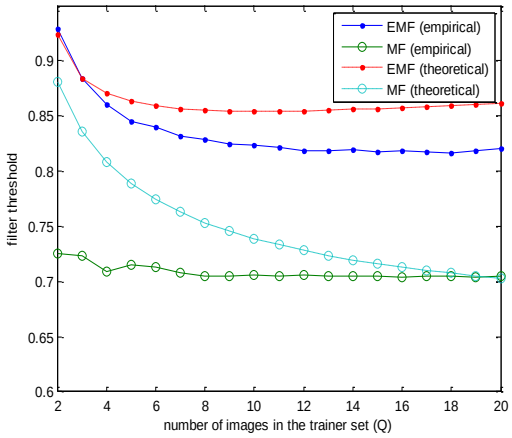


FIGURE 15: Comparisons between empirical and theoretical values of EMF and MF thresholds.

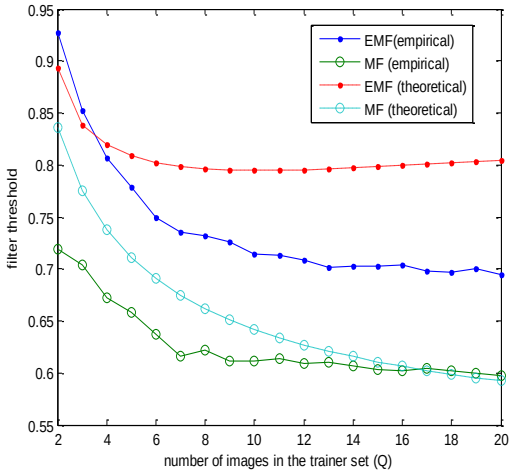


FIGURE 16: Comparisons between empirical and theoretical values of EMF and MF thresholds.

6. Comparison Between EMF and SDF Performance

This section presents a concise performance comparison between two image classifiers based on the enhanced matched filter (EMF) of Section 2 and a synthetic discriminant function (SDF) proposed in reference [47]. The grayscale images of six ALOI objects [57, 58] were used as the basis of the training and test images for this experiment. The original database contains seventy-two 144×192 -pixel images of each object viewed at the same distance (range) and equally spaced view angles five degrees apart. Figure 17 shows six samples of each object viewed at sixty-degree increments. The original images were cropped to minimize the black background and were resized to 128×128-pixel. To synthesize the image bank for training and test images for this experiment, each image was subsequently rescaled to create five different range (distance) scales equally distributed between ten percent below and above the nominal range, thus creating 360 images of each object.

Thirty images of one object consisting of six consecutive view angles, comprising a contiguous twenty-five- degree angular sector and five distances, constitute the target class set of images. Likewise, 1800 images of the five remaining objects represent the non-target class of images. Fifty percent of the target class images are randomly selected to form the training set of images, from which the EMF and the SDF filters are constructed. The remaining target class images are used as non-trained-on target class test images.

The synthesized EMF and SDF filters are used to assign a binary label, namely target or non-target, to each of the thirty test images consisting of fifteen non-trained-on target class and fifteen non-target class images, which are randomly selected from the 1800-image non-target class image set described above. The assigned labels are subsequently compared to the actual class disposition of each test image. The detection and misclassification rates are, respectively, the percentage of target class test images that are correctly classified, and the percentage of non-target class test images that are erroneously classified as target class.

The experiment described above was performed many times, each time using one of the ALOI objects of Figure 17 in a designated twenty-five-degree angular sector as representing the target class and repeating the experiment multiple times by selecting different combinations of the target-class images as trainers, and the remaining as the non-trained-on target class test images. The EMF and SDF filter performance results were averaged across all instantiations of the experiment. The experiments comprised choosing one ALOI object in a designated contiguous twenty-five-degree angular sector as target across all six objects and across many angular sectors. Figure 18 shows the performance results of the EMF and the SDF classifiers in the experiment described above. It is seen that in this experiment the SDF outperforms the EMF for low misclassification rates. However, the EMF achieves higher detections rates at lower misclassification rates compared to the SDF classifier. The performance of the MF, which is not shown here, was much inferior to both the EMF and the SDF.

The set of experiments explained above was repeated, where in each instantiation of the experiment one-third of the target-class images were utilized as trainers and the remaining two-third as non-trained on target class test images. The performance results of the EMF and SDF classifiers for this experiment are also shown in Figure 18. It is shown that when the number of training images was reduced from fifteen to ten, the EMF clearly outperforms the SDF.

The limited experimental results presented here show that in application areas where the number of

training images are small, and where deep learning convolutional neural network-based classifier solutions may not be viable, the EMF classifier could present a viable solution.

FIGURE 17: Samples of ALOI images of objects-136, 138, 148, 151, 154, 160 used for training and testing of image classifiers based on EMF and SDF.

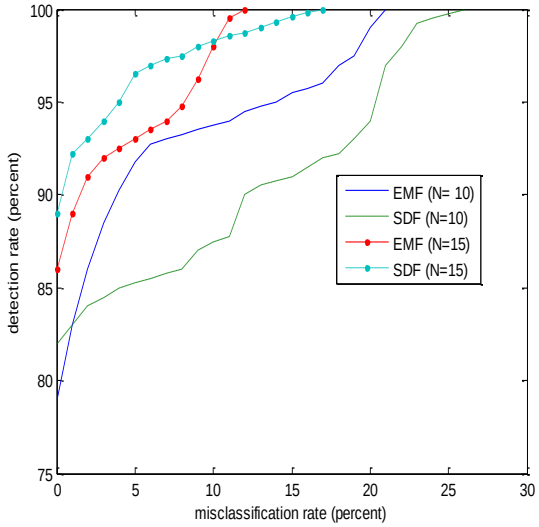


FIGURE 18: Comparison between EMF and SDF performance. Numbers in the legend denote number of training images in each instantiation of the experiment.

7. Conclusions

An efficient algorithm for the synthesis of an adaptive and robust distortion tolerant correlation filter for autonomous vision system applications has been presented. The enhanced matched filter (EMF) is obtained by pixel-wise addition of the appropriately compensated target class training set of images associated with the prescribed target states and view conditions. The EMF replaces the training set of images, for detecting and locating the target of interest in new sensor images, thereby reducing the operation phase memory requirement and computational complexity. The efficient construction mechanism allows the EMF to be adaptively improved as new episodes of the target of interest are detected, and this makes EMF unique among the wide array of distortion tolerant correlation filters reported in the literature. The mathematical formulas for computing the expected values of the EMF and the matched filter (MF) thresholds have been developed. Computed values of the EMF and the MF thresholds based on analytical formulas have been compared to empirical results. Performance comparisons between the EMF and MF using analytical formulations as well as synthetic and real images have been presented. Performance comparisons between the EMF and the SDF classifier have also been presented.

Future work will include investigation of the effects of dynamic adaptation of the EMF and sensor noise on the filter performance. Development and implementation of algorithms for optimal partitioning of the large-scale training set of images into multiple smaller-scale clusters of trainers with constrained zones of effectiveness will also be investigated. A properly devised cluster of trainers fashioned from a larger trainer set characterizes the target of interest subject to constrained states and view conditions. The EMF based on a particular cluster of trainers has a zone of effectiveness which corresponds to the respective target states and view conditions. The zone of effectiveness of a suitably designed set of EMFs comprising multiple filters is equivalent to the target states and view conditions encompassed by the large-scale set of trainers.

Acknowledgements

The author was partially supported by U.S. Department of Defense through contract W911NF-13-1-0136.

References

[1] E.R. Davies. *Computer and Machine Vision: Theory, Algorithms, Practicalities*, 2nd ed. Waltham: Academic Press, 2012.
[2] D.A. Forsyth, J. Ponce. *Computer Vision: A Modern Approach*, 2nd Ed. Upper Saddle River: Prentice Hall, 2012.
[3] R. Szeliski. *Computer Vision: Algorithms and Applications*. New York: Springer, 2011.
[4] B.G. Batchelor. *Machine Vision Handbook*, pp.565-623. New York: Springer, 2012.
[5] C. Steger, M. Ulrich, C. Wiedemann. *Machine Vision Algorithms and Applications*. Hoboken: Wiley, 2007.
[6] K. He, G. Gkioxari, P. Dollar, R. Girshick, " Mask R-CNN." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[7] A. Bochkovskiy, C.Y.Wang, H. M, Liao, " YOLOv4: Optimal Speed and Accuracy of Object Detection."

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[8] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 6154–6162, 2018.
[9] M. Lokanath, K.S. Kumar, E.S. Keerthi" Accurate object classification and detection by faster-RCNN." IOP Conference Series Materials Science and Engineering 263(5), 2017.
[10] A. Karpathy, T. Leung, R. Sukthankar, G. Toderici, S. Shetty, L. Fei-Fei. "Large-Scale Classification with Convolutional Neural Networks." IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[11] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, A. Oliva. "Learning Deep Features for Scene Recognition using Places Database." Advances in Neural Information Processing Systems in NIPS Proceedings 27, 2014.
[12] Y. LeCun, Y. Bengio, Y., G. Hinton. "Deep Learning." Nature, Vol. 521, No.7553, pp 436-444, 2015.
[13] A. Shrivastava, T. Pfister, O. Tuzel, J. Suskind, W. Wang, R. Webb. "Learning from Simulated and Unsupervised Images Through Adversarial Training." Computer Vision and Pattern Recognition CVPR, pp. 2107-2116, 2017.
[14] K. Li, W. Ma, U. Sajid, Y. Wu, G. Wang, " Object Detection with Convolutional Neural Networks." Computer Vision and Pattern Recognition (CVPR) arXiv:1912.01844, 2019.
[15] J. Chen, Z. Wang, H. Li. " Real-time object segmentation based on convolutional neural network with saliency optimization for picking." Journal of Systems Engineering and Elec., Volume 16, Issue (6) pp. 1300-1307, 2018.
[16] A. Kumar, S. Srivastava, " Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector." Procedia Computer Science, Vol 171, pp. 2610-2617 (2020).
[17] G. Turin. "An introduction to matched filters." IRE Transactions on Information Theory, vol. 6, Issue 3, pp. 311-329, 1960.
[18] A. Papoulis. Signal Analysis. New York, McGraw-Hill,1977.
[19] J.W. Goodman. Introduction to Fourier Optics, 2nd Ed. New York, McGraw Hill, 1996.
[20] H.L. Van Trees. Detection, Estimation, and Modulation Theory, Volume 1. Hoboken, John Wiley & Sons, 2001.
[21] L.M. Novak, G.L. Owirka, C.M. Netishen. "Radar Target Identification using Spatial Matched Filters." Pattern Recognition, vol. 27, Issue 4, pp. 607-617, 1994.
[22] A. VanderLugt. Optical Signal Processing. Hoboken, Wiley,1992.
[23] A. VanderLugt. "Signal detection by complex filters." IEEE Trans. Information Theory, vol. 10, Issue 2, pp. 139-145, 1964.
[24] D.L. Flannery, J.L. Horner. "Fourier optical signal processor." Proc. IEEE vol. **77,** Issue 10, **pp.** 1511-1527, 1989.
[25] J.L. Horner, P.D. Gianino. "Phase-only matched filtering." Applied Optics, vol. 23, Issue 6, pp. 812-816, 1984.
[26] K. Heidary, H.J. Caulfield. "Nonlinear Fourier Correlation." Proc. SPIE 7340, Optical Pattern Recognition XX, pp. 1-11, 2009. doi:10.1117/12.818256
[27] V. Kober, M. Mozerov, I.A. Ovseevich. "Adaptive correlation filters for pattern recognition." Pattern Recognition and Image Analysis, vol. 16, Issue 3, pp. 425-431, 2006.
[28] P.N. Belhumeur, D.J. Kriegman. "What is the set of images of an object under all possible illumination conditions?" International Journal of Computer Vision, vol. 28, Issue3, pp. 1-16, 1998.
[29] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman. "From few to many: Illumination cone models for face recognition under variable lighting and pose." IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, Issue 6, pp. 643-660, 2001.
[30] K.J. Dana, B.V. Ginneken, S.K. Nayar, J.J. Koenderink. "Reflectance and texture of real-world surfaces." ACM Trans. Graphics, vol. 18, no. 1, pp. 1-34, 1999.
[31] T.M. Caelli, Z.Q. Liu. "On the minimum number of templates required for shift, rotation and size invariant pattern recognition." Pattern Recognition, vol. 21, no. 3, pp. 205-216, 1988.
[32] P.W. Hallinan. "A Deformable Model for Face Recognition Under Arbitrary Lighting Conditions." Ph.D. dissertation, Harvard Univ., Cambridge, MA,1995.
[33] R.C. Hoover, A.A. Maciejewski, E.G. Roberts. "Fast Eigenspace Decomposition of Images of Objects with Variation in Illumination and Pose." IEEE Trans. Systems, Man, and Cybernetics, vol. 41, Issue 2, pp. 318-329, 2011.
[34] D. Hao, S.M. Seitz, N. Snavely. "The Dimensionality of Scene Appearance." Proc. 12th IEEE Int. Conf. Computer Vision, pp. 1917-1924, 2009.
[35] S. Landeau, T. Dagobert. "Image database generation using image metric constraints: an application within the CALADIOM project." Proc. SPIE 623410, Automatic Target Recognition XV1, pp. 1-12, 2006.
[36] H.J. Caulfield, W.T. Maloney. "Improved discrimination in optical character recognition." Applied Optics, vol. 8, Issue 11, pp. 2354-2356, 1969.
[37] C.F. Hester, D. Casasent. "Multivariant technique for multiclass pattern recognition." Applied Optics, vol. 19, Issue 11, pp. 1758-1761, 1980.
[38] D. Casasent, D. Psaltis. "Position, Rotation, and Scale Invariant Optical Correlation." Applied Optics, Volume 15, Issue 7, pp. 1795-1799, 1976.
[39] Y.N. Hsu, H.H. Arsenault, G. April. "Rotation-invariant digital pattern recognition using circular harmonic expansion." Applied Optics, vol. 21, Issue 22, pp. 4012-4015, 1982.
[40] D. Mendlovic, E. Marom, N. Konforti. "Shift and scale invariant pattern recognition using Mellin radial harmonics." Optics Communications, vol. 67, Issue 3, pp. 172-176, 1988.
[41] K. Heidary, H.J. Caulfield. "Needles in a Haystack: Fast spatial search for targets in similar-looking backgrounds." Journal of the Franklin Institute, vol. 349, Issue10, pp. 2935-2955, 2012.
[42] K. Heidary, H.J. Caulfield. "Application of supergeneralized matched filters to target classification." Applied Optics, vol. 44, Issue 1, pp. 47-54, 2005.

[43] Y. Sheng, J. Duvernoy. "Circular-Fourier-radial-Mellin transform descriptors for pattern recognition." JOSA A, vol. 3, Issue 6, pp. 885-888, 1986.
[44] S.H. Hong, B. Javidi. "Optimum nonlinear composite filter for distortion-tolerant pattern recognition." Applied Optics, vol. 41, issue. 11, pp. 2172-2178, 2002.
[45] H.J. Caulfield, M.H. Weinberg. "Computer recognition of 2-D patterns using generalized matched filters." Applied Optics, vol. 21, Issue 9, pp. 1699-1704, 1982.
[46] R.B. Johnson, K. Heidary. "A unified approach for database analysis and application to ATR performance metrics." Proc. SPIE 7696, Automatic Target Recognition XX, 2011. doi:10.1117/12.849942
[47] B.V.K. Kumar. "Minimum-variance synthetic discriminant functions." JOSA A, vol. 3, Issue 10, pp. 1579-1584, 1986.
[48] A. Mahalanobis, B.V.K. Kumar, D. Casasent. "Minimum average correlation energy filters." Applied Optics vol. 26, Issue 17, pp. 3633-3640, 1987.
[49] D. Casasent, G. Ravichandran, S. Bollapragada. "Gaussian-minimum average correlation energy filters." Applied Optics, vol. 30, Issue 35, pp. 5176-5181, 1991.
[50] D. Casasent, G. Ravichandran. "Advanced distortion-invariant minimum average correlation energy (MACE) filters." Applied Optics, vol. 31, Issue 8, pp. 1109-1116, 1992.
[51] M. Savvides, B.V.K. Kumar. "Quad Phase Minimum Average Correlation Energy Filters for Reduced Memory Illumination Tolerant Face Authentication." *Lecture Notes in Computer Science*, vol. 2688, pp. 1056-1065, 2003.
[52] N. Muller, B.M. Herbst. "On the use of SDF-type Filters for Distortion Invariant Image Location." Proceedings of the 15th International Conference on Pattern Recognition, Volume 3, 2000.
[53] N. Muller, B.M. Herbst. "On the use of SDF-type filters for distortion parameter estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, issue 11, pp. 1521-1528, 2002.
[54] K. Heidary. "Distortion tolerant correlation filter design." Applied Optics, vol. 52, No. 12, pp. 2570-2576, 2013.
[55] K. Heidary. "Design of State Invariant Target Classifiers." Proceedings of the International Conference on Computer and Information Science and Technology, Ottawa, Ontario, Canada, 2015.
[56] A. Papoulis, S.U. Pillali. *Probability, Random Variables and Stochastic Processes, 4th Edition*. New York, McGraw Hill, 2002.
[57] J.M. Geusebroek, G.L. Burghouts, A.W.M. Smeulders. "The Amsterdam library of object images." Int. J. Comput. Vision, vol. 61, Issue 1,pp. 103-112, 2005.
[58] https://aloi.science.uva.nl/

Appendix A

The enhanced matched filter and the corresponding threshold are computed in accordance with the procedure outlined below.

Input: Training image set $g_q(m,n)$.

Output: The EMF comprising the filter template and corresponding threshold $h_{EMF}(m,n), T_{EMF}$.

1. Normalize each trainer image such that for each image the sum of pixel intensities is zero and the sum of squares of pixel intensities is one.
2. For the normalized images of step-one, obtain the peak cross correlations among all image pairs. Identify the anchor trainer. This is the normalized trainer with the greatest minimum peak cross-correlation with respect to all the other normalized trainers. If more than one such trainer exists, pick one of them randomly and designate it as the anchor trainer.
3. Spatially shift each normalized trainer such that its peak cross correlation with respect to the anchor trainer is at pixel-(1,1) and renormalize the spatially shifted images in accordance with Step-1.
4. Add all the spatially shifted trainers, including the anchor trainer, and renormalize to obtain h_{EMF}.
5. The minimum peak cross correlation of the h_{EMF} with respect to all normalized trainers is T_{EMF}

Appendix B

In the analysis of Section 3.2 the following integral is encountered.

$$I_{N,P} = \int_0^1 (1-x^N)^P dx \qquad \text{(B-1)}.$$

Where, N, P are positive integers. As far as the author can ascertain, the closed-form expression for the integral in (B-1) has not been listed in any of the standard integral tables. The integral can be evaluated in closed form as shown below.

$$I_{N,P} = \int_0^1 (1-x^N)^{P-1} dx - \int_0^1 x x^{N-1}(1-x^N)^{P-1} dx \qquad \text{(B-2)}.$$

Applying integration by parts to the second integral in (B-2), after simplification, one arrives at the following.

$$I_{N,P} = \int_0^1 (1-x^N)^{P-1} dx - \frac{1}{PN} I_{N,P} \qquad \text{(B-3)}.$$

$$I_{N,P} = \frac{1}{1+\frac{1}{NP}} \int_0^1 (1-x^N)^{P-1} dx \qquad \text{(B-4)}.$$

$$I_{N,P} = \frac{NP}{1+NP} I_{N,(P-1)} \qquad \text{(B-5)}.$$

The reduction formula of (B-5) is repeatedly applied until one arrives at $I_{N,1}$ for the integral at the right-hand side. This leads to a closed-form expression for the integral in (B-1).

$$I_{N,1} = \frac{N}{N+1} \qquad \text{(B-6)}.$$

$$I_{N,P} = \frac{N^P P!}{\prod_{p=1}^{P}(1+pN)} \qquad \text{(B-7)}.$$