A machine learning approach to hyperspectral detection of solid targets

Amanda Ziemann[a], Michal Kucer[a,b], and James Theiler[a]

[a]Intelligence and Space Research, Los Alamos National Laboratory, Los Alamos, NM 87545
[b]Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, NY 14623

ABSTRACT

We describe and compare two approaches to solid subpixel target detection in hyperspectral imagery. The first approach requires explicit models for both the target and the background, and employs a generalized likelihood ratio in order to obtain a detector that is optimized to those specific models. When this approach is most successful, a closed-form solution is obtained that permits the detector to be efficiently applied. A specific example of this approach is outlined in some detail, leading to the elliptically-contoured finite-target matched filter (EC-FTMF), a variant of the classical FTMF algorithm that uses a multivariate t-distribution instead of a Gaussian as the model for the background. The second approach also requires an explicit model of the target, but does not need a model for the background. In this second approach, matched pairs of data samples are created: for each pixel in the original hyperspectral image, a corresponding pixel is generated by implanting the target into the original pixel. These matched pairs are used as training data for a machine learning algorithm to classify pixels as either non-target or target. Here we use a support vector machine, but the matched pair machine learning (MPML) framework does not restrict the choice of classifier type. Detectors using both approaches are applied both to simulated data (with Gaussian and with multivariate t distributed backgrounds) and to real hyperspectral data with known, referenced targets.

Keywords: hyperspectral, target detection, replacement model, matched pairs, machine learning, EC-FTMF

1. INTRODUCTION

Many materials that are of civilian and military interest are difficult to distinguish using traditional cameras, as those cameras are designed to recreate what we see visually (by measuring the "red"-ness, "green"-ness, and "blue"-ness of the pixels in the scene). These materials, however, often possess rich information in narrow visible and non-visible channels of the electromagnetic spectrum, which can be captured by hyperspectral cameras. When deployed on airborne or spaceborne platforms, the high spectral resolution of hyperspectral imagery enables remote discrimination of materials. In the target detection regime, we exploit this spectral resolution for the detection of image pixels that are likely to contain the target material. In the work presented here, although we limit our examples to detection of opaque targets, the focus is in fact this: the development of a broader framework for target detection that enables us to use more sophisticated physics-based and data-driven models to produce detection algorithms that are not confined to simplistic assumptions.

1.1 Traditional Target Detection Algorithms

Traditional target detection algorithms (not restricted to HSI data) begin with specific models of the target, of the background, and of the interaction between the two. The target is usually the most well known *ab initio*, sometimes characterized by a unique spectral signature, sometimes by a family of signatures (with the family in turn sometimes characterized by a subspace or a simplex or a probability distribution), and often informed by known physical or chemical properties. The background, by contrast, exhibits a complexity that is not straightforwardly amenable to direct modeling. Instead, the background is generally characterized by some kind of parametric or nonparametric data-driven model, which can be very effective because there is often a lot of data available to drive that model, which is generally (though not exclusively) some kind of probability distribution. Finally, there is the interaction of target and background. The most common interaction model is additive, with

---

Author emails: AZ, ziemann@lanl.gov; MK, mxk7721@rit.edu; JT, jt@lanl.gov.

the observed pixel containing signal that is simply the background *plus* what the target adds to that signal. A linear replacement model can be more appropriate for solid subpixel targets wherein the target signal is a fraction of the total pixel signal, and even more complicated models might be appropriate for heavy chemical plumes or thin layers of semi-transparent powder. In general, if $\mathbf{b} \in \mathbb{R}^d$ is the observed spectrum of the background clutter, we will write $\mathbf{x} = \xi(\mathbf{b})$ as the observed pixel spectrum when target is present.

1.2 Additive Target Model

In particular, when the target is additive, we write

$$\mathbf{x} = \xi(\mathbf{b}) = \mathbf{b} + \epsilon \mathbf{t} \tag{1}$$

where $\mathbf{t}$ is the target signal (modeled here as a unique spectral signature), and ϵ is the measure of the strength of that signal. Depending on the model for the background, the additive model can lead to the adaptive matched filter (AMF),[1–3] the adaptive coherence (or "cosine") estimator (ACE),[4,5] or the elliptically-contoured generalized likelihood ratio test (EG-GLRT) detector.[6]

Although it makes sense for radar signal processing,[7] and is plausible for modeling the effect of weak gas-phase chemical plumes,[8–11] the additive model is "phenomenologically inconsistent" with opaque target materials.[12]

1.3 Replacement Target Model

The replacement target model is given by

$$\mathbf{x} = \xi(\mathbf{b}) = (1 - \alpha)\mathbf{b} + \alpha \mathbf{t} \tag{2}$$

where $0 \le \alpha \le 1$ is the fraction of a pixel covered by the opaque target with spectral signature $\mathbf{t}$. Algorithms designed to respect the phenomenology of opaque targets include the mixture-tuned matched filter,[13] the false-alarm mitigating matched filter,[14,15] and the adaptive residual[16] (note that if we are in a regime where α is known to be small, then the matched filters designed for the additive target model can still be effective). A rigorous derivation of a detector that is based explicitly on the replacement model in Eq. (2) and on the assumptions of a fixed target and a Gaussian backgound is the finite-target matched filter (FTMF) of Schaum and Stocker.[17] This has been extended to a variable target, though with the restriction that target covariance is proportional to the background covariance,[15] a situation described by Schaum as *ampliskedastic*.[18] Variants of this extension have also been derived by using different flavors of clairvoyant fusion instead of the more traditional generalized likelihood ratio test.[18,19]

1.4 EC-FTMF

In this section, we sketch out the derivation of an elliptically-contoured finite-target matched filter (EC-FTMF),[20] which follows the derivation of the FTMF, but applied to a background with an elliptically-contoured distribution. We also show that the EC-FTMF reduces to the FTMF in the limit as the background becomes more Gaussian. We begin with an expression for the multivariate t distribution:

$$t_\nu(\mathbf{b}|\boldsymbol{\mu}, R) = \frac{\Gamma\left((\nu + d)/2\right)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|R|^{1/2}} \left[1 + \frac{1}{\nu - 2}(\mathbf{b} - \boldsymbol{\mu})^\mathsf{T} R^{-1}(\mathbf{b} - \boldsymbol{\mu})\right]^{-(\nu+d)/2} \tag{3}$$

where ν represents the degrees of freedom, and μ and R are mean and covariance of the data respectively. The utility of EC distributions for hyperspectral data has been previously noted.[6,10,21–23] The detector is formulated via the hypothesis testing framework, with the null and alternative hypotheses given by

$$\mathcal{H}_0 : \alpha = 0, \tag{4}$$

$$\mathcal{H}_1 : \alpha > 0. \tag{5}$$

If $p_x(\mathbf{x}|\mathcal{H}_0)$ and $p_x(\mathbf{x}|\mathcal{H}_1)$ are the conditional probability distributions for the observed spectrum $\mathbf{x}$ under the null and alternative hypotheses respectively, then the likelihood ratio test provides a formula for the detector:

$$\mathcal{D}(\mathbf{x}) = \frac{p_x(\mathbf{x}|\mathcal{H}_1)}{p_x(\mathbf{x}|\mathcal{H}_0)}. \tag{6}$$

Since α is unknown to us beforehand, we use the Generalized Likelihood Ratio Testing framework[24] and use the maximum likelihood estimate of α; thus,

$$\mathcal{D}(\mathbf{x}) = \frac{\max_\alpha p_x(\mathbf{x}|\alpha)}{p_x(\mathbf{x}|0)} = \frac{p_x(\mathbf{x}|\hat\alpha)}{p_x(\mathbf{x}|0)}, \tag{7}$$

where $\hat\alpha$ is the maximum likelihood estimator for α, given by

$$\hat\alpha = \operatorname*{argmax}_\alpha p_x(\mathbf{x}|\alpha). \tag{8}$$

Expressing the background pixel $\mathbf{b} = (\mathbf{x} - \alpha\mathbf{t})/(1-\alpha)$ as a function of fraction α, pixel measurement $\mathbf{x}$, and target spectrum $\mathbf{t}$, we get that

$$p_x(\mathbf{x}|0) = p_b(\mathbf{x}), \quad \text{and} \tag{9}$$

$$p_x(\mathbf{x}|\alpha) = (1-\alpha)^{-d} p_b\left(\frac{\mathbf{x} - \alpha\mathbf{t}}{1-\alpha}\right), \tag{10}$$

where $p_b(\mathbf{x}) = t_\nu(\mathbf{x}|\boldsymbol{\mu}, R)$ is the background distribution. In particular, the expression for $p_x(\mathbf{x}|\alpha)$ is given by

$$p_x(\mathbf{x}|\alpha) = c \times (1-\alpha)^{-d}\left[1 + \frac{(1-\alpha)^{-2}}{\nu - 2}(\mathbf{z} - \alpha\mathbf{u})^\mathsf{T}(\mathbf{z} - \alpha\mathbf{u})\right]^{-(\nu+d)/2} \tag{11}$$

where c is a constant that depends on ν, p, and $|R|$, and $\mathbf{z} = R^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ and $\mathbf{u} = R^{-1/2}(\mathbf{t} - \boldsymbol{\mu})$ are the whitened versions of the measured pixel and the target. To find the α that maximizes $p_x(\mathbf{x}|\alpha)$, we take the derivative and set it to 0. When this is done (excruciating details are in the Appendix), we obtain a quadratic equation:

$$A(1-\alpha)^2 + B(1-\alpha) + C = 0 \tag{12}$$

with coefficients given by[20]

$$A = (\mathbf{t} - \boldsymbol{\mu})^\mathsf{T} R^{-1}(\mathbf{t} - \boldsymbol{\mu}) + (\nu - 2) \tag{13}$$

$$B(\mathbf{x}) = \left(1 - \frac{\nu}{d}\right)(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{t} - \boldsymbol{\mu}) \tag{14}$$

$$C(\mathbf{x}) = -\frac{\nu}{d}(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{x} - \mathbf{t}). \tag{15}$$

Taking the limit $\nu \to \infty$, we recover the coefficients corresponding to the FTMF model:[17]

$$A = 1 \tag{16}$$

$$B(\mathbf{x}) = -(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{t} - \boldsymbol{\mu})/d \tag{17}$$

$$C(\mathbf{x}) = -(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{x} - \mathbf{t})/d. \tag{18}$$

From Eq. (12), we have

$$\hat\alpha(\mathbf{x}) = 1 - \frac{-B(\mathbf{x}) + \sqrt{B^2(\mathbf{x}) - 4AC(\mathbf{x})}}{2A} \tag{19}$$

and finally, the EC-FTMF detector is given by

$$\mathcal{D}(\mathbf{x}) = \frac{p_x(\mathbf{x}|\hat\alpha(\mathbf{x}))}{p_x(\mathbf{x}|0)} \tag{20}$$

with p_x defined in Eq. (11).

1.5 Limitations of Traditional Target Detection Algorithms

A strength of this traditional approach to the design of target detection algorithms is that one can adapt different models of target and background and target/background interaction to the specific physical scenarios of interest. A problem that arises in practice, however, is that expressions for these different models can lead to unwieldy algebra, and their combination into a single likelihood (or generalized likelihood) ratio can be difficult. Closed-form solutions are desirable, but in order to achieve them, simplifying assumptions are sometimes made, even when the practitioner knows better (*e.g.*, that the background isn't really Gaussian, that the target isn't really fixed at a single spectral signature, or that the background-target interaction isn't really linear).

2. MATCHED-PAIR FRAMEWORK FOR TARGET DETECTION

There is a growing abundance of hyperspectral image data available from a variety of sources in both government and commercial sectors. Ground truth for this data is not, however, growing as rapidly. Machine learning has the potential to add enormous value, but the usual scenario for machine learning exploitation – as illustrated in Fig. 1(b) – requires an adequate number of *labeled* samples to estimate the statistics of the classes of interest. For target detection, those classes are background and target, and while background pixels are plentiful, reliably-labeled target pixels are rare to nonexistent – the situation illustrated in Fig. 1(c). An example of this more challenging situation occurs when we consider the problem of target detection in the SHARE 2012 dataset,[25] which we will describe in more detail in Section 3.2. Here, the number of samples for the target is much smaller than for the background (*e.g.*, the blue felt panels comprise a target class with only 72 pixels out of the total ∼48,000 pixels in the test image, or roughly 0.15% of the area of the image).

Using the Matched Pair Machine Learning (MPML) framework,[26, 27] we couple statistical sampling with physical modeling (which is incorporated into the ξ function), and transform the target detection problem back into the realm of classification – as illustrated in Fig. 1(d) – allowing us to use contemporary algorithms to obtain powerful detectors. This amounts to *implanting* target into the background scene, but instead of being used to evaluate the quality of the imagery[28] or algorithm,[29] it is used as a tool to optimize the performance of the detection algorithm.

2.1 Matched-Pair Machine Learning

In the classification framework, we seek a function $f(\mathbf{x})$ that assigns to each data sample $\mathbf{x} \in \mathcal{X}$ a label $y \in \{-1, +1\}$. Formally, we assume we are given a set of data $\mathcal{X} = \{\mathbf{x}_i\}$, with $\mathbf{x}_i \in \mathbb{R}^d$. For MPML, we use the physics-informed function $\xi : \mathbb{R}^d \to \mathbb{R}^d$ to describe the effect of the "treatment" that modifies a target-free sample $\mathbf{x} \in \mathcal{X}$ by implanting a target in that sample. In this formulation, the treatment ξ is known to us, but we do not assume or have any knowledge of the underlying data distribution p_x from which $\mathbf{x}$ is presumed to be drawn. With $\xi(\mathbf{x})$ we create matched pairs of data, where a dataset

$$\mathcal{X}_{matched} = \{(\mathbf{x}_i, -1)\} \cup \{(\xi(\mathbf{x}_i), +1)\}$$

is used to learn a function $f : \mathbb{R}^d \to \{-1, +1\}$ that allows us to distinguish $\mathbf{x}$ and $\xi(\mathbf{x})$. In the case of hyperspectral detection, the treatment ξ will be modeled after the effects that the various types of targets have on the spectrum of a pixel.

The problem is set up as follows: we are given a hyperspectral image $\mathcal{I}$, where each pixel $\mathbf{x} \in \mathbb{R}^d$ captures the spectral reflectance at d different wavelengths. We treat the image as a dataset of n pixels $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, where each $\mathbf{x}_i$ can be said to have a label $y_i \in \{-1, +1\}$. In most practical instances, the vast majority of these pixels have a label of $y_i = -1$ (no target); but it is often the case that the labels y_i are not available at all, and in that case the pixels are all treated as background and the label $y = -1$ is assigned to all the pixels. For each pixel $\mathbf{x}$, a new pixel $\xi(\mathbf{x})$ is created, and the label $y = +1$ is assigned to that pixel. In generating our matched pair data, we take advantage of the physics knowledge, encapsulated in $\xi(\mathbf{x})$, that we have with respect to how the target affects the pixel, whether it is additive, replacement, Beer's law, *etc.* Traditionally, the presence of a target has been modeled by the additive model, where it is assumed that if the target is present in the picture with some "strength" ϵ, the resultant spectrum will be given by $\mathbf{x} = \mathbf{b} + \epsilon \mathbf{t}$. Such models are not appropriate for opaque targets present in the pixels,[12] and thus in order to model the function $\xi(\mathbf{x})$, we will use the replacement
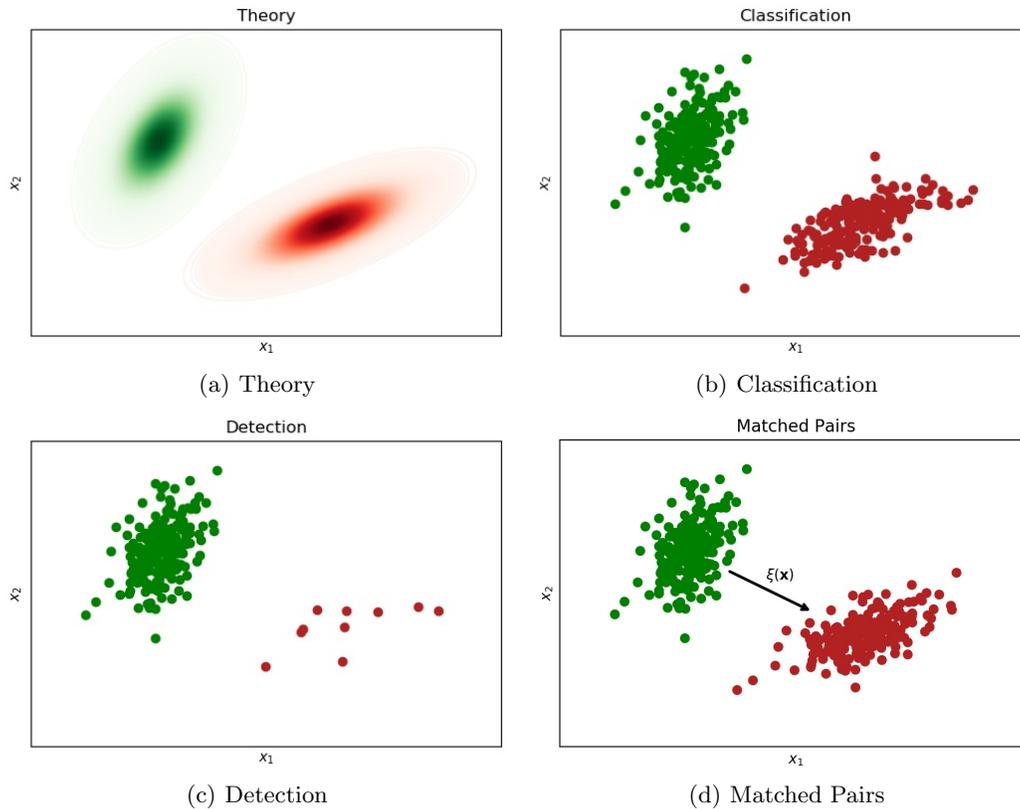
Figure 1. Following an original figure from Manolakis *et al.*,[24] which depicts an important distinction between classification and detection (namely, that the number of detectable samples are at most few, which challenges traditional machine learning), we show how using knowledge about the target/background interaction – as encapsulated in the function $\xi(\mathbf{x})$ – leads to matched pairs of background and target pixels, thereby enabling classification-style machine learning tools to be employed. Here, red is background, and green is target. (a) Theoretical models for the distributions. (b) Ideally, what the problem would look like from a classification standpoint. (c) What the problem actually looks like from a classification standpoint. (d) An approach for using matched-pair machine learning to generate more in-scene labeled target data.

model (*i.e.*, $\mathbf{x} = (1-\alpha)\mathbf{b} + \alpha\mathbf{t}$), where α is the fraction of the background spectrum $\mathbf{b}$ that will be replaced by the target spectrum $\mathbf{t}$ (here, α can also be thought of as the areal fraction of the pixel taken up by the target). This allows us to generate two sets of data:

$$\mathcal{X}_{-1} = \{(\mathbf{x}, -1) \mid \mathbf{x} \in \mathcal{X}\}, \text{ and} \tag{21}$$

$$\mathcal{X}_{+1} = \{((1-\alpha)\mathbf{x} + \alpha\mathbf{t}, +1) \mid \mathbf{x} \in \mathcal{X}\} \tag{22}$$

where $\mathbf{x}$ corresponds to individual pixels of the image, which are treated as background. Note that the background class $\mathcal{X}_{-1}$ includes any target-containing pixels, and the expectation is that by using robust classifiers we are going to be able to learn the correct decision surface that will separate the background and target pixels. This allows us to transform the detection problem (see Fig. 1(c)) to one of supervised classification (see Fig. 1(d)), and thus can draw on the plethora of tools developed to learn the decision function in a data-driven fashion.

2.2 Support Vector Classification

For our investigation, we chose the Support Vector Machine (SVM) as our classifier.[30] For completeness, in this section we give the basic formulation of the SVM and explore the different kernels we investigated. In its most basic formulation, the SVM tries to find the best linear classifier of the form

$$f(\mathbf{x}) = \mathbf{w}^{\mathsf{T}}\mathbf{x} + b \tag{23}$$

where $\mathbf{x_i}$ is assigned to class $y_i = 1$ if $f(\mathbf{x}) \geq 0$, else $y_i = -1$. Intuitively, SVM tries to maximize the margin, *i.e.*, the distance from the decision surface to the closest examples in either of the classes, and can be expressed:

$$\underset{\mathbf{w}}{\text{minimize}} \quad \mathbf{w}^\mathsf{T}\mathbf{w}$$

$$\text{subject to} \quad \left(\mathbf{w}^\mathsf{T}\mathbf{x}_j + b\right) y_j \geq 1, \ \forall j \in \{1, \ldots, m\}.$$

The constraint $\left(\mathbf{w}^\mathsf{T}\mathbf{x}_j + b\right) y_j \geq 1$ ensures that $f(\mathbf{x}_j)$ correctly assigns the label y_j for all j in the training set. A variant of of this approach that uses "slack variables" softens this constraint, and enables solutions that only assign *most* samples the correct label. The Represter Theorem states that the solution $\mathbf{w}$ can be expressed as:

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \implies f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i \left(\mathbf{x}_i^\mathsf{T}\mathbf{x}\right) + b \tag{24}$$

where we see that the decision function is a function of $\mathbf{x}_i^\mathsf{T}\mathbf{x}$, which can be thought of as a measure of similarity of the new example compared to previously seen examples. Using the "kernel trick," we can replace the dot product with any kernel function $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$ to measure the similarity, which implicitly maps the instances to higher dimensional spaces and performs the inner product in that space. In this case, the function $f(\mathbf{x})$ in Eq. (24) becomes

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i y_i \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) + b. \tag{25}$$

In our experiments, we explored the following kernel functions $\mathcal{K}$:

- **Linear Kernel** - given by a simple dot product between two vectors in $\mathbb{R}^d$:

$$\mathcal{K}_{\text{lin}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{n} \mathbf{x}_i^k \mathbf{x}_j^k = \mathbf{x}_i^\mathsf{T}\mathbf{x}_j \tag{26}$$

- **Cosine Similarity Kernel** - given by the cosine of an angle between two vectors in $\mathbb{R}^d$:

$$\mathcal{K}_{\text{cos}}(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i^\mathsf{T}\mathbf{x}_j}{\|\mathbf{x}_i\|\|\mathbf{x}_j\|} \tag{27}$$

- **Radial Basis Function (RBF) Kernel** - measures similarity of two vectors in $\mathbb{R}^d$:

$$\mathcal{K}_{\text{rbf}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \tag{28}$$

3. EXPERIMENTS

In our exploration of the EC-FTMF and the MPML detectors, we compared them on simulated data where we varied the data statistics, and on real data with available ground truth targets.[25] For each dataset, we also tested the detectors that are theoretically optimal for the various background models (*i.e.*, AMF, ACE, FTMF).

3.1 Synthetic Data with Implanted Targets

We chose to explore two variations of synthetic data, where the background data were drawn from the following distributions:

- Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, R)$, and

- Elliptically-contoured multivariate t-distribution $t_\nu(\boldsymbol{\mu}, R)$.

And for the target/background interaction, we used the replacement model:

$$\mathbf{x} = (1 - \alpha)\mathbf{b} + \alpha\mathbf{t}.$$

The mean $\boldsymbol{\mu}$ and covariance R were computed from the SHARE 2012 data and used to generate the synthetic samples (that are 229-dimensional) in order to better approximate the detector performance on the real data.

3.2 Real Data with Deployed Targets

In addition to the experiments with synthetic data, we used imagery from the RIT SHARE 2012 campaign[25] over Avon, NY. In particular, we used a hyperspectral image in which a collection of red and blue felt panels (2×2 m and 3×3 m in size) were placed in various states of illumination and occlusion.* The aerial image was captured by the SpecTIR VS sensor spanning 360 spectral bands in the VNIR-SWIR, covering the range 400-2400 nm. After removing the "bad bands," the final image contained a total of 229 spectral bands. The ground sample distance (GSD) of the sensor was approximately 1 m, and the georectified image used in this study was 170×280 pixels. The image was atmospherically compensated to approximate ground surface reflectance, and the target spectra were field-measured with an ASD spectrometer during the campaign, and are also in reflectance.
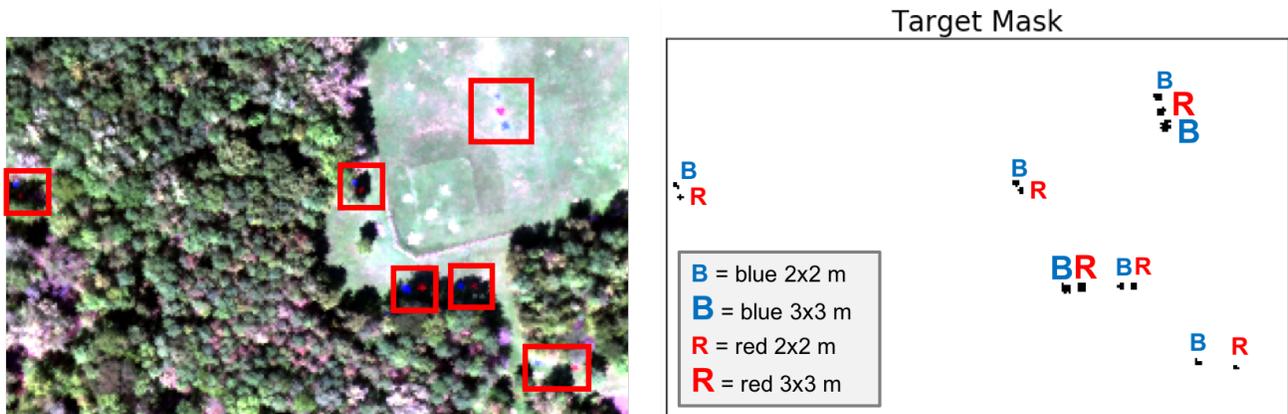
Figure 2. RGB of the SHARE 2012 hyperspectral image (left) and target mask showing the red and blue felt panel locations as well as the sizes of the deployed panels (right).

4. RESULTS

4.1 Synthetic Data

In Fig. 3(a), we show the Receiver Operating Characteristic (ROC) curves for the case when the background was Gaussian. The best performing detector is the FTMF, as was expected as it was derived to be the ideal detector under such conditions (*i.e.*, Gaussian background and replacement target model). We can see however that both the EC-FTMF detector and the RBF kernel SVM matched-pair approach (rbfMPML) closely trail the FTMF and do comparatively well. That is not the case for the other detectors, which achieve much poorer performance. In particular, the AMF detector – which also assumes a Gaussian background – performs very poorly; AMF also assumes an additive instead of a replacement model, and that incorrect assumption is very costly in this case. The linear kernel SVM matched-pair approach (linMPML) does slightly better than AMF, but still performs poorly, suggesting that a linear model will not work well in this scenario.

On the other hand, for the elliptically-contoured background, we see in Fig. 3(b) a substantial performance drop for the FTMF. Meanwhile, EC-FTMF is the best performing detector (which is not surprising, as it is optimized for this background). ACE and the nonlinear MPML detectors (cosMPML, rbfMPML) follow closely behind in performance.

While FTMF is the best detector when the background is Gaussian, it is evidently sensitive to deviations from this background model assumption. In contrast, EC-FTMF is relatively robust even as the background deviates from the specific EC distribution for which it was derived. This robustness is shared by the rbfMPML detector, which does not assume a particular background distribution, and instead adapts to the available data.

---

*The full SHARE 2012 data set is available at: `https://www.rit.edu/cos/share2012/`
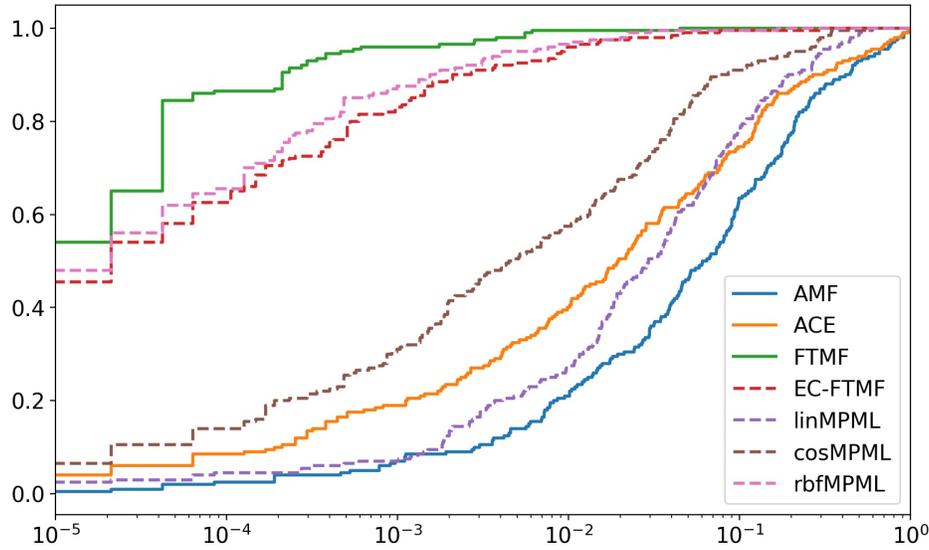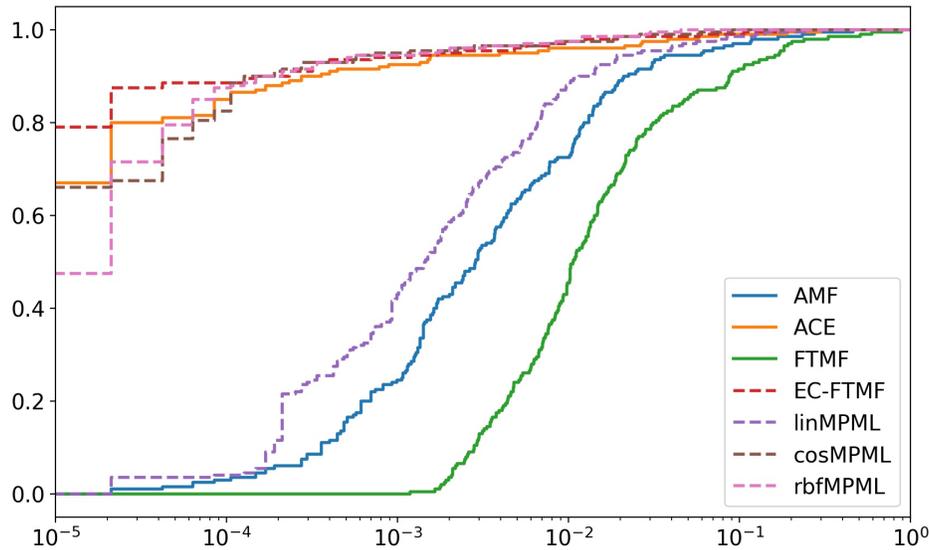
Figure 3. ROC curves for the various detectors with synthetic data generated using the replacement model, where the background was assumed to be (a) Gaussian and (b) elliptically-contoured ($\nu = 20$).

4.2 SHARE 2012 Data

In Fig. 4 we see ROC curves and detection maps for SHARE 2012 and the red felt targets, and Fig. 5 shows the ROC curves and detection maps for the blue felt targets. Because this is real hyperspectral data, the background is complicated (and non-Gaussian), and the opaque felt targets more accurately fit the replacement target model (as opposed to the additive model). For both panel types, the FTMF detector performs least well at the low false positive rates. We also observe here that, for the matched-pair approaches, cosMPML performs very well and rbfMPML performs poorly. ACE and EC-FTMF also perform well, and nearly identically. The performance of ACE in this case is especially impressive, since it was originally derived with the assumptions of an additive target model and a Gaussian background with unknown variance. (In fact, here the target is not additive, the background is not Gaussian, and the variance *is* known.)
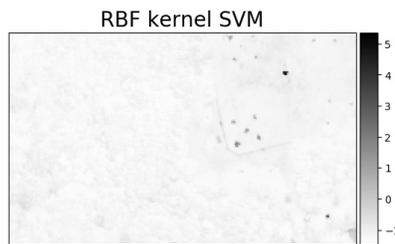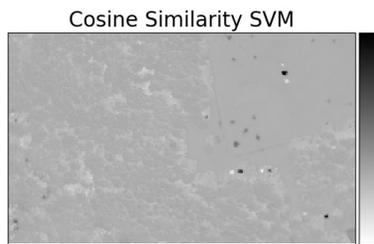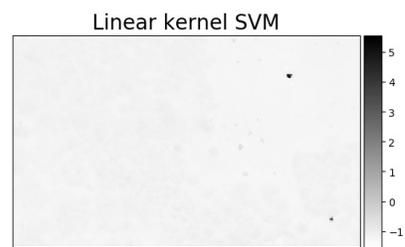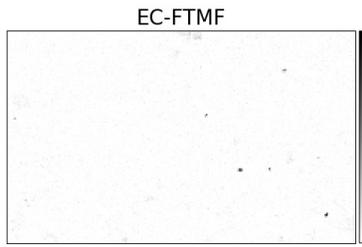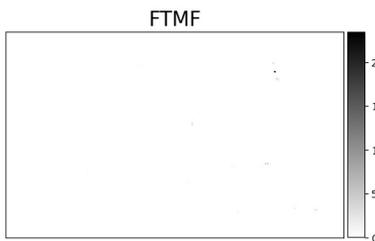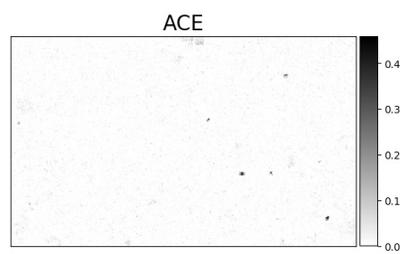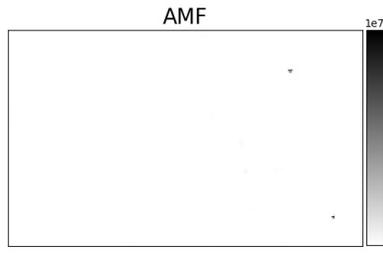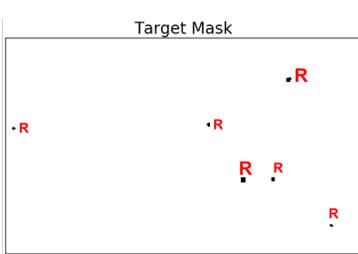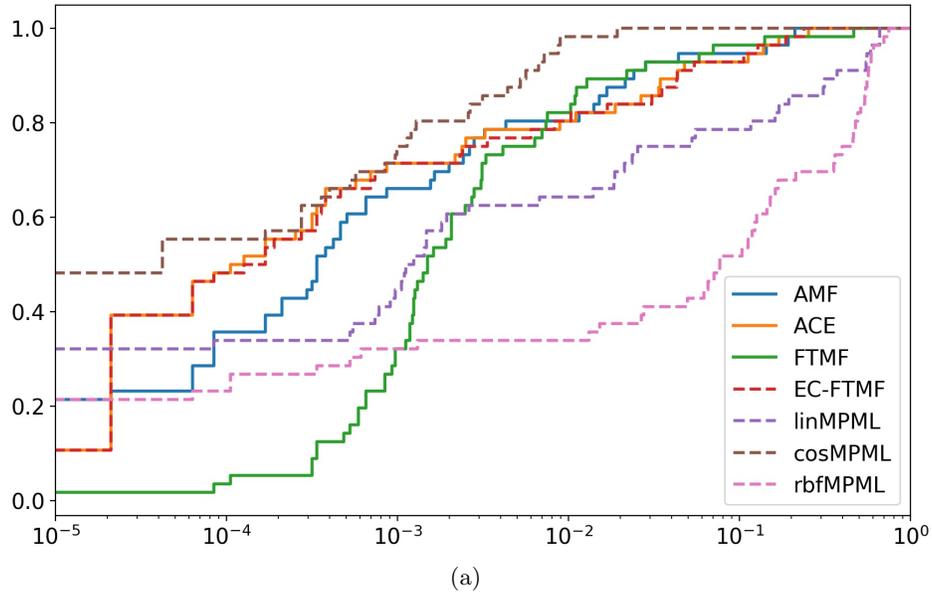
Figure 4. Results for the SHARE 2012 data with red felt targets: (a) ROC curves, (b) truth mask showing red felt panel locations, (c) - (i) corresponding detection maps for the various detectors.
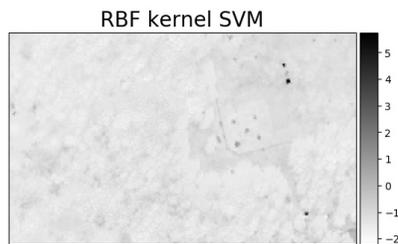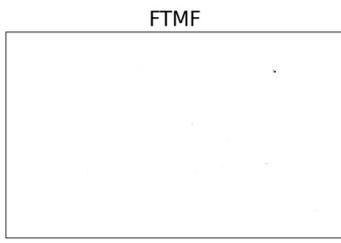
Figure 5. Results for the SHARE 2012 data with blue felt targets: (a) ROC curves, (b) truth mask showing blue felt panel locations, (c) - (i) corresponding detection maps for the various detectors.
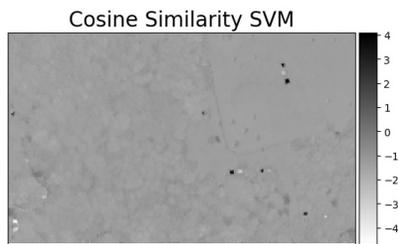
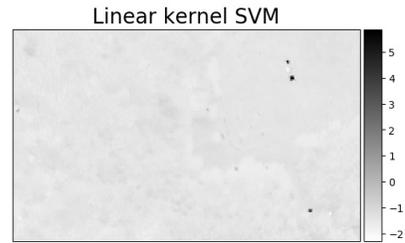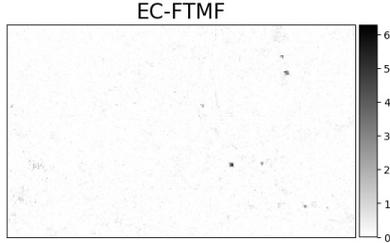## 5. CONCLUSIONS AND FUTURE WORK

This paper explored the use of machine learning to create classifier-based target detectors, and presented a derivation of the EC-FTMF detector. In the MPML f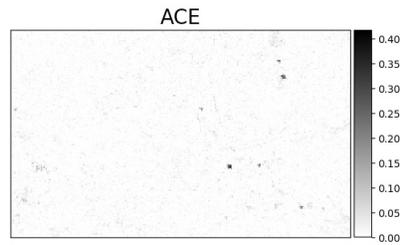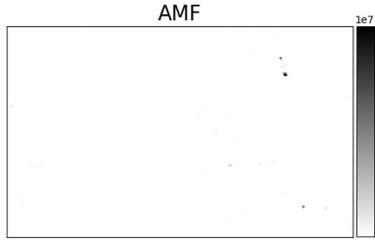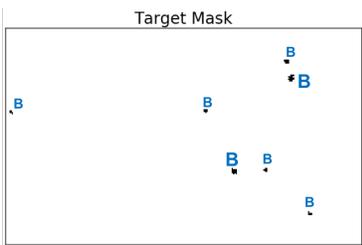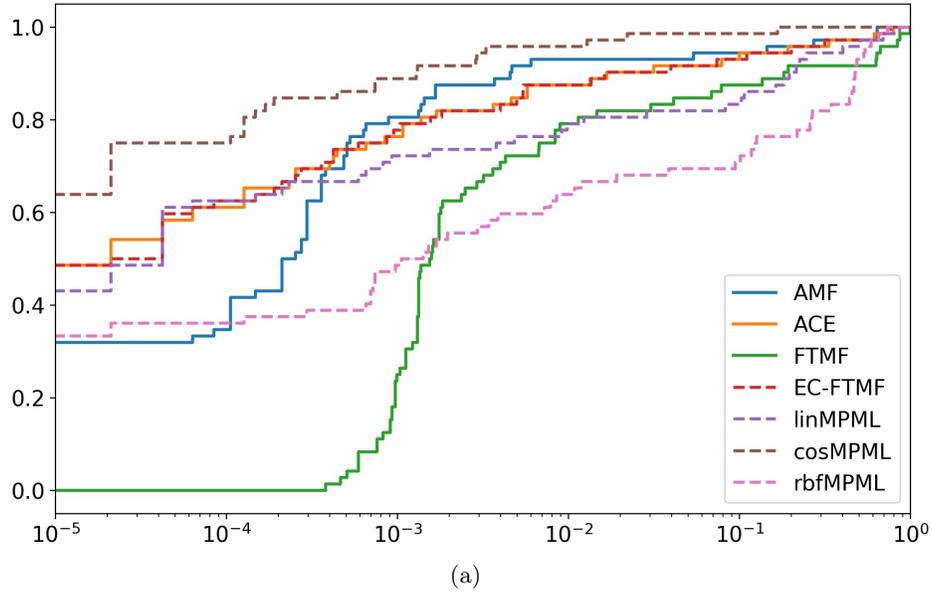ramework, we remedy the problem of a low number of target examples by creating *matched pairs*, where we apply a "treatment" to individual data samples in order to create pairs of samples (each pair comprising one untreated and one treated sample). We then use these pairs as labeled training data to enable the learning of a decision function $f(\mathbf{x})$ that maps all samples to either the background (untreated) or target (treated) class.

We also observed that relaxing the assumption of Gaussian background degraded the performance of the FTMF significantly. We proposed and derived a closed-form solution for EC-FTMF, a new detector which assumes an EC-tailed background. Results showed that EC-FTMF performs well in all of the tested cases, even in spite of the ground-truth test data not having an EC-tailed distribution. The performance of the Cosine Kernel SVM matched-pair detector suggests that the MPML target detection framework is promising, as it achieves the best results in both target cases for the ground-truthed data.

Most noteworthy is that this framework allows us to treat variability of the target in a more uniform way, so that a whole new algorithm (such as Simplex ACE[31–33] or Ampliskedastic FTMF[14, 18]) need not be developed for every new model of target variability, with the concern that the new algorithm is also making specific assumptions about the background distribution and target/background interaction. Instead, one replaces the deterministic $\xi(\mathbf{x})$ with a stochastic variant that samples from the target variability, and then just lets a classifier do the rest.

6. ACKNOWLEDGMENTS

REFERENCES

[1] Reed, I. S., Mallett, J. D., and Brennan, L. E., "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerospace and Electronic Systems* **10**, 853–863 (1974).

[2] Kelly, E. J., "Performance of an adaptive detection algorithm: rejection of unwanted signals," *IEEE Trans. Aerospace and Electronic Systems* **25**, 123–133 (1989).

[3] Robey, F. C., Fuhrmann, D. R., Kelly, E. J., and Nitzberg, R., "A CFAR adaptive matched filter detector," *IEEE Trans. Aerospace and Electronic Systems* **28**, 208–216 (1992).

[4] Scharf, L. L. and McWhorter, L. T., "Adaptive matched subspace detectors and adaptive coherence estimators," in [*Proc. Asilomar Conference on Signals, Systems, and Computers*], (1996).

[5] Kraut, S., Scharf, L. L., and Butler, R. W., "The adaptive coherence estimator: a uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Processing* **53**, 427–438 (2005).

[6] Theiler, J. and Foy, B. R., "EC-GLRT: Detecting weak plumes in non-Gaussian hyperspectral clutter using an elliptically-contoured generalized likelihood ratio test," in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], I:221 (2008).

[7] McWhorter, L., Scharf, L. L., and Griffiths, L., "Adaptive coherence estimation for radar signal processing," *Proc. Thirtieth Asilomar Conference on Signals, Systems and Computers* **1**, 536–540 (1996).

[8] Hayden, A., Niple, E., and Boyce, B., "Determination of trace-gas amounts in plumes by the use of orthogonal digital filtering of thermal-emission spectra," *Applied Optics* **35**, 2802–2809 (1996).

[9] Villeneuve, P. V., Fry, H. A., Theiler, J., Smith, B. W., and Stocker, A. D., "Improved matched-filter detection techniques," *Proc. SPIE* **3753**, 278–285 (1999).

[10] Theiler, J., Foy, B. R., and Fraser, A. M., "Characterizing non-Gaussian clutter and detecting weak gaseous plumes in hyperspectral imagery," *Proc. SPIE* **5806**, 182–193 (2005).

[11] Manolakis, D. G. and D'Amico, F. M., "A taxonomy of algorithms for chemical vapor detection with hyperspectral imaging spectroscopy," *Proc. SPIE* **5795**, 125–133 (2005).

[12] Schaum, A., "Enough with the additive target model," *Proc. SPIE* **9088**, 90880C (2014).

[13] Boardman, J. W. and Kruse, F. A., "Analysis of imaging spectrometer data using N-dimensional geometry and a mixture-tuned matched filtering approach," *IEEE Trans. Geoscience and Remote Sensing* **49**, 4138–4152 (2011).

[14] DiPietro, R. S., Manolakis, D. G., Lockwood, R. B., Cooley, T., and Jacobson, J., "Performance evaluation of hyperspectral detection algorithms for sub-pixel objects," *Proc. SPIE* **7695**, 76951W (2010).

[15] DiPietro, R. S., Manolakis, D. G., Lockwood, R. B., Cooley, T., and Jacobson, J., "Hyperspectral matched filter with false-alarm mitigation," *Optical Engineering* **51**, 016202 (2012).

[16] Theiler, J. and Ziemann, A., "Local background estimation and the replacement target model," *Proc. SPIE* **10198**, 101980V (2017).

[17] Schaum, A. and Stocker, A., "Spectrally selective target detection," in [*Proc. ISSSR (International Symposium on Spectral Sensing Research)*], 23 (1997).

[18] Schaum, A., "Continuum fusion versions of the finite target matched filter for sub-pixel detection," in [*Proc. Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*], (2013).

[19] Schaum, A., "Continuum fusion solutions for replacement target models in electro-optic detection," *Applied Optics* **53**, C25–C31 (2014).

[20] Theiler, J., Zimmer, B., and Ziemann, A., "Closed-form detector for solid sub-pixel targets in multivariate t-distributed background clutter," *arXiv:1804.02062* (2018).

[21] Manolakis, D., Marden, D., Kerekes, J., and Shaw, G., "On the statistics of hyperspectral imaging data," *Proc. SPIE* **4381**, 308–316 (2001).

[22] Marden, D. B. and Manolakis, D., "Using elliptically contoured distributions to model hyperspectral imaging data and generate statistically similar synthetic data," *Proc. SPIE* **5425**, 558–572 (2004).

[23] Theiler, J., Scovel, C., Wohlberg, B., and Foy, B. R., "Elliptically-contoured distributions for anomalous change detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters* **7**, 271–275 (2010).

[24] Manolakis, D., Marden, D., and A. Shaw, G., "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal* **14**, 79–116 (2003).

[25] Giannandrea, A., Raqueno, N., Messinger, D. W., Faulring, J., Kerekes, J. P., van Aardt, J., Canham, K., Hagstrom, S., Ontiveros, E., Gerace, A., Kaufman, J., Vongsy, K. M., Griffith, H., Bartlett, B. D., Ientilucci, E., Meola, J., Scarff, L., and Daniel, B., "The SHARE 2012 data campaign," *Proc. SPIE* **8743**, 87430F (2013).

[26] Theiler, J., "Matched-pair machine learning," *Technometrics* **55**(4), 536–547 (2013).

[27] Theiler, J., "Transductive and matched-pair machine learning for difficult target detection problems," *Proc. SPIE* **9088**, 90880E (2014).

[28] Stefanou, M. S. and Kerekes, J. P., "A method for assessing spectral image utility," *IEEE Trans. Geoscience and Remote Sensing* **47**, 1698–1706 (2009).

[29] Basener, W. F., Nance, E., and Kerekes, J., "The target implant method for predicting target difficulty and detector performance in hyperspectral imagery," *Proc. SPIE* **8048**, 80481H (2011).

[30] Cortes, C. and Vapnik, V., "Support vector networks," *Machine Learning* **20**, 273 (1995).

[31] Broadwater, J. and Chellappa, R., "Hybrid detectors for subpixel targets," *IEEE Trans. Pattern Analysis and Machine Intelligence* **29**, 1891–1903 (2007).

[32] Adler-Golden, S., Gruninger, J., and Sundberg, R., "Hyperspectral detection and identification with constrained target subspaces," in [*Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*], **II**, 465–468 (2008).

[33] Ziemann, A. and Theiler, J., "Simplex ACE: a constrained subspace detector," *Optical Engineering* **56**, 081808 (2017).

7. APPENDIX: DERIVATION OF COEFFICIENTS FOR EC-FTMF

Our aim in this Appendix is to derive the expressions for the coefficients A, B, and C given in the text by Eqs. (13-15). We begin with Eq. (11), which we recapitulate here:

$$p(\mathbf{x}|\alpha) = c \times (1 - \alpha)^{-d} \left[1 + \frac{(1 - \alpha)^{-2}}{\nu - 2}(\mathbf{z} - \alpha\mathbf{u})^{\mathsf{T}}(\mathbf{z} - \alpha\mathbf{u})\right]^{-(\nu+d)/2}$$

where c is a constant, while $\mathbf{z} = R^{-1/2}(\mathbf{x}-\boldsymbol{\mu})$ and $\mathbf{u} = R^{-1/2}(\mathbf{t}-\boldsymbol{\mu})$ are the whitened versions of the measurement pixel and the target. To reduce the algebraic clutter, let $\beta = 1-\alpha$ and $\mathbf{w} = \mathbf{z}-\mathbf{u}$ (Note: $\mathbf{z}-\alpha\mathbf{u} = \mathbf{z}-\mathbf{u}+\mathbf{u}-\alpha\mathbf{u} = \mathbf{w} + \beta\mathbf{u}$). We want $\hat{\beta} = \mathrm{argmax}_\beta\, p(\mathbf{x}|\beta)$, but since the logarithm is monotonic, it is equivalant to write

$$\hat{\beta} = \mathrm{argmax}_\beta\, \log p(\mathbf{x}|\beta) \tag{29}$$

where

$$\log p(\mathbf{x}|\beta) = \log(c) - d\log(\beta) - \left(\frac{\nu+d}{2}\right)\log\left[1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})\right]. \tag{30}$$

To find $\hat{\beta}$, we will take the derivative of $\log p(\mathbf{x}|\beta)$, and set it to zero.

$$0 = \frac{\partial}{\partial\beta}\log p(\mathbf{x}|\beta)$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\partial}{\partial\beta}\log\left[1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})\right]$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{\partial}{\partial\beta}\left[1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{1}{\nu - 2}\frac{\partial}{\partial\beta}\left[\beta^{-2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{1}{\nu - 2}\left[-2\beta^{-3}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u}) + \beta^{-2}\frac{\partial}{\partial\beta}\left[(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})\right]\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{1}{\nu - 2}\left[-2\beta^{-3}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u}) + 2\beta^{-2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}\mathbf{u}\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{1}{\nu - 2}\left[-2\beta^{-3}(\mathbf{w}^{\mathsf{T}}\mathbf{w} + 2\beta\mathbf{w}^{\mathsf{T}}\mathbf{u} + \beta^{2}\mathbf{u}^{\mathsf{T}}\mathbf{u}) + 2\beta^{-2}(\mathbf{w}^{\mathsf{T}}\mathbf{u} + \beta\mathbf{u}^{\mathsf{T}}\mathbf{u})\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} - \left(\frac{\nu+d}{2}\right)\frac{\frac{1}{\nu - 2}\left[-2\beta^{-3}\mathbf{w}^{\mathsf{T}}\mathbf{w} - 2\beta^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{u}\right]}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

$$= -d\beta^{-1} + \left(\frac{\nu+d}{\nu - 2}\right)\frac{\beta^{-3}\mathbf{w}^{\mathsf{T}}\mathbf{w} + \beta^{-2}\mathbf{w}^{\mathsf{T}}\mathbf{u}}{1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^{\mathsf{T}}(\mathbf{w} + \beta\mathbf{u})}$$

Multiplying the equations through with the denominator of the second part gives us the following equation:

$$0 = -d\beta^{-1}\left[1 + \frac{\beta^{-2}}{\nu - 2}(\mathbf{w} + \beta\mathbf{u})^\mathsf{T}(\mathbf{w} + \beta\mathbf{u})\right] + \left(\frac{\nu + d}{\nu - 2}\right)\left[\beta^{-3}\mathbf{w}^\mathsf{T}\mathbf{w} + \beta^{-2}\mathbf{w}^\mathsf{T}\mathbf{u}\right] \tag{31}$$

Multiplying both sides with $-\beta^3(\nu - 2)/d$ gives

$$\begin{aligned}
0 &= (\nu - 2)\beta^2 + (\mathbf{w} + \beta\mathbf{u})^\mathsf{T}(\mathbf{w} + \beta\mathbf{u}) - (\nu/d + 1)\left[\mathbf{w}^\mathsf{T}\mathbf{w} + \beta\mathbf{w}^\mathsf{T}\mathbf{u}\right] \\
&= \left[(\nu - 2) + \mathbf{u}^\mathsf{T}\mathbf{u}\right]\beta^2 + \left[2\mathbf{w}^\mathsf{T}\mathbf{u} - (\nu/d + 1)\mathbf{w}^\mathsf{T}\mathbf{u}\right]\beta + \left[\mathbf{w}^\mathsf{T}\mathbf{w} - (\nu/d + 1)\mathbf{w}^\mathsf{T}\mathbf{w}\right] \\
&= \left[\mathbf{u}^\mathsf{T}\mathbf{u} + (\nu - 2)\right]\beta^2 + \left[(1 - \nu/d)\mathbf{w}^\mathsf{T}\mathbf{u}\right]\beta + \left[-(\nu/d)\mathbf{w}^\mathsf{T}\mathbf{w}\right].
\end{aligned}$$

Recall $\beta = 1 - \alpha$, so we now have the quadratic equation $0 = A(1 - \alpha)^2 + B(1 - \alpha) + C$ given in Eq. (12), with

$$\begin{aligned}
A &= \mathbf{u}^\mathsf{T}\mathbf{u} + (\nu - 2) \\
B &= (1 - \nu/d)\mathbf{w}^\mathsf{T}\mathbf{u} \\
C &= (-\nu/d)\mathbf{w}^\mathsf{T}\mathbf{w}
\end{aligned}$$

Going back to our original unwhitened variables, we have

$$\begin{aligned}
\mathbf{u} &= R^{-1/2}(\mathbf{t} - \boldsymbol{\mu}) \\
\mathbf{w} &= \mathbf{z} - \mathbf{u} \\
&= R^{-1/2}(\mathbf{x} - \boldsymbol{\mu}) - R^{-1/2}(\mathbf{t} - \boldsymbol{\mu}) \\
&= R^{-1/2}(\mathbf{x} - \mathbf{t})
\end{aligned}$$

which leads to

$$\begin{aligned}
A &= (\mathbf{t} - \boldsymbol{\mu})^\mathsf{T} R^{-1}(\mathbf{t} - \boldsymbol{\mu}) + (\nu - 2) \\
B &= (1 - \nu/2)(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{t} - \boldsymbol{\mu}) \\
C &= (-\nu/d)(\mathbf{x} - \mathbf{t})^\mathsf{T} R^{-1}(\mathbf{x} - \mathbf{t})
\end{aligned}$$

as reported in Eqs. (13,14,15).