

Research Article

A Comparative Analysis of Kernel Subspace Target Detectors for Hyperspectral Imagery

Heesung Kwon and Nasser M. Nasrabadi

US Army Research Laboratory, ATTN: AMSRL-SE-SE, 2800 Powder Mill Road, Adelphi, MD 20783-1197, USA

Received 30 September 2005; Revised 11 May 2006; Accepted 18 May 2006

Recommended by Kostas Berberidis

Several linear and nonlinear detection algorithms that are based on spectral matched (subspace) filters are compared. Nonlinear (*kernel*) versions of these spectral matched detectors are also given and their performance is compared with linear versions. Several well-known matched detectors such as matched subspace detector, orthogonal subspace detector, spectral matched filter, and adaptive subspace detector are extended to their corresponding kernel versions by using the idea of kernel-based learning theory. In kernel-based detection algorithms the data is assumed to be implicitly mapped into a high-dimensional kernel feature space by a nonlinear mapping, which is associated with a kernel function. The expression for each detection algorithm is then derived in the feature space, which is *kernelized* in terms of the kernel functions in order to avoid explicit computation in the high-dimensional feature space. Experimental results based on simulated toy examples and real hyperspectral imagery show that the kernel versions of these detectors outperform the conventional linear detectors.

Copyright © 2007 H. Kwon and N. M. Nasrabadi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Detecting signals of interest, particularly with wide signal variability, in noisy environments has long been a challenging issue in various fields of signal processing. Among a number of previously developed detectors, the well-known matched subspace detector (MSD) [1], orthogonal subspace detector (OSD) [1, 2], spectral matched filter (SMF) [3, 4], and adaptive subspace detectors (ASD) also known as adaptive cosine estimator (ACE) [5, 6], have been widely used to detect a desired signal (target).

Matched signal detectors, such as spectral matched filter and matched subspace detectors (whether adaptive or non-adaptive), only exploit second-order correlations, thus completely ignoring nonlinear (higher-order) spectral interband correlations that could be crucial to discriminate between targets and background. In this paper, our goal is to provide a complete comparative analysis of the kernel-based versions of MSD, OSD, SMF, and ASD detectors [7–10] which have equivalent nonlinear versions in the input domain. Each kernel detector is obtained by defining a corresponding model in a high- (possibly infinite) dimensional feature space associated with a certain nonlinear mapping of the input data. This

nonlinear mapping of the input data into a high-dimensional feature space is often expected to increase the data separability and provide simpler decision rules for data discrimination [11]. These kernel-based detectors exploit the higher-order spectral interband correlations in a feature space which is implicitly achieved via a kernel function implementation [12].

The nonlinear versions of a number of signal processing techniques such as principal component analysis (PCA) [13], Fisher discriminant analysis [14], clustering in feature space [15], linear classifiers [16], nonlinear feature extraction based on kernel orthogonal centroid method [17], matched signal detectors for target detection [7–10], anomaly detection [18], classification in nonlinear subspace [19], and classifiers based on kernel Bayes rule [20] have already been defined in kernel space. Furthermore, in [21] kernels were used as generalized dissimilarity measures for classification and in [22] kernel methods were applied to face recognition.

This paper is organized as follows. Section 2 provides the background to the kernel-based learning methods and kernel trick. Section 3 introduces the linear matched subspace detector and its kernel version. The orthogonal subspace detector is defined in Section 4 as well as its kernel version. In Section 5 we describe the conventional spectral

matched filter and its kernel version in the feature space in terms of the kernel function using the kernel trick. Finally, in Section 6 the adaptive subspace detector and its kernel version are introduced. Performance comparison between the conventional and the kernel version of these algorithms is provided in Section 7. Conclusions are given in Section 8.

2. KERNEL METHODS AND KERNEL TRICK

The basic principle behind kernel-based algorithms is that a nonlinear mapping is used to extend the input space to a higher-dimensional feature space. Implementing a simple algorithm in the feature space then corresponds to a nonlinear version of the algorithm in the original input space. The algorithm is efficiently implemented in the feature space by using a Mercer kernel function [11] which uses the so-called kernel trick property [12]. Suppose that the input hyperspectral data is represented by the data space ($\mathcal{X} \subseteq \mathcal{R}^p$) and \mathcal{F} is a feature space associated with \mathcal{X} by a nonlinear mapping function ϕ

$$\phi : \mathcal{X} \rightarrow \mathcal{F}, \quad \mathbf{x} \mapsto \phi(\mathbf{x}), \quad (1)$$

where \mathbf{x} is an input vector in \mathcal{X} which is mapped into a potentially much higher (could be infinite) dimensional feature space. Due to the high dimensionality of the feature space \mathcal{F} , it is computationally not feasible to implement any algorithm directly in feature space. However, kernel-based learning algorithms use an effective kernel trick given by (2) to implement dot products in feature space by employing kernel functions [12]. The idea in kernel-based techniques is to obtain a nonlinear version of an algorithm defined in the input space by implicitly redefining it in the feature space and then converting it in terms of dot products. The kernel trick is then used to implicitly compute the dot products in \mathcal{F} without mapping the input vectors into \mathcal{F} ; therefore, in the kernel methods, the mapping ϕ does not need to be identified.

The kernel representation for the dot products in \mathcal{F} is expressed as

$$k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j), \quad (2)$$

where k is a kernel function in terms of the original data. There are a large number of Mercer kernels that have the kernel trick property; see [12] for detailed information about the properties of different kernels and kernel-based learning. Our choice of kernel in this paper is the Gaussian radial basis function (RBF) kernel and the associated nonlinear function ϕ with this kernel generates a feature space of infinite dimensionality.

3. LINEAR MSD AND KERNEL MSD

3.1. Linear MSD

In this model the target pixel vectors are expressed as a linear combination of target spectral signature and background spectral signature, which are represented by subspace target spectra and subspace background spectra, respectively. The

hyperspectral target detection problem in a p -dimensional input space is expressed as two competing hypotheses \mathbf{H}_0 and \mathbf{H}_1 :

$$\begin{aligned} \mathbf{H}_0 : \mathbf{y} &= \mathbf{B}\boldsymbol{\zeta} + \mathbf{n}, \quad \text{target absent,} \\ \mathbf{H}_1 : \mathbf{y} &= \mathbf{T}\boldsymbol{\theta} + \mathbf{B}\boldsymbol{\zeta} + \mathbf{n} = \begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\zeta} \end{bmatrix} + \mathbf{n}, \quad \text{target present,} \end{aligned} \quad (3)$$

where \mathbf{T} and \mathbf{B} represent orthogonal matrices whose p -dimensional orthonormal columns span the target and background subspaces, respectively; $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ are unknown vectors whose entries are coefficients that account for the abundances of the corresponding column vectors of \mathbf{T} and \mathbf{B} , respectively; \mathbf{n} represents Gaussian random noise ($\mathbf{n} \in \mathcal{R}^p$) distributed as $\mathcal{N}(0, \sigma^2 \mathbf{I})$; and $\begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix}$ is a concatenated matrix of \mathbf{T} and \mathbf{B} . The numbers of the column vectors of \mathbf{T} and \mathbf{B} , N_t , and N_b , respectively, are usually smaller than p ($N_t, N_b < p$).

The generalized likelihood ratio test (GLRT) for model (3) was derived in [1], given as

$$\mathbf{L}_2(\mathbf{y}) = \frac{\mathbf{y}^T (\mathbf{I} - \mathbf{P}_B) \mathbf{y}}{\mathbf{y}^T (\mathbf{I} - \mathbf{P}_{TB}) \mathbf{y}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (4)$$

where $\mathbf{P}_B = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T = \mathbf{B} \mathbf{B}^T$ is a projection matrix associated with the N_b -dimensional background subspace (\mathbf{B}); \mathbf{P}_{TB} is a projection matrix associated with the ($N_{bt} = N_b + N_t$)-dimensional target-and-background subspace (\mathbf{TB}):

$$\mathbf{P}_{TB} = \begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix} \left[\begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix}^T \begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix} \right]^{-1} \begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix}^T. \quad (5)$$

$\mathbf{L}_2(\mathbf{y})$ is compared to η to make a final decision about which hypothesis best relates to \mathbf{y} . In general, any sets of orthonormal basis vectors that span the corresponding subspace can be used as the column vectors of \mathbf{T} and \mathbf{B} . In this paper, the significant eigenvectors (normalized by the square root of their corresponding eigenvalues) of the target and background covariance matrices \mathbf{C}_T and \mathbf{C}_B are used to create the column vectors of \mathbf{T} and \mathbf{B} , respectively.

3.2. Linear MSD in the feature space and its kernel version

The hyperspectral detection problem based on the target and background subspaces can be described in the feature space \mathcal{F} as

$$\begin{aligned} \mathbf{H}_{0_\phi} : \phi(\mathbf{y}) &= \mathbf{B}_\phi \boldsymbol{\zeta}_\phi + \mathbf{n}_\phi, \quad \text{target absent,} \\ \mathbf{H}_{1_\phi} : \phi(\mathbf{y}) &= \mathbf{T}_\phi \boldsymbol{\theta}_\phi + \mathbf{B}_\phi \boldsymbol{\zeta}_\phi + \mathbf{n}_\phi \\ &= \begin{bmatrix} \mathbf{T}_\phi & \mathbf{B}_\phi \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_\phi \\ \boldsymbol{\zeta}_\phi \end{bmatrix} + \mathbf{n}_\phi, \quad \text{target present,} \end{aligned} \quad (6)$$

where \mathbf{T}_ϕ and \mathbf{B}_ϕ represent matrices whose orthonormal columns span target and background subspaces (\mathbf{B}_ϕ) and (\mathbf{T}_ϕ) in \mathcal{F} , respectively; $\boldsymbol{\theta}_\phi$ and $\boldsymbol{\zeta}_\phi$ are unknown vectors

whose entries are coefficients that account for the abundances of the corresponding column vectors of \mathbf{T}_ϕ and \mathbf{B}_ϕ , respectively; \mathbf{n}_ϕ represents Gaussian random noise; and $[\mathbf{T}_\phi \ \mathbf{B}_\phi]$ is a concatenated matrix of \mathbf{T}_ϕ and \mathbf{B}_ϕ . The significant eigenvectors (normalized) of the target and background covariance matrices ($\mathbf{C}_{\mathbf{T}_\phi}$ and $\mathbf{C}_{\mathbf{B}_\phi}$) in \mathcal{F} form the column vectors of \mathbf{T}_ϕ and \mathbf{B}_ϕ , respectively. It should be pointed out that the above model (6) in the feature space is not exactly the same as applying the nonlinear map ϕ to the additive model given in (3). However, this model in the feature space is equivalent to a specific nonlinear model in the input space which is capable of modeling the nonlinear interband relationships within the data. Therefore, defining MSD using the model (6) is the same as developing an MSD for an equivalent nonlinear model in the input space.

Using a similar reasoning as described in the previous subsection, the GLRT of the hyperspectral detection problem depicted by the model in (6) as shown in [7] is given by

$$\mathbf{L}_2(\phi(\mathbf{y})) = \frac{\phi(\mathbf{y})^T (\mathbf{P}_{\mathbf{I}_\phi} - \mathbf{P}_{\mathbf{B}_\phi}) \phi(\mathbf{y})}{\phi(\mathbf{y})^T (\mathbf{P}_{\mathbf{I}_\phi} - \mathbf{P}_{\mathbf{T}_\phi \mathbf{B}_\phi}) \phi(\mathbf{y})} \stackrel{H_{1\phi}}{\geq} \eta_\phi, \quad (7)$$

where $\mathbf{P}_{\mathbf{I}_\phi}$ represents an identity projection operator in \mathcal{F} ; $\mathbf{P}_{\mathbf{B}_\phi} = \mathbf{B}_\phi (\mathbf{B}_\phi^T \mathbf{B}_\phi)^{-1} \mathbf{B}_\phi^T = \mathbf{B}_\phi \mathbf{B}_\phi^T$ is a background projection matrix; and $\mathbf{P}_{\mathbf{T}_\phi \mathbf{B}_\phi}$ is a joint target-and-background projection matrix in \mathcal{F} :

$$\begin{aligned} \mathbf{P}_{\mathbf{T}_\phi \mathbf{B}_\phi} &= [\mathbf{T}_\phi \ \mathbf{B}_\phi] \left[[\mathbf{T}_\phi \ \mathbf{B}_\phi]^T \ [\mathbf{T}_\phi \ \mathbf{B}_\phi] \right]^{-1} [\mathbf{T}_\phi \ \mathbf{B}_\phi]^T \\ &= [\mathbf{T}_\phi \ \mathbf{B}_\phi] \left[\begin{array}{cc} \mathbf{T}_\phi^T \mathbf{T}_\phi & \mathbf{T}_\phi^T \mathbf{B}_\phi \\ \mathbf{B}_\phi^T \mathbf{T}_\phi & \mathbf{B}_\phi^T \mathbf{B}_\phi \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{T}_\phi^T \\ \mathbf{B}_\phi^T \end{array} \right]. \end{aligned} \quad (8)$$

To kernelize (7) we will separately kernelize the numerator and denominator. First consider its numerator:

$$\phi(\mathbf{y})^T (\mathbf{P}_{\mathbf{I}_\phi} - \mathbf{P}_{\mathbf{B}_\phi}) \phi(\mathbf{y}) = \phi(\mathbf{y})^T \mathbf{P}_{\mathbf{I}_\phi} \phi(\mathbf{y}) - \phi(\mathbf{y})^T \mathbf{B}_\phi \mathbf{B}_\phi^T \phi(\mathbf{y}). \quad (9)$$

Using (A.3), as shown in the appendix, \mathbf{B}_ϕ and \mathbf{T}_ϕ can be written in terms of their corresponding data spaces as

$$\mathbf{B}_\phi = [\mathbf{e}_b^1 \ \mathbf{e}_b^2 \ \cdots \ \mathbf{e}_b^{N_b}] = \phi_{\mathbf{Z}_B} \widetilde{\mathcal{B}}, \quad (10)$$

$$\mathbf{T}_\phi = [\mathbf{e}_t^1 \ \mathbf{e}_t^2 \ \cdots \ \mathbf{e}_t^{N_t}] = \phi_{\mathbf{Z}_T} \widetilde{\mathcal{T}}, \quad (11)$$

where \mathbf{e}_b^i and \mathbf{e}_t^i are the significant eigenvectors of $\mathbf{C}_{\mathbf{B}_\phi}$ and $\mathbf{C}_{\mathbf{T}_\phi}$, respectively; $\phi_{\mathbf{Z}_B} = [\phi(\mathbf{y}_1) \ \phi(\mathbf{y}_2) \ \cdots \ \phi(\mathbf{y}_M)]$, $\mathbf{y}_i \in \mathbf{Z}_B$ is the background reference data and $\phi_{\mathbf{Z}_T} = [\phi(\mathbf{y}_1) \ \phi(\mathbf{y}_2) \ \cdots \ \phi(\mathbf{y}_N)]$, $\mathbf{y}_i \in \mathbf{Z}_T$ is the target reference data; and the column vectors of $\widetilde{\mathcal{B}}$ and $\widetilde{\mathcal{T}}$ represent only the significant eigenvectors ($\widetilde{\beta}_1, \widetilde{\beta}_2, \dots, \widetilde{\beta}_{N_b}$) and ($\widetilde{\alpha}_1, \widetilde{\alpha}_2, \dots, \widetilde{\alpha}_{N_t}$) of the background centered kernel matrix $\mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_B) = (\mathbf{K})_{ij} = k(\mathbf{y}_i, \mathbf{y}_j)$, $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Z}_B$, and target centered kernel matrix $\mathbf{K}(\mathbf{Z}_T, \mathbf{Z}_T) = (\mathbf{K})_{ij} = k(\mathbf{y}_i, \mathbf{y}_j)$, $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Z}_T$, normalized by the square root of their associated eigenvalues, respectively. Using (10) the projection of $\phi(\mathbf{y})$ onto \mathbf{B}_ϕ

becomes $\mathbf{B}_\phi^T \phi(\mathbf{y}) = \widetilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y})$ and, similarly, using (11) the projection onto \mathbf{T}_ϕ is $\mathbf{T}_\phi^T \phi(\mathbf{y}) = \widetilde{\mathcal{T}}^T \mathbf{k}(\mathbf{Z}_T, \mathbf{y})$, where $\mathbf{k}(\mathbf{Z}_B, \mathbf{y})$ and $\mathbf{k}(\mathbf{Z}_T, \mathbf{y})$, referred to as the empirical kernel maps in the machine learning literature [12], are column vectors whose entries are $k(\mathbf{x}_i, \mathbf{y})$ for $\mathbf{x}_i \in \mathbf{Z}_B$ and $\mathbf{x}_i \in \mathbf{Z}_T$, respectively. Now we can write

$$\phi(\mathbf{y})^T \widetilde{\mathcal{B}} \widetilde{\mathcal{B}}^T \phi(\mathbf{y}) = \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T \widetilde{\mathcal{B}} \widetilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}). \quad (12)$$

The projection onto the identity operator $\phi(\mathbf{y})^T \mathbf{P}_{\mathbf{I}_\phi} \phi(\mathbf{y})$ also needs to be kernelized. $\mathbf{P}_{\mathbf{I}_\phi}$ is defined as $\mathbf{P}_{\mathbf{I}_\phi} := \mathbf{\Omega}_\phi \mathbf{\Omega}_\phi^T$, where $\mathbf{\Omega}_\phi = [\mathbf{e}_q^1 \ \mathbf{e}_q^2 \ \cdots]$ is a matrix whose columns are all the eigenvectors with $\lambda \neq 0$ that are in the span of $\phi(\mathbf{y}_i)$, $\mathbf{y}_i \in \mathbf{Z}_T \cup \mathbf{Z}_B := \mathbf{Z}_{\mathbf{T}\mathbf{B}}$. From (A.3) $\mathbf{\Omega}_\phi$ can similarly be expressed as

$$\mathbf{\Omega}_\phi = [\mathbf{e}_q^1 \ \mathbf{e}_q^2 \ \cdots \ \mathbf{e}_q^{N_{bt}}] = \phi_{\mathbf{Z}_{\mathbf{T}\mathbf{B}}} \widetilde{\Delta}, \quad (13)$$

where $\phi_{\mathbf{Z}_{\mathbf{T}\mathbf{B}}} = \phi_{\mathbf{Z}_T} \cup \phi_{\mathbf{Z}_B}$ and $\widetilde{\Delta}$ is a matrix whose columns are the eigenvectors ($\kappa_1, \kappa_2, \dots, \kappa_{N_{bt}}$) of the centered kernel matrix $\mathbf{K}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{Z}_{\mathbf{T}\mathbf{B}}) = (\mathbf{K})_{ij} = k(\mathbf{y}_i, \mathbf{y}_j)$, $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Z}_{\mathbf{T}\mathbf{B}}$, with nonzero eigenvalues, normalized by the square root of their associated eigenvalues. Using $\mathbf{P}_{\mathbf{I}_\phi} = \mathbf{\Omega}_\phi \mathbf{\Omega}_\phi^T$ and (13),

$$\begin{aligned} \phi(\mathbf{y})^T \mathbf{P}_{\mathbf{I}_\phi} \phi(\mathbf{y}) &= \phi(\mathbf{y})^T \phi_{\mathbf{Z}_{\mathbf{T}\mathbf{B}}} \widetilde{\Delta} \widetilde{\Delta}^T \phi_{\mathbf{Z}_{\mathbf{T}\mathbf{B}}}^T \phi(\mathbf{y}) \\ &= \mathbf{k}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{y})^T \widetilde{\Delta} \widetilde{\Delta}^T \mathbf{k}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{y}); \end{aligned} \quad (14)$$

$\mathbf{k}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{y})$ is the concatenated vector $[\mathbf{k}(\mathbf{Z}_T, \mathbf{y})^T \ \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T]^T$. The kernelized numerator of (7) is now given by

$$\mathbf{k}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{y})^T \widetilde{\Delta} \widetilde{\Delta}^T \mathbf{k}(\mathbf{Z}_{\mathbf{T}\mathbf{B}}, \mathbf{y}) - \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T \widetilde{\mathcal{B}} \widetilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}). \quad (15)$$

We now kernelize $\phi(\mathbf{y})^T \mathbf{P}_{\mathbf{T}_\phi \mathbf{B}_\phi} \phi(\mathbf{y})$ in the denominator of (7) to complete the kernelization process. Using (8), (10), and (11) we have

$$\begin{aligned} \phi(\mathbf{y})^T \mathbf{P}_{\mathbf{T}_\phi \mathbf{B}_\phi} \phi(\mathbf{y}) &= \phi(\mathbf{y})^T [\mathbf{T}_\phi \ \mathbf{B}_\phi] \left[\begin{array}{cc} \mathbf{T}_\phi^T \mathbf{T}_\phi & \mathbf{T}_\phi^T \mathbf{B}_\phi \\ \mathbf{B}_\phi^T \mathbf{T}_\phi & \mathbf{B}_\phi^T \mathbf{B}_\phi \end{array} \right]^{-1} \left[\begin{array}{c} \mathbf{T}_\phi^T \\ \mathbf{B}_\phi^T \end{array} \right] \phi(\mathbf{y}) \\ &= [\mathbf{k}(\mathbf{Z}_T, \mathbf{y})^T \widetilde{\mathcal{T}} \ \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T \widetilde{\mathcal{B}}] \\ &\quad \times \left[\begin{array}{cc} \widetilde{\mathcal{T}}^T \mathbf{K}(\mathbf{Z}_T, \mathbf{Z}_T) \widetilde{\mathcal{T}} & \widetilde{\mathcal{T}}^T \mathbf{K}(\mathbf{Z}_T, \mathbf{Z}_B) \widetilde{\mathcal{B}} \\ \widetilde{\mathcal{B}}^T \mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_T) \widetilde{\mathcal{T}} & \widetilde{\mathcal{B}}^T \mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_B) \widetilde{\mathcal{B}} \end{array} \right]^{-1} \\ &\quad \times \left[\begin{array}{c} \widetilde{\mathcal{T}}^T \mathbf{k}(\mathbf{Z}_T, \mathbf{y}) \\ \widetilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}) \end{array} \right]. \end{aligned} \quad (16)$$

Finally, substituting (12), (14), and (16) into (7) the kernelized GLRT is given by

$$\mathbf{L}_{2K} = \frac{(\mathbf{k}(\mathbf{Z}_{TB}, \mathbf{y})^T \tilde{\Delta} \tilde{\Delta}^T \mathbf{k}(\mathbf{Z}_{TB}, \mathbf{y}) - \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T \tilde{\mathcal{B}} \tilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}))}{\left(\mathbf{k}(\mathbf{Z}_{TB}, \mathbf{y})^T \tilde{\Delta} \tilde{\Delta}^T \mathbf{k}(\mathbf{Z}_{TB}, \mathbf{y}) - [\mathbf{k}(\mathbf{Z}_T, \mathbf{y})^T \tilde{\mathcal{F}} \quad \mathbf{k}(\mathbf{Z}_B, \mathbf{y})^T \tilde{\mathcal{B}}] \Lambda_1^{-1} \times \begin{bmatrix} \tilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_T, \mathbf{y}) \\ \tilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}) \end{bmatrix} \right)}, \quad (17)$$

where

$$\Lambda_1 = \begin{bmatrix} \tilde{\mathcal{F}}^T \mathbf{K}(\mathbf{Z}_T, \mathbf{Z}_T) \tilde{\mathcal{F}} & \tilde{\mathcal{F}}^T \mathbf{K}(\mathbf{Z}_T, \mathbf{Z}_B) \tilde{\mathcal{B}} \\ \tilde{\mathcal{B}}^T \mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_T) \tilde{\mathcal{F}} & \tilde{\mathcal{B}}^T \mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_B) \tilde{\mathcal{B}} \end{bmatrix}. \quad (18)$$

In the above derivation (17) we assumed that each mapped input data $\phi(\mathbf{x}_i)$ in the feature space was centered $\phi_c(\mathbf{x}_i) = \phi(\mathbf{x}_i) - \hat{\boldsymbol{\mu}}_\phi$, where $\hat{\boldsymbol{\mu}}_\phi$ represents the estimated mean in the feature space given by $\hat{\boldsymbol{\mu}}_\phi = (1/N) \sum_{i=1}^N \phi(\mathbf{x}_i)$. However, the original data is usually not centered and the estimated mean in the feature space can not be explicitly computed, therefore, the kernel matrices have to be properly centered as shown by (A.14) in the appendix. The empirical kernel maps $\mathbf{k}(\mathbf{Z}_T, \mathbf{y})$, $\mathbf{k}(\mathbf{Z}_B, \mathbf{y})$, and $\mathbf{k}(\mathbf{Z}_{TB}, \mathbf{y})$ have to be centered by removing their corresponding empirical kernel map means (e.g., $\tilde{\mathbf{k}}(\mathbf{Z}_T, \mathbf{y}) = \mathbf{k}(\mathbf{Z}_T, \mathbf{y}) - (1/N) \sum_{i=1}^N k(\mathbf{y}_i, \mathbf{y}) \cdot \vec{\mathbf{1}}$, $\mathbf{y}_i \in \mathbf{Z}_T$, where $\vec{\mathbf{1}} = (1, 1, \dots, 1)^T$ is an N-dimensional vector).

4. OSP AND KERNEL OSP ALGORITHMS

4.1. Linear spectral mixture model

The OSP algorithm [2] is based on maximizing the signal-to-noise ratio (SNR) in the subspace orthogonal to the background subspace. It does not provide directly an estimate of the abundance measure for the desired end member in the mixed pixel. However, in [23] it is shown that the OSP classifier is related to the unconstrained least-squares estimate or the maximum-likelihood estimate (MLE) (similarly derived by [1]) of the unknown signature abundance by a scaling factor.

A linear mixture model for pixel \mathbf{y} consisting of p spectral bands is described by

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}, \quad (19)$$

where the $(p \times l)$ matrix \mathbf{M} represent l endmembers spectra, $\boldsymbol{\alpha}$ is a $(p \times 1)$ column vector whose elements are the coefficients that account for the proportions (abundances) of each end-member spectrum contributing to the mixed pixel, and \mathbf{n} is a $(p \times p)$ vector representing an additive zero-mean noise. Assuming now we want to identify one particular signature (e.g., a military target) with a given spectral signature \mathbf{d} and a corresponding abundance measure α_l , we can represent \mathbf{M} and $\boldsymbol{\alpha}$ in partition form as $\mathbf{M} = (\mathbf{U} : \mathbf{d})$ and $\boldsymbol{\alpha} = [\begin{smallmatrix} \gamma \\ \alpha_l \end{smallmatrix}]$ then the model (19) can be rewritten as

$$\mathbf{r} = \mathbf{d}\alpha_l + \mathbf{B}\boldsymbol{\gamma} + \mathbf{n}, \quad (20)$$

where the columns of \mathbf{B} represent the undesired spectral signatures (background signatures or eigenvectors) and the column vector $\boldsymbol{\gamma}$ is the abundance measure for the undesired spectral signatures. The reason for rewriting the model (19) as (20) is to separate \mathbf{B} from \mathbf{M} in order to show how to annihilate \mathbf{B} from an observed input pixel prior to classification.

To remove the undesired signature, the background rejection operator is given by the $(p \times p)$ matrix

$$\mathbf{P}_B^\perp = \mathbf{I} - \mathbf{B}\mathbf{B}^\#, \quad (21)$$

where $\mathbf{B}^\# = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T$ is the pseudoinverse of \mathbf{B} . Applying \mathbf{P}_B^\perp to the model (20) results in

$$\mathbf{P}_B^\perp \mathbf{r} = \mathbf{P}_B^\perp \mathbf{d}\alpha_l + \mathbf{P}_B^\perp \mathbf{n}. \quad (22)$$

The operator \mathbf{w} that maximizes the signal-to-noise ratio (SNR) of the filter output $\mathbf{w}\mathbf{P}_B^\perp \mathbf{y}$,

$$\text{SNR}(\mathbf{w}) = \frac{[\mathbf{w}^T \mathbf{P}_B^\perp \mathbf{d}] \alpha_l^2 [\mathbf{d}^T \mathbf{P}_B^\perp \mathbf{w}]}{\mathbf{w}^T \mathbf{P}_B^\perp \mathbf{E}[\mathbf{nn}^T] \mathbf{P}_B^\perp \mathbf{w}}, \quad (23)$$

as shown in [2], is given by the matched filter $\mathbf{w} = \kappa \mathbf{d}$, where κ is a constant. The OSP operator is now given by

$$\mathbf{q}_{\text{OSP}}^T = \mathbf{d}^T \mathbf{P}_B^\perp \quad (24)$$

which consists of a background signature rejecter followed by a matched filter. The output of the OSP classifier is given by

$$D_{\text{OSP}} = \mathbf{q}_{\text{OSP}}^T \mathbf{r} = \mathbf{d}^T \mathbf{P}_B^\perp \mathbf{y}. \quad (25)$$

4.2. OSP in feature space and its kernel version

A new mixture model in the high-dimensional feature space \mathcal{F} is now defined which has an equivalent nonlinear model in the input space. The new model is given by

$$\phi(\mathbf{r}) = \mathbf{M}_\phi \boldsymbol{\alpha}_\phi + \mathbf{n}_\phi, \quad (26)$$

where \mathbf{M}_ϕ is a matrix whose columns are the endmember spectra in the feature space; $\boldsymbol{\alpha}_\phi$ is a coefficient vector that accounts for the abundances of each endmember spectrum in the feature space; \mathbf{n}_ϕ is an additive zero-mean noise. Again this new model is not quite the same as explicitly mapping the model (19) by a nonlinear function into a feature space. But it is capable of representing the nonlinear relationships within the hyperspectral bands for classification. The model (26) can also be rewritten as

$$\phi(\mathbf{r}) = \phi(\mathbf{d})\alpha_{p_\phi} + \mathbf{B}_\phi \boldsymbol{\gamma}_\phi + \mathbf{n}_\phi, \quad (27)$$

where $\phi(\mathbf{d})$ represents the spectral signature of the desired target in the feature space with the corresponding abundance α_{p_ϕ} and the columns of \mathbf{B}_ϕ represent the undesired background signatures in the feature space which are obtained by finding the significant normalized eigenvectors of the background covariance matrix.

The output of the OSP classifier in the feature space is given by

$$D_{\text{OSP}_\phi} = \mathbf{q}_{\text{OSP}_\phi}^T \mathbf{r} = \phi(\mathbf{d})^T (\mathbf{I}_\phi - \mathbf{B}_\phi \mathbf{B}_\phi^T) \phi(\mathbf{r}), \quad (28)$$

where \mathbf{I}_ϕ is the identity matrix in the feature space. This output (28) is very similar to the numerator of (7). It can easily be shown [8] that the kernelized version of (28) is now given by

$$D_{\text{KOSP}} = \mathbf{k}(\mathbf{Z}_{Bd}, \mathbf{d})^T \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T \mathbf{k}(\mathbf{Z}_{Bd}, \mathbf{y}) - \mathbf{k}(\mathbf{Z}_B, \mathbf{d})^T \tilde{\mathcal{B}} \tilde{\mathcal{B}}^T \mathbf{k}(\mathbf{Z}_B, \mathbf{y}), \quad (29)$$

where $\mathbf{Z}_B = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]$ corresponds to N -input background spectral signatures and $\tilde{\mathcal{B}} = (\tilde{\beta}^1, \tilde{\beta}^2, \dots, \tilde{\beta}^{N_b})^T$ are the N_b significant eigenvectors of the centered kernel matrix (Gram matrix) $\mathbf{K}(\mathbf{Z}_B, \mathbf{Z}_B)$ normalized by the square root of their corresponding eigenvalues. $\mathbf{k}(\mathbf{Z}_B, \mathbf{r})$ and $\mathbf{k}(\mathbf{Z}_B, \mathbf{d})$ are column vectors whose entries are $k(\mathbf{x}_i, \mathbf{y})$ and $k(\mathbf{x}_i, \mathbf{d})$ for $\mathbf{x}_i \in \mathbf{Z}_B$, respectively. $\mathbf{Z}_{Bd} = \mathbf{Z}_B \cup d$ and $\tilde{\mathbf{Y}}$ is a matrix whose columns are the N_{bd} eigenvectors $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{N_{bd}})$ of the centered kernel matrix $\mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{Z}_{Bd}) = (\mathbf{K})_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{Z}_B \cup \mathbf{d}$, with nonzero eigenvalues, normalized by the square root of their associated eigenvalues. Also $\mathbf{k}(\mathbf{Z}_{Bd}, \mathbf{y})$ is the concatenated vector $[\mathbf{k}(\mathbf{Z}_B, \mathbf{r})^T \ k(\mathbf{d}, \mathbf{y})^T]^T$ and $\mathbf{k}(\mathbf{Z}_{Bd}, \mathbf{d})$ is the concatenated vector $[\mathbf{k}(\mathbf{Z}_B, \mathbf{d})^T \ k(\mathbf{d}, \mathbf{d})^T]^T$. In the above derivation (29) we assumed that the mapped input data was centered in the feature space. For noncentered data the kernel matrices and the empirical kernel maps have to be properly centered as is shown in the appendix.

5. LINEAR SMF AND KERNEL MSF

5.1. Linear SMF

In this section, we introduce the concept of linear SMF. The constrained least-squares approach is used to derive the linear SMF. Let the input spectral signal \mathbf{x} be $\mathbf{x} = [x(1), x(2), \dots, x(p)]^T$ consisting of p spectral bands. We can model each spectral observation as a linear combination of the target spectral signature and noise:

$$\mathbf{x} = a\mathbf{s} + \mathbf{n}, \quad (30)$$

where a is an attenuation constant (target abundance measure). When $a = 0$ no target is present and when $a > 0$ a target is present, the vector $\mathbf{s} = [s(1), s(2), \dots, s(p)]^T$ contains the spectral signature of the target and vector \mathbf{n} contains the additive background clutter noise.

Let us define \mathbf{X} to be a $p \times N$ matrix of the N background reference pixels obtained from the input test image. Let each observation spectral pixel to be represented as a column in the sample matrix \mathbf{X}

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]. \quad (31)$$

We can design a linear matched filter $\mathbf{w} = [w(1), w(2), \dots, w(p)]^T$ such that the desired target signal \mathbf{s} is passed through while the average filter output energy is minimized. This constrained filter design is equivalent to a constrained least-squares minimization problem, as was shown in [24–27], which is given by

$$\min_{\mathbf{w}} \{\mathbf{w}^T \hat{\mathbf{R}} \mathbf{w}\} \quad \text{subject to } \mathbf{s}^T \mathbf{w} = 1, \quad (32)$$

where minimization of $\min_{\mathbf{w}} \{\mathbf{w}^T \hat{\mathbf{R}} \mathbf{w}\}$ ensures that the background clutter noise is suppressed by the filter \mathbf{w} , and the constrain condition $\mathbf{s}^T \mathbf{w} = 1$ makes sure that the filter gives an output of unity when a target is detected.

The solution to this constrained least-squares minimization problem is given by

$$\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1} \mathbf{s}}{\mathbf{s}^T \hat{\mathbf{R}}^{-1} \mathbf{s}}, \quad (33)$$

where $\hat{\mathbf{R}}$ represents the estimated correlation matrix for the reference data. The above expression is referred to as minimum variance distortionless response (MVDR) beamformer in the array processing literature [24, 28], and more recently the same expression was also obtained for hyperspectral target detection and was called constrained energy minimization (CEM) filter or correlation-based matched filter [25, 26]. The output of the linear filter for the test input \mathbf{r} , given the estimated correlation matrix, is given by

$$y_r = \mathbf{w}^T \mathbf{r} = \frac{\mathbf{s}^T \hat{\mathbf{R}}^{-1} \mathbf{r}}{\mathbf{s}^T \hat{\mathbf{R}}^{-1} \mathbf{s}}. \quad (34)$$

If the observation data is centered a similar expression is obtained for the centered data which is given by

$$y_r = \mathbf{w}^T \mathbf{r} = \frac{\mathbf{s}^T \hat{\mathbf{C}}^{-1} \mathbf{r}}{\mathbf{s}^T \hat{\mathbf{C}}^{-1} \mathbf{s}}, \quad (35)$$

where $\hat{\mathbf{C}}$ represents the estimated covariance matrix for the reference centered data. Similarly, in [4, 5] it was shown that using the GLRT, a similar expression as in MVDR or CEM (35) can be obtained if \mathbf{n} is assumed to be the background Gaussian random noise distributed as $\mathcal{N}(0, \mathbf{C})$ where \mathbf{C} is the expected covariance matrix of only the background noise. This filter is referred to as matched filter in the signal processing literature or Capon method [29] in the array processing literature. In this paper, we implemented the matched filter given by the expression (35).

5.2. SMF in feature space and its kernel version

We now consider a model in the kernel feature space which has an equivalent nonlinear model in the original input space

$$\phi(\mathbf{x}) = a_\phi \phi(\mathbf{s}) + \mathbf{n}_\phi, \quad (36)$$

where ϕ is the nonlinear mapping associated with a kernel function k , a_ϕ is an attenuation constant (abundance measure), the high-dimensional vector $\phi(\mathbf{s})$ contains the spectral signature of the target in the feature space, and vector \mathbf{n}_ϕ contains the additive noise in the feature space.

Using the constrained least-squares approach that was explained in the previous section it can easily be shown that the equivalent matched filter \mathbf{w}_ϕ in the feature space is given by

$$\mathbf{w}_\phi = \frac{\hat{\mathbf{R}}_\phi^{-1} \phi(\mathbf{s})}{\phi(\mathbf{s})^T \hat{\mathbf{R}}_\phi^{-1} \phi(\mathbf{s})}, \quad (37)$$

where $\hat{\mathbf{R}}_\phi$ is the estimated correlation matrix in the feature space. The estimated correlation matrix is given by

$$\hat{\mathbf{R}}_\phi = \frac{1}{N} \mathbf{X}_\phi \mathbf{X}_\phi^T, \quad (38)$$

where $\mathbf{X}_\phi = [\phi(\mathbf{x}_1) \ \phi(\mathbf{x}_2) \ \cdots \ \phi(\mathbf{x}_N)]$ is a matrix whose columns are the mapped input reference data in the feature space. The matched filter in the feature space (37) is equivalent to a nonlinear matched filter in the input space and its output for an input $\phi(\mathbf{r})$ is given by

$$y_{\phi(\mathbf{r})} = \mathbf{w}_\phi^T \phi(\mathbf{r}) = \frac{\phi(\mathbf{s})^T \hat{\mathbf{R}}_\phi^{-1} \phi(\mathbf{r})}{\phi(\mathbf{s})^T \hat{\mathbf{R}}_\phi^{-1} \phi(\mathbf{s})}. \quad (39)$$

If the data was centered the matched filter for the centered data in the feature space would be

$$y_{\phi(\mathbf{r})} = \mathbf{w}_\phi^T \phi(\mathbf{r}) = \frac{\phi(\mathbf{s})^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{r})}{\phi(\mathbf{s})^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{s})}. \quad (40)$$

We now show how to kernelize the matched filter expression (40), where the resulting nonlinear matched filter is called the kernel matched filter. It is shown in the appendix that the pseudoinverse (inverse) of the estimated background covariance matrix can be written as

$$\hat{\mathbf{C}}_\phi^\# = \mathbf{X}_\phi \mathcal{B} \Lambda^{-2} \mathcal{B}^T \mathbf{X}_\phi^T. \quad (41)$$

Inserting (41) into (40) it can be rewritten as

$$y_{\phi(\mathbf{r})} = \frac{\phi(\mathbf{s})^T \mathbf{X}_\phi \mathcal{B} \Lambda^{-2} \mathcal{B}^T \mathbf{X}_\phi^T \phi(\mathbf{r})}{\phi(\mathbf{s})^T \mathbf{X}_\phi \mathcal{B} \Lambda^{-2} \mathcal{B}^T \mathbf{X}_\phi^T \phi(\mathbf{s})}. \quad (42)$$

Also using the properties of the kernel PCA as shown by (A.13) in the appendix, we have the relationship

$$\mathbf{K}^{-2} = \frac{1}{N^2} \mathcal{B} \Lambda^{-2} \mathcal{B}^T. \quad (43)$$

We denote $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X}) = (\mathbf{K})_{ij}$ an $N \times N$ Gram kernel matrix whose entries are the dot products $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. Substituting (43) into (42) the kernelized version of SMF is given by

$$y_{\mathbf{K}} = \frac{\mathbf{k}(\mathbf{X}, \mathbf{s})^T \mathbf{K}^{-2} \mathbf{k}(\mathbf{X}, \mathbf{r})}{\mathbf{k}(\mathbf{X}, \mathbf{s})^T \mathbf{K}^{-2} \mathbf{k}(\mathbf{X}, \mathbf{s})} = \frac{\mathbf{k}_s^T \mathbf{K}^{-2} \mathbf{k}_r}{\mathbf{k}_s^T \mathbf{K}^{-2} \mathbf{k}_s}, \quad (44)$$

where $\mathbf{k}_s = \mathbf{k}(\mathbf{X}, \mathbf{s})$ and $\mathbf{k}_r = \mathbf{k}(\mathbf{X}, \mathbf{r})$ are the empirical kernel maps for \mathbf{s} and \mathbf{r} , respectively. As in the previous section, the kernel matrix \mathbf{K} as well as the empirical kernel maps, \mathbf{k}_s and \mathbf{k}_r , need to be properly centered if the original data was not centered.

6. ASD AND KERNEL ASD

6.1. Linear adaptive subspace detector

In this section, the GLRT under the two competing hypotheses (\mathbf{H}_0 and \mathbf{H}_1) for a certain mixture model is described. The subpixel detection model for a measurement \mathbf{x} is expressed as

$$\begin{aligned} \mathbf{H}_0 : \mathbf{x} &= \mathbf{n}, & \text{target absent,} \\ \mathbf{H}_1 : \mathbf{x} &= \mathbf{U}\boldsymbol{\theta} + \sigma\mathbf{n}, & \text{target present,} \end{aligned} \quad (45)$$

where \mathbf{U} represents an orthogonal matrix whose orthonormal columns are the normalized eigenvectors that span the target subspace $\langle \mathbf{U} \rangle$; $\boldsymbol{\theta}$ is an unknown vector whose entries are coefficients that account for the abundances of the corresponding column vectors of \mathbf{U} and \mathbf{n} represents Gaussian random noise distributed as $\mathcal{N}(0, \mathbf{C})$.

In model (45), \mathbf{x} is assumed to be a background noise under \mathbf{H}_0 and a linear combination of a target subspace signal and a scaled background noise, distributed as $\mathcal{N}(\mathbf{U}\boldsymbol{\theta}, \sigma^2\mathbf{C})$, under \mathbf{H}_1 . The background noise under the two hypotheses is represented by the same covariance but different variances because of the existence of subpixel targets under \mathbf{H}_1 . The GLRT for the subpixel problem described by (45), the so-called ASD [5], is given by

$$D_{\text{ASD}}(\mathbf{x}) = \frac{\mathbf{x}^T \hat{\mathbf{C}}^{-1} \mathbf{U} (\mathbf{U}^T \hat{\mathbf{C}}^{-1} \mathbf{U})^{-1} \mathbf{U}^T \hat{\mathbf{C}}^{-1} \mathbf{x}}{\mathbf{x}^T \hat{\mathbf{C}}^{-1} \mathbf{x}} \underset{H_0}{\overset{H_1}{\geq}} \eta_{\text{ASD}}, \quad (46)$$

where $\hat{\mathbf{C}}$ is the MLE of the covariance \mathbf{C} and η_{ASD} represents a threshold. Expression (46) has a constant false alarm rate (CFAR) property and is also referred to as the adaptive cosine estimator because (46) measures the angle between $\tilde{\mathbf{x}}$ and $\langle \tilde{\mathbf{U}} \rangle$, where $\tilde{\mathbf{x}} = \hat{\mathbf{C}}^{-1/2} \mathbf{x}$ and $\tilde{\mathbf{U}} = \hat{\mathbf{C}}^{-1/2} \mathbf{U}$.

6.2. ASD in the feature space and its kernel version

We define a new subpixel model in a high-dimensional feature space \mathcal{F} given by

$$\begin{aligned} \mathbf{H}_{0_\phi} : \phi(\mathbf{x}) &= \mathbf{n}_\phi, & \text{target absent,} \\ \mathbf{H}_{1_\phi} : \phi(\mathbf{x}) &= \mathbf{U}_\phi \boldsymbol{\theta}_\phi + \sigma_\phi \mathbf{n}_\phi, & \text{target present,} \end{aligned} \quad (47)$$

where \mathbf{U}_ϕ represents a matrix whose M_1 orthonormal columns are the normalized eigenvectors that span target subspace $\langle \mathbf{U}_\phi \rangle$ in \mathcal{F} ; $\boldsymbol{\theta}_\phi$ is unknown vectors whose entries are coefficients that account for the abundances of the corresponding column vectors of \mathbf{U}_ϕ ; \mathbf{n}_ϕ represents Gaussian random noise distributed by $\mathcal{N}(0, \mathbf{C}_\phi)$; and σ_ϕ is the noise variance under \mathbf{H}_{1_ϕ} . The GLRT for the model (47) in \mathcal{F} is now

given by

$$D(\phi(\mathbf{x})) = \frac{\phi(\mathbf{x})^T \hat{\mathbf{C}}_\phi^{-1} \mathbf{U}_\phi (\mathbf{U}_\phi^T \hat{\mathbf{C}}_\phi^{-1} \mathbf{U}_\phi)^{-1} \mathbf{U}_\phi^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{x})}{\phi(\mathbf{x})^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{x})}, \quad (48)$$

where $\hat{\mathbf{C}}_\phi$ is the MLE of \mathbf{C}_ϕ .

We now show how to kernelize the ASD expression (48) in the feature space. The inverse (pseudoinverse) background covariance matrix in (48) can be represented by its eigenvector decomposition (see the appendix) given by the expression

$$\hat{\mathbf{C}}_\phi^\# = \mathbf{X}_\phi \mathcal{B} \Lambda^{-2} \mathcal{B}^T \mathbf{X}_\phi^T, \quad (49)$$

where $\mathbf{X}_\phi = [\phi_c(\mathbf{x}_1) \ \phi_c(\mathbf{x}_2) \ \cdots \ \phi_c(\mathbf{x}_N)]$ represents the centered vectors in the feature space corresponding to N independent background spectral signatures, $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_N]$ and $\mathcal{B} = [\boldsymbol{\beta}^1 \ \boldsymbol{\beta}^2 \ \cdots \ \boldsymbol{\beta}^{N_1}]$ are the nonzero eigenvectors of the centered kernel matrix (Gram matrix) $\mathbf{K}(\mathbf{X}, \mathbf{X})$. Similarly, \mathbf{U}_ϕ is given by

$$\mathbf{U}_\phi = \mathbf{Y}_\phi \tilde{\mathcal{F}}, \quad (50)$$

where $\mathbf{Y}_\phi = [\phi_c(\mathbf{y}_1) \ \phi_c(\mathbf{y}_2) \ \cdots \ \phi_c(\mathbf{y}_M)]$ are the centered vectors in the feature space corresponding to the M independent target spectral signatures $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_M]$, and $\tilde{\mathcal{F}} = [\tilde{\boldsymbol{\alpha}}^1 \ \tilde{\boldsymbol{\alpha}}^2 \ \cdots \ \tilde{\boldsymbol{\alpha}}^{M_1}]$, $M_1 < M$, is a matrix consisting of the M_1 eigenvectors of the kernel matrix $\mathbf{K}(\mathbf{Y}, \mathbf{Y})$ normalized by the square root of their corresponding eigenvalues. Now, the term $\phi(\mathbf{x})^T \hat{\mathbf{C}}_\phi^{-1} \mathbf{U}_\phi$ in the numerator of (48) becomes

$$\begin{aligned} \phi(\mathbf{x})^T \hat{\mathbf{C}}_\phi^{-1} \mathbf{U}_\phi &= \phi(\mathbf{x})^T \mathbf{X}_\phi \mathcal{B} \Lambda^{-2} \mathcal{B}^T \mathbf{X}_\phi^T \mathbf{Y}_\phi \tilde{\mathcal{F}} \\ &= \mathbf{k}(\mathbf{x}, \mathbf{X})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{K}(\mathbf{X}, \mathbf{Y}) \tilde{\mathcal{F}} \equiv \mathbf{K}_x, \end{aligned} \quad (51)$$

where $\mathcal{B} \Lambda^{-2} \mathcal{B}^T$ is replaced by $\mathbf{K}(\mathbf{X}, \mathbf{X})^{-2}$ using the relationship (A.13), as shown in the appendix.

Similarly,

$$\begin{aligned} \mathbf{U}_\phi^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{x}) &= \tilde{\mathcal{F}}^T \mathbf{K}(\mathbf{X}, \mathbf{Y})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{k}(\mathbf{x}, \mathbf{X}) = \mathbf{K}_x^T, \\ \mathbf{U}_\phi^T \hat{\mathbf{C}}_\phi^{-1} \mathbf{U}_\phi &= \tilde{\mathcal{F}}^T \mathbf{K}(\mathbf{X}, \mathbf{Y})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{K}(\mathbf{X}, \mathbf{Y}) \tilde{\mathcal{F}}. \end{aligned} \quad (52)$$

The denominator of (48) is also expressed as

$$\phi(\mathbf{x})^T \hat{\mathbf{C}}_\phi^{-1} \phi(\mathbf{x}) = \mathbf{k}(\mathbf{x}, \mathbf{X})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{k}(\mathbf{x}, \mathbf{X}). \quad (53)$$

Finally, the kernelized expression of (48) is given by

$$D_{\text{KASD}}(\mathbf{x}) = \frac{\mathbf{K}_x [\tilde{\mathcal{F}}^T \mathbf{K}(\mathbf{X}, \mathbf{Y})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{K}(\mathbf{X}, \mathbf{Y}) \tilde{\mathcal{F}}]^{-1} \mathbf{K}_x^T}{\mathbf{k}(\mathbf{x}, \mathbf{X})^T \mathbf{K}(\mathbf{X}, \mathbf{X})^{-2} \mathbf{k}(\mathbf{x}, \mathbf{X})}. \quad (54)$$

As in the previous sections all the kernel matrices $\mathbf{K}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{K}(\mathbf{X}, \mathbf{X})$ as well as the empirical kernel maps need to be properly centered.

7. EXPERIMENTAL RESULTS

The proposed kernel-based matched signal detectors, the kernel MSD (KMSD), kernel ASD (KASD), kernel OSP (KOSP), and kernel SMF (KSMF) as well as the corresponding conventional detectors are implemented based on two different types of data sets—illustrative toy data sets and real-hyperspectral images that contain military targets. The Gaussian RBF kernel, $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/c)$, was used to implement the kernel-based detectors, where c represents the width of the Gaussian distribution. The value of c was chosen such that the overall data variations can be fully exploited by the Gaussian RBF function; the value for c was determined experimentally.

7.1. Illustrative toy examples

Figures 1 and 2 show contour and surface plots of the conventional detectors and the kernel-based detectors, on two different types of two-dimensional toy data sets: a Gaussian mixture in Figure 1 and nonlinearly mapped data in Figure 2. In the contour and surface plots, data points for the desired target were represented by the star-shaped symbol and the background points were represented by the circles. In Figure 2 the two-dimensional data points $\mathbf{x} = (x, y)$ for each class were obtained by nonlinearly mapping the original Gaussian mixture data points $\mathbf{x}_0 = (x_0, y_0)$ in Figure 1. All the data points in Figure 2 were nonlinearly mapped by $\mathbf{x} = (x, y) = (x_0, x_0^2 + y_0)$. In the new data set the second component of each data point is nonlinearly related to its first component.

For both data sets, the contours generated by the kernel-based detectors are highly nonlinear and naturally following the dispersion of the data and thus successfully separating the two classes, as opposed to the linear contours obtained by the conventional detectors. Therefore, the kernel-based detectors clearly provided significantly improved discrimination over the conventional detectors for both the Gaussian mixture and nonlinearly mapped data. Among the kernel-based detectors, KMSD and KASD outperform KOSP and KSMF mainly because targets in KMSD and KASD are better represented by the associated target subspace than by a single spectral signature used in KOSP and KSMF. Note that the contour plots for MSD (Figures 1(a) and 2(a)) represent only the numerator of (4) because the denominator becomes unstable for the two-dimensional cases; that is, the value inside the brackets $(\mathbf{I} - \mathbf{P}_{\text{TB}})$ becomes zero for the two-dimensional data.

7.2. Hyperspectral images

In this section, hyperspectral digital imagery collection experiment (HYDICE) images from the desert radiance II data collection (DR-II) and forest radiance I data collection (FR-I) were used to compare detection performance between the kernel-based and conventional methods. The HYDICE imaging sensor generates 210 bands across the whole spectral

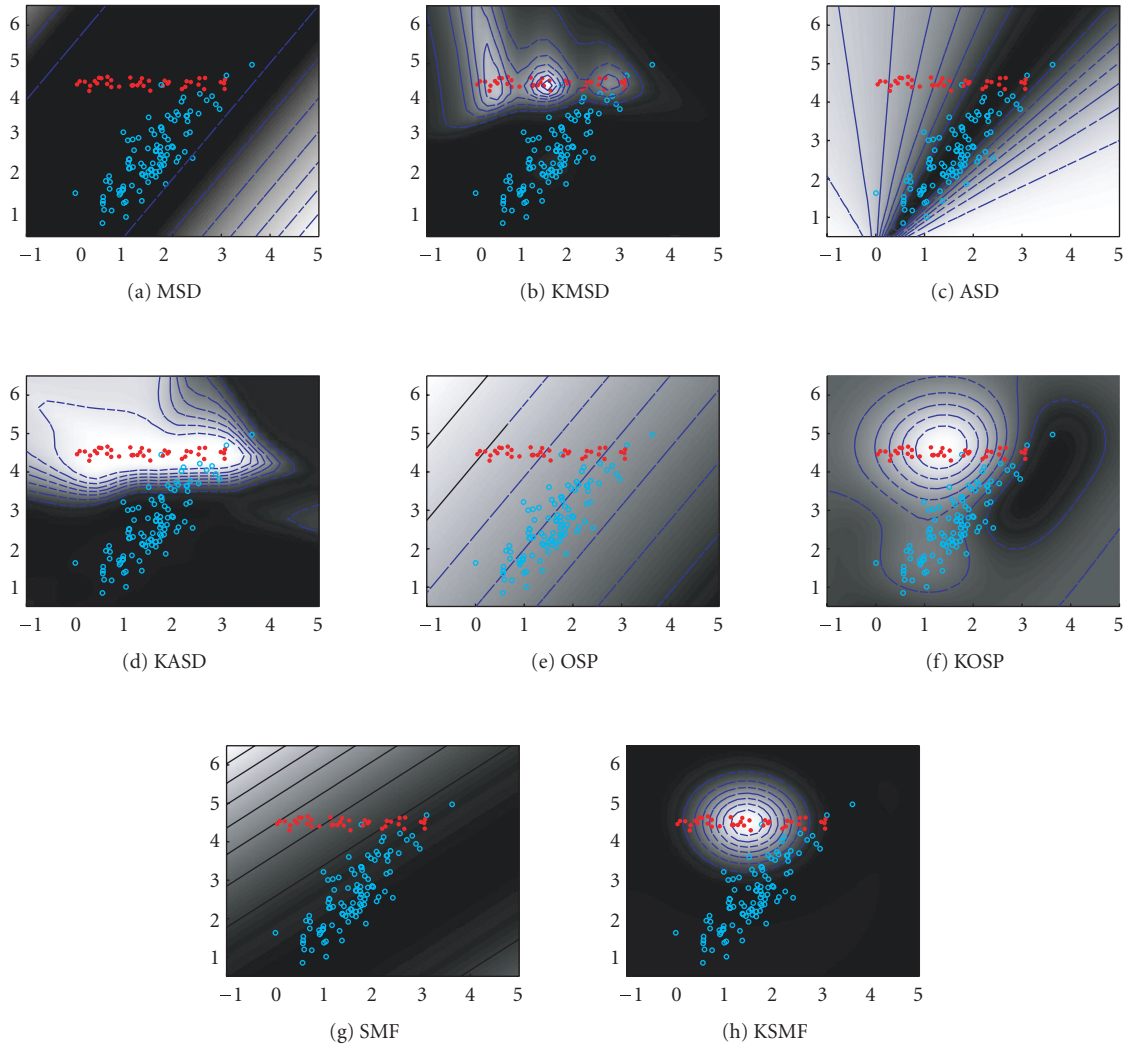


FIGURE 1: Contour and surface plots of the conventional matched signal detectors and their corresponding kernel versions on a toy dataset (a mixture of Gaussian).

range ($0.4\text{--}2.5\mu\text{m}$) which includes the visible and short-wave infrared (SWIR) bands. But we only use 150 bands by discarding water absorption and low-SNR bands; the spectral bands used are the 23rd–101st, 109th–136th, and 152nd–194th for the HYDICE images. The DR-II image includes 6 military targets along the road and the FR-I image includes total 14 targets along the tree line, as shown in the sample band images in Figure 3. The detection performance of the DR-II and FR-I images was provided in both the qualitative and quantitative—the receiver operating characteristics (ROC) curves—forms. The spectral signatures of the desired target and undesired background signatures were directly collected from the given hyperspectral data to implement both the kernel-based and conventional detectors.

All the pixel vectors in a test image are first normalized by a constant, which is a maximum value obtained from all the spectral components of the spectral vectors in the

corresponding test image, so that the entries of the normalized pixel vectors fit into the interval of spectral values between zero and one. The rescaling of pixel vectors was mainly performed to effectively utilize the dynamic range of Gaussian RBF kernel.

Figures 4–7 show the detection results including the ROC curves generated by applying the kernel-based and conventional detectors to the DR-II and FR-I images. In general, the detected targets by the kernel-based detectors are much more evident than the ones detected by the conventional detectors, as shown in Figures 4 and 5. Figures 6 and 7 show the ROC curve plots for the kernel-based and conventional detectors for the DR-II and FR-I images; in general, the kernel-based detectors outperformed the conventional detectors. In particular, KMSD performed the best of all kernel-based detectors detecting all the targets and significantly suppressing the background. The performance superiority of KMSD is mainly attributed to the utilization of both the target and

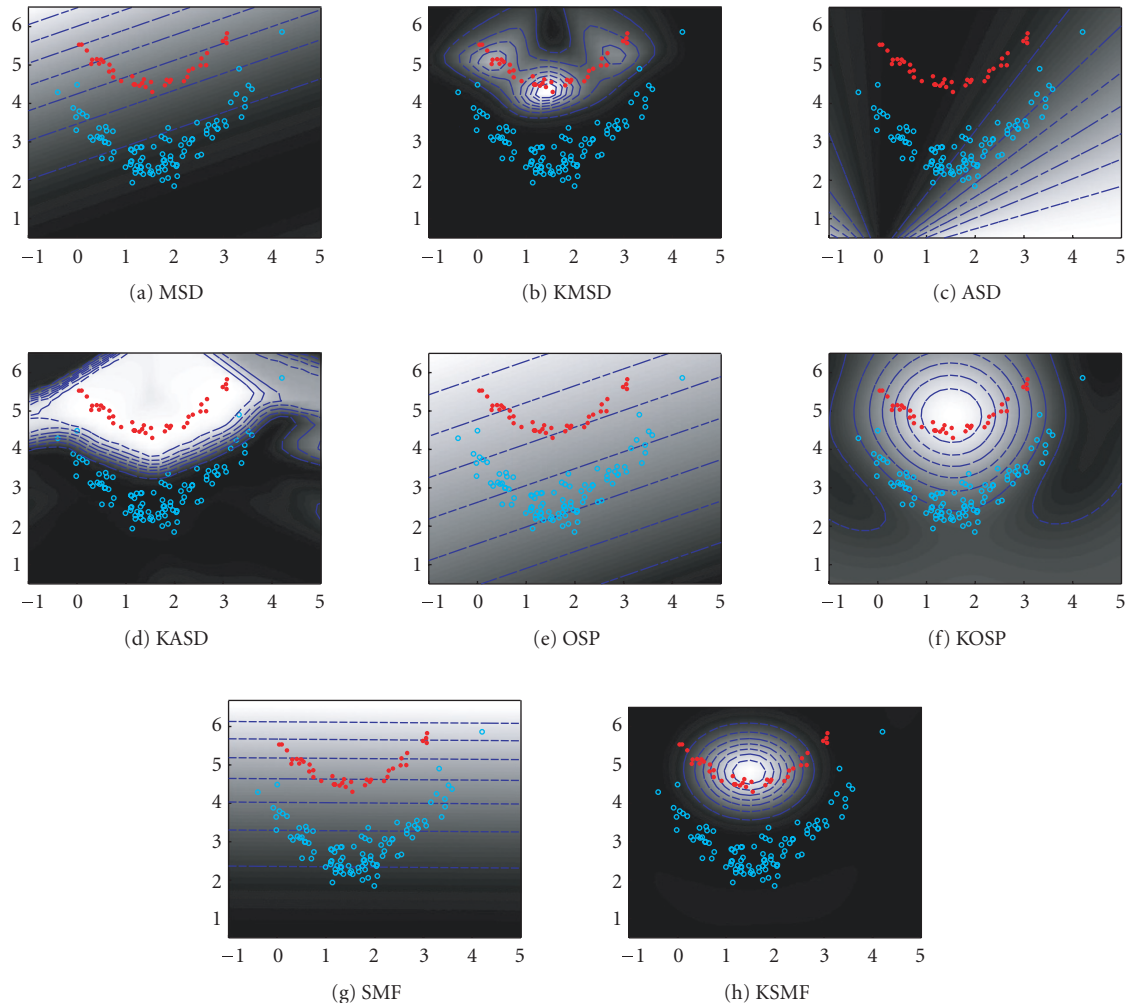


FIGURE 2: Contour and surface plots of the conventional matched signal detectors and their corresponding kernel versions on a toy dataset: in this toy example, the Gaussian mixture data shown in Figure 1 was modified to generate nonlinearly mixed data.

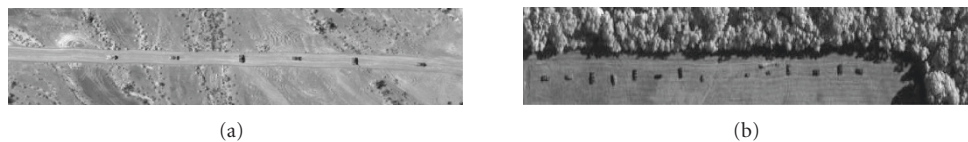


FIGURE 3: Sample band images from (a) the DR-II image and (b) the FR-I image.

background kernel subspaces representing the target and background signals in the feature space, respectively.

8. CONCLUSIONS

In this paper, kernel versions of several matched signal detectors, such as KMSD, KOSP, KSMF, and KASD have been implemented using the kernel-based learning theory. Performance comparison between the matched signal detectors and their corresponding nonlinear versions was conducted based

on two-dimensional toy examples as well as real hyperspectral images. It is shown that the kernel-based nonlinear versions of these detectors outperform the linear versions.

APPENDIX

KERNEL PCA

In this appendix we will show the derivation of the kernel PCA and its properties. Our goal is to prove the relationships

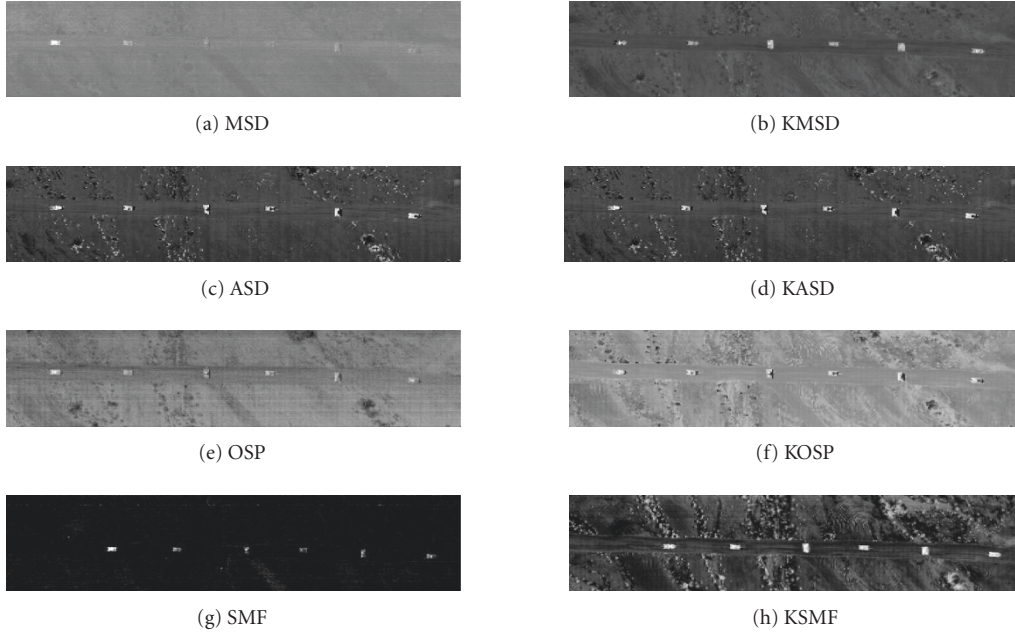


FIGURE 4: Detection results for the DR-II image using the conventional detectors and the corresponding kernel versions.

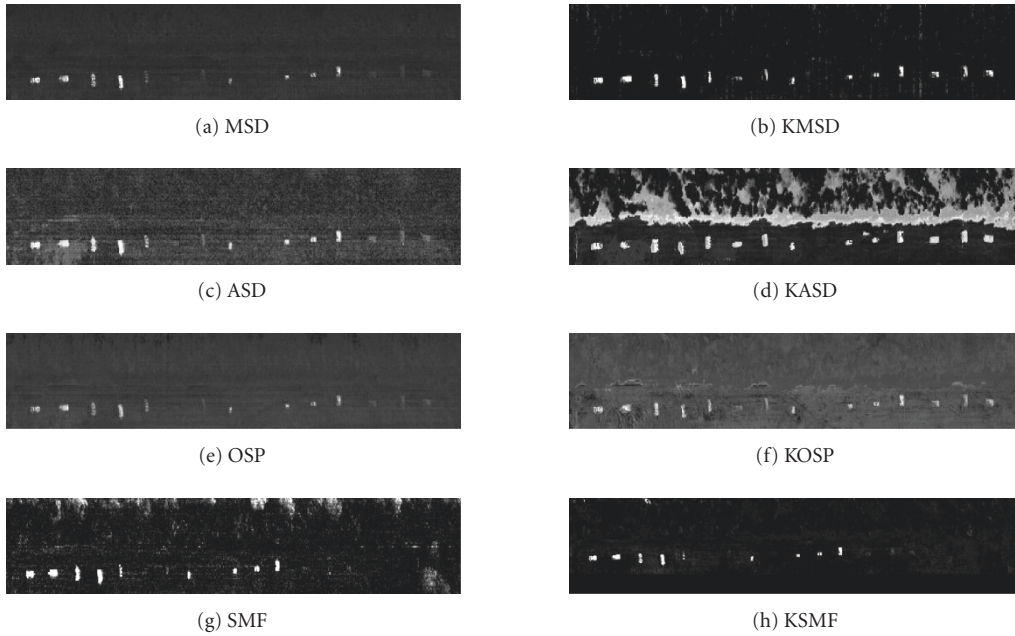


FIGURE 5: Detection results for the FR-I image using the conventional detectors and the corresponding kernel versions.

(49) and (A.13) from the kernel PCA properties. To drive the kernel PCA consider the estimated background clutter covariance matrix in the feature space and assume that the input data has been normalized (centered) to have zero mean. The estimated covariance matrix in the feature space is given by

$$\hat{\mathbf{C}}_{\phi} = \frac{1}{N} \mathbf{X}_{\phi} \mathbf{X}_{\phi}^T. \quad (\text{A.1})$$

The PCA eigenvectors are computed by solving the eigenvalue problem

$$\begin{aligned} \lambda \mathbf{v}_{\phi} &= \hat{\mathbf{C}}_{\phi} \mathbf{v}_{\phi} = \frac{1}{N} \sum_{i=1}^N \phi(\mathbf{x}_i) \phi(\mathbf{x}_i)^T \mathbf{v}_{\phi} \\ &= \frac{1}{N} \sum_{i=1}^N \langle \phi(\mathbf{x}_i), \mathbf{v}_{\phi} \rangle \phi(\mathbf{x}_i), \end{aligned} \quad (\text{A.2})$$

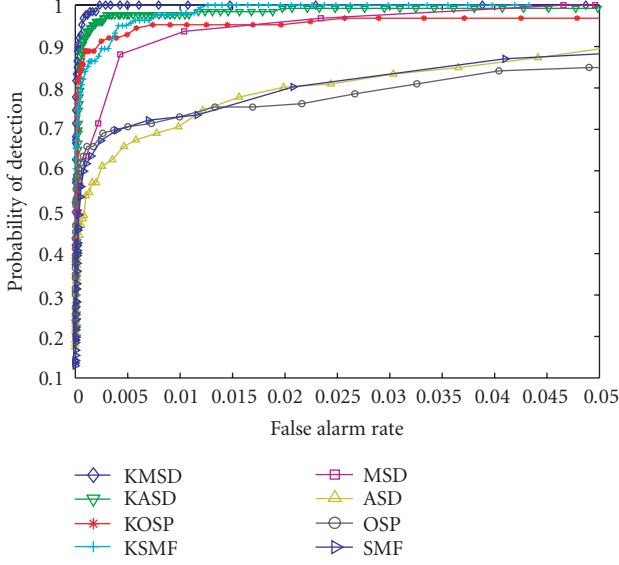


FIGURE 6: ROC curves obtained by conventional detectors and the corresponding kernel versions for the DR-II image.

where \mathbf{v}_ϕ is an eigenvector in \mathcal{F} with a corresponding nonzero eigenvalue λ . Equation (A.2) indicates that each eigenvector \mathbf{v}_ϕ with corresponding $\lambda \neq 0$ are spanned by $\phi(\mathbf{x}_1), \dots, \phi(\mathbf{x}_N)$; that is,

$$\mathbf{v}_\phi = \sum_{i=1}^N \lambda^{-1/2} \beta_i \phi(\mathbf{x}_i) = \mathbf{X}_\phi \boldsymbol{\beta} \lambda^{-1/2}, \quad (\text{A.3})$$

where $\mathbf{X}_\phi = [\phi(\mathbf{x}_1) \ \phi(\mathbf{x}_2) \ \dots \ \phi(\mathbf{x}_N)]$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)^T$. Substituting (A.3) into (A.2) and multiplying with $\phi(\mathbf{x}_n)^T$ yields

$$\begin{aligned} & \lambda \sum_{i=1}^N \beta_i \langle \phi(\mathbf{x}_n), \phi(\mathbf{x}_i) \rangle \\ &= \frac{1}{N} \sum_{i=1}^N \beta_i \phi(\mathbf{x}_n) \phi(\mathbf{x}_i) \phi(\mathbf{x}_i)^T \sum_{i=1}^N \phi(\mathbf{x}_i) \end{aligned} \quad (\text{A.4})$$

$$= \frac{1}{N} \sum_{i=1}^N \beta_i \langle \phi(\mathbf{x}_n), \sum_{j=1}^N \phi(\mathbf{x}_j) \left\langle \phi(\mathbf{x}_j), \phi(\mathbf{x}_i) \right\rangle \rangle$$

$$\forall n = 1, \dots, N.$$

We denote by $\mathbf{K} = \mathbf{K}(\mathbf{X}, \mathbf{X}) = (\mathbf{K})_{ij}$ the $N \times N$ kernel matrix whose entries are the dot products $\langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$. Equation (A.4) can be rewritten as

$$N\lambda\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta}, \quad (\text{A.5})$$

where $\boldsymbol{\beta}$ turn out to be the eigenvectors with nonzero eigenvalues of the centered kernel matrix \mathbf{K} . Therefore, the Gram matrix can be written in terms of it eigenvector decomposition as

$$\mathbf{K} = \mathbf{B}\boldsymbol{\Omega}\mathbf{B}^T, \quad (\text{A.6})$$

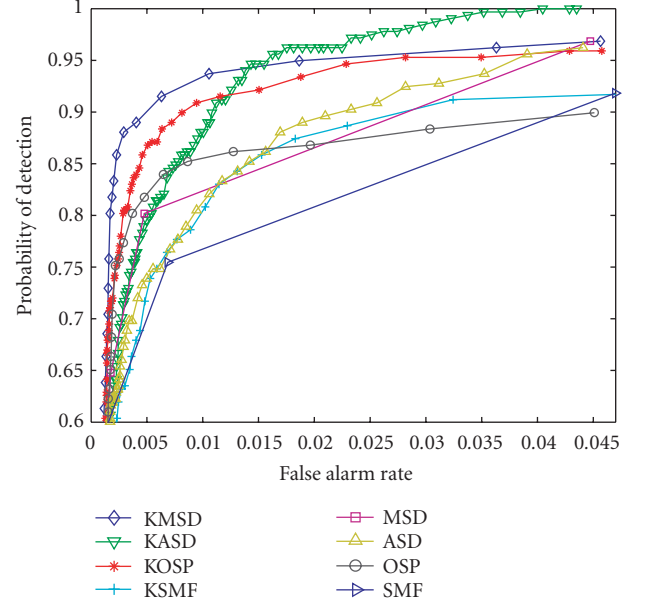


FIGURE 7: ROC curves obtained by conventional detectors and the corresponding kernel versions for the FR-I image.

where $\mathbf{B} = [\boldsymbol{\beta}^1 \ \boldsymbol{\beta}^2 \ \dots \ \boldsymbol{\beta}^N]$ are the eigenvectors of the kernel matrix and $\boldsymbol{\Omega}$ is a diagonal matrix with diagonal values equal to the nonzero eigenvalues of the kernel matrix \mathbf{K} .

Similarly, from the definition of PCA in the feature space (A.2) the estimated background covariance matrix is decomposed as

$$\hat{\mathbf{C}}_\phi = \mathbf{V}_\phi \boldsymbol{\Lambda} \mathbf{V}_\phi^T, \quad (\text{A.7})$$

where $\mathbf{V}_\phi = [\mathbf{v}_\phi^1 \ \mathbf{v}_\phi^2 \ \dots \ \mathbf{v}_\phi^N]$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with its diagonal elements being the nonzero eigenvalues of $\hat{\mathbf{C}}_\phi$. From (A.2) and (A.5) the eigenvalues of the covariance matrix $\boldsymbol{\Lambda}$ in the feature space and the eigenvalues of the kernel matrix $\boldsymbol{\Omega}$ are related by

$$\boldsymbol{\Lambda} = \frac{1}{N} \boldsymbol{\Omega}. \quad (\text{A.8})$$

Substituting (A.8) into (A.6) we obtain the relationship

$$\mathbf{K} = N\mathbf{B}\boldsymbol{\Lambda}\mathbf{B}^T, \quad (\text{A.9})$$

where N is a constant representing the total number of background clutter samples, which can be ignored.

The sample covariance matrix in the feature space is rank deficient consisting of N columns and the number of its rows is the same as the dimensionality of the feature space which could be infinite. Therefore, its inverse cannot be obtained but its pseudoinverse can be written as [30]

$$\hat{\mathbf{C}}_\phi^\# = \mathbf{V}_\phi \boldsymbol{\Lambda}^{-1} \mathbf{V}_\phi^T, \quad (\text{A.10})$$

where $\boldsymbol{\Lambda}^{-1}$ consists of only the reciprocals of the nonzero eigenvalues (which is determined by the effective rank of the

covariance matrix [30]). The eigenvectors \mathbf{V}_ϕ in the feature space can be represented as

$$\mathbf{V}_\phi = \mathbf{X}_\phi \mathcal{B} \mathbf{\Lambda}^{-1/2} = \mathbf{X}_\phi \widetilde{\mathcal{B}}, \quad (\text{A.11})$$

then the pseudoinverse background covariance matrix $\widehat{\mathbf{C}}_\phi^\#$ can be written as

$$\widehat{\mathbf{C}}_\phi^\# = \mathbf{V}_\phi \mathbf{\Lambda}^{-1} \mathbf{V}_\phi^T = \mathbf{X}_\phi \mathcal{B} \mathbf{\Lambda}^{-2} \mathcal{B}^T \mathbf{X}_\phi^T. \quad (\text{A.12})$$

The maximum number of eigenvectors in the pseudoinverse is equal to the number of nonzero eigenvalues (or the number of independent data samples), which cannot be exactly determined due to round-off error in the calculations. Therefore, the effective rank [30] is determined by only including the eigenvalues that are above a small threshold. Similarly, the inverse Gram matrix \mathbf{K}^{-1} can also be written as

$$\mathbf{K}^{-1} = \frac{1}{N} \mathcal{B} \mathbf{\Lambda}^{-1} \mathcal{B}^T. \quad (\text{A.13})$$

If the data samples are not independent then the pseudoinverse of the Gram matrix has to be used, which is the same as (A.13) except only that the eigenvectors with eigenvalues above a small threshold are included in order to obtain a numerically stable inverse.

In the derivation of the kernel PCA we assumed that the data has already been centered in the feature space by removing the sample mean. However, the sample mean cannot be directly removed in the feature space due to the high dimensionality of \mathcal{F} . That is the kernel PCA needs to be derived in terms of the original uncentered input data. Therefore, the kernel matrix $\widehat{\mathbf{K}}$ needs to be properly centered [12]. The effect of centering on the kernel PCA can be seen by replacing the uncentered \mathbf{X}_ϕ with the centered $\mathbf{X}_\phi - \mu_\phi$ (where μ_ϕ is the mean of the reference input data) in the estimation of the covariance matrix expression (A.1). The resulting centered $\widehat{\mathbf{K}}$ is shown in [12] to be given by

$$\widehat{\mathbf{K}} = (\mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K} \mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N), \quad (\text{A.14})$$

where the $N \times N$ matrix $(\mathbf{1}_N)_{ij} = 1/N$. In the above (A.6) and (A.13) the kernel matrix \mathbf{K} needs to be replaced by the centered kernel matrix $\widehat{\mathbf{K}}$.

REFERENCES

- [1] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2156, 1994.
- [2] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, 1994.
- [3] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI*, vol. 4049 of *Proceedings of SPIE*, pp. 2–17, Orlando, Fla, USA, April 2000.
- [4] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, 1992.
- [5] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2538–2541, 1999.
- [6] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, 2001.
- [7] H. Kwon and N. M. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 178–194, 2006.
- [8] H. Kwon and N. M. Nasrabadi, "Kernel orthogonal subspace projection for hyperspectral signal classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2952–2962, 2005.
- [9] H. Kwon and N. M. Nasrabadi, "Kernel adaptive subspace detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 3, no. 2, pp. 271–275, 2006.
- [10] H. Kwon and N. M. Nasrabadi, "Kernel spectral matched filter for hyperspectral target detection," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. 665–668, Philadelphia, Pa, USA, March 2005.
- [11] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1999.
- [12] B. Schölkopf and A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, Mass, USA, 2002.
- [13] B. Schölkopf, A. J. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [14] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [15] M. Girolami, "Mercer kernel-based clustering in feature space," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 780–784, 2002.
- [16] A. Ruiz and P. E. Lopez-de-Teruel, "Nonlinear kernel-based statistical pattern analysis," *IEEE Transactions on Neural Networks*, vol. 12, no. 1, pp. 16–32, 2001.
- [17] C. H. Park and H. Park, "Nonlinear feature extraction based on centroids and kernel functions," *Pattern Recognition*, vol. 37, no. 4, pp. 801–810, 2004.
- [18] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.
- [19] E. Maeda and H. Murase, "Multi-category classification by kernel based nonlinear subspace method," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '99)*, vol. 2, pp. 1025–1028, Phoenix, Ariz, USA, March 1999.
- [20] M. M. Dundar and D. A. Landgrebe, "Toward an optimal supervised classifier for the analysis of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 271–277, 2004.
- [21] E. Pekalska, P. Paclik, and R. P. W. Duin, "A generalized kernel approach to dissimilarity based classification," *Journal of Machine Learning Research*, vol. 2, pp. 175–211, 2001.
- [22] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Transactions on Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.
- [23] J. J. Settle, "On the relationship between spectral unmixing and subspace projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 4, pp. 1045–1046, 1996.

- [24] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [25] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Ph.D. dissertation, Department of Computer Science & Electrical Engineering, University of Maryland, Baltimore, Md, USA, 1993.
- [26] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Kluwer Academic / Plenum, New York, NY, USA, 2003.
- [27] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, Reading, Mass, USA, 1991.
- [28] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.
- [29] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [30] G. Strang, *Linear Algebra and Its Applications*, Harcourt Brace, Orlando, Fla, USA, 1986.

Heesung Kwon received the B.S. degree in electronic engineering from Sogang University, Seoul, Korea, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo in 1995 and 1999, respectively. From 1983 to 1993, he was with Samsung Electronics Corp., where he worked as an engineer. Since 1996, he has been working at the US Army Research Laboratory, Adelphi, Md. His interests include hyperspectral image analysis, pattern recognition, statistical learning, and image/video compression. He has published over 45 papers on these topics in leading journals and conferences.



Nasser M. Nasrabadi received the B.S. (Eng.) and Ph.D. degrees in electrical engineering from the Imperial College of Science and Technology, University of London, London, England, in 1980 and 1984, respectively. From October 1984 to December 1984 he worked for IBM (UK) as a Senior Programmer. During 1985 to 1986 he worked with Philips Research Laboratory in NY as a Member of technical staff. From 1986 to 1991 he was an Assistant Professor in the Department of Electrical Engineering at Worcester Polytechnic Institute, Worcester, Mass. From 1991 to 1996 he was an Associate Professor with the Department of Electrical and Computer Engineering at State University of New York at Buffalo, Buffalo, NY. Since September 1996 he has been a Senior Research Scientist (ST) with the US Army Research Laboratory working on image processing and automatic target recognition. He has served as an Associate Editor for the IEEE Transactions on Image Processing, the IEEE Transactions on Circuits, Systems, and Video Technology, and the IEEE Transactions on Neural Networks. He is also a Fellow of IEEE and SPIE. His current research interests are in kernel-based learning algorithms, automatic target recognition, and neural networks applications to image processing.

