Kernel Matched Signal Detectors for Hyperspectral Target Detection

Heesung Kwon and Nasser M. Nasrabadi
U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD 20783-1197

Abstract

In this paper, we compare several detection algorithms that are based on spectral matched (subspace) filters. Nonlinear (*kernel*) versions of these spectral matched (subspace) detectors are also discussed and their performance is compared with the linear versions. These kernel-based detectors exploit the nonlinear correlations between the spectral bands that are ignored by the conventional detectors. Several well-known matched detectors, such as matched subspace detector, orthogonal subspace detector, spectral matched filter and adaptive subspace detector (adaptive cosine estimator) are extended to their corresponding kernel versions by using the idea of kernel-based learning theory. In kernel-based detection algorithms the data is implicitly mapped into a high dimensional kernel feature space by a nonlinear mapping which is associated with a kernel function. The 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.

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 inter-band correlations that could be crucial to discriminate between target and background. In this paper, our aim is to introduce nonlinear versions of MSD, OSD, SMF and ASD detectors which effectively exploits the higher order spectral inter-band correlations in a high (possibly infinite) dimensional feature space associated with a certain nonlinear mapping via kernel-based learning methods [7]. A nonlinear mapping of the input data into a high dimensional feature space is often expected to increase the data separability and reduce the complexity of the corresponding data structure. The nonlinear versions of a number of signal processing techniques such as principal component analysis (PCA) [8], Fisher discriminant analysis [9], linear classifiers [10], and kernel-based anomaly detection [11] have already been defined in a kernel space.

This paper is organized as follows. Section 2 provides the background to the kernel-based learning methods and kernel trick. Section 3 introduces a linear matched subspace 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 ad its kernel version in the feature space and reformulate the the expression 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 versions of these algorithms is provided in Section 7 and conclusions are given in Section 8.

2 Kernel-based Learning and Kernel Trick

Suppose that the input hyperspectral data is represented by the data space ($\mathcal{X} \subseteq \mathcal{R}^l$) and $\mathcal{F}$ is a feature space associated with $\mathcal{X}$ by a nonlinear mapping function Φ

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

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 Eq. (2) to implement dot products in feature space by employing kernel functions [7]. The idea in kernel-based tech-

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 IEEE

niques 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

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j), \tag{2}$$

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 [7] for detailed information about the properties of different kernels and kernel-based learning. Our choice of kernel in this paper is the Gaussian 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 $\mathbf{H_1}$

$$\begin{aligned} &\mathbf{H}_0 : \mathbf{y} = \mathbf{B}\zeta + \mathbf{n}, \\ &\mathbf{H}_1 : \mathbf{y} = \mathbf{T}\theta + \mathbf{B}\zeta + \mathbf{n} = \begin{bmatrix} \mathbf{T} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \theta \\ \zeta \end{bmatrix} + \mathbf{n}, \end{aligned} \tag{3}$$

where $\mathbf{T}$ and $\mathbf{B}$ represent orthogonal matrices whose p-dimensional column vectors span the target and background subspaces, respectively; θ and ζ 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 the model (3) was derived in [1], given as

$$\mathbf{L}_2(\mathbf{y}) = \frac{\mathbf{y}^T(\mathbf{I} - \mathbf{P_B})\,\mathbf{y}}{\mathbf{y}^T(\mathbf{I} - \mathbf{P_{TB}})\,\mathbf{y}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{4}$$

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 $< \mathbf{TB} >$

$$\mathbf{P_{TB}} = \begin{bmatrix} \mathbf{T}\,\mathbf{B} \end{bmatrix} \begin{bmatrix} [\mathbf{T}\,\mathbf{B}]^T [\mathbf{T}\,\mathbf{B}] \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{T}\,\mathbf{B} \end{bmatrix}^T. \tag{5}$$

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

$$\begin{aligned} &\mathbf{H}_{0_\Phi} : \Phi(\mathbf{y}) = \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi, \\ &\mathbf{H}_{1_\Phi} : \Phi(\mathbf{y}) = \mathbf{T}_\Phi \theta_\Phi + \mathbf{B}_\Phi \zeta_\Phi + \mathbf{n}_\Phi \\ &\qquad = \begin{bmatrix} \mathbf{T}_\Phi & \mathbf{B}_\Phi \end{bmatrix} \begin{bmatrix} \theta_\Phi \\ \zeta_\Phi \end{bmatrix} + \mathbf{n}_\Phi, \end{aligned} \tag{6}$$

where $\mathbf{T}_\Phi$ and $\mathbf{B}_\Phi$ represent full-rank matrices whose column vectors span target and background subspaces $< \mathbf{B}_\Phi >$ and $< \mathbf{T}_\Phi >$ in $\mathcal{F}$, respectively; θ_Φ and ζ_Φ are unknown vectors whose entries are coefficients that account for the abundances of the corresponding column vectors of $\mathbf{T}_\Phi$ and $\mathbf{B}_\Phi$, respectively; $\mathbf{n}_\Phi$ represents Gaussian random noise; and $\begin{bmatrix} \mathbf{T}_\Phi & \mathbf{B}_\Phi \end{bmatrix}$ is a concatenated matrix of $\mathbf{T}_\Phi$ and $\mathbf{B}_\Phi$. Using a similar reasoning as described in the previous subsection, the GLRT of the hyperspectral detection problem depicted by the model in (6) is given by

$$\mathbf{L}_2(\Phi(\mathbf{y})) = \frac{\Phi(\mathbf{y})^T(\mathbf{P}_{\mathbf{I}_\Phi} - \mathbf{P}_{\mathbf{B}_\Phi})\Phi(\mathbf{y})}{\Phi(\mathbf{y})^T(\mathbf{P}_{\mathbf{I}_\Phi} - \mathbf{P}_{\mathbf{T}_\Phi\mathbf{B}_\Phi})\Phi(\mathbf{y})}, \tag{7}$$

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 $\mathcal{F}$

$$\begin{aligned} \mathbf{P}_{\mathbf{T}_\Phi\mathbf{B}_\Phi} &= \begin{bmatrix} \mathbf{T}_\Phi\,\mathbf{B}_\Phi \end{bmatrix} \begin{bmatrix} [\mathbf{T}_\Phi\,\mathbf{B}_\Phi]^T[\mathbf{T}_\Phi\,\mathbf{B}_\Phi] \end{bmatrix}^{-1} \times \\ &\quad \begin{bmatrix} \mathbf{T}_\Phi\,\mathbf{B}_\Phi \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{T}_\Phi\,\mathbf{B}_\Phi \end{bmatrix} \begin{bmatrix} \mathbf{T}_\Phi^T\mathbf{T}_\Phi & \mathbf{T}_\Phi^T\mathbf{B}_\Phi \\ \mathbf{B}_\Phi^T\mathbf{T}_\Phi & \mathbf{B}_\Phi^T\mathbf{B}_\Phi \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{T}_\Phi^T \\ \mathbf{B}_\Phi^T \end{bmatrix}. \end{aligned} \tag{8}$$

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

$$\begin{aligned} \Phi(\mathbf{y})^T(\mathbf{P}_{\mathbf{I}_\Phi} - \mathbf{P}_{\mathbf{B}_\Phi})\Phi(\mathbf{y}) = \Phi(\mathbf{y})^T\mathbf{P}_{\mathbf{I}_\Phi}\Phi(\mathbf{y}) - \\ \Phi(\mathbf{y})^T\mathbf{B}_\Phi\mathbf{B}_\Phi^T\Phi(\mathbf{y}). \end{aligned} \tag{9}$$

Each column of $\mathbf{B}_\Phi$ and $\mathbf{T}_\Phi$ can be written in terms of its corresponding data space [7] as

$$\mathbf{B}_\Phi = \begin{bmatrix} \mathbf{e_b}^1 & \mathbf{e_b}^2 & \dots & \mathbf{e_b}^{N_b} \end{bmatrix} = \Phi_{\mathbf{Z_B}}\mathcal{B}, \tag{10}$$

$$\mathbf{T}_\Phi = \begin{bmatrix} \mathbf{e_t}^1 & \mathbf{e_t}^2 & \dots & \mathbf{e_t}^{N_t} \end{bmatrix} = \Phi_{\mathbf{Z_T}}\mathcal{T}, \tag{11}$$

where $\mathbf{e}_\mathbf{b}^i$ and $\mathbf{e}_\mathbf{t}^j$ are the significant eigenvectors of the target and background covariance metrics $\mathbf{C}_{\mathbf{B}_\Phi}$ and $\mathbf{C}_{\mathbf{T}_\Phi}$, respectively; $\Phi_{\mathbf{Z_B}} = \left[\begin{array}{cccc} \Phi(\mathbf{y}_1) & \Phi(\mathbf{y}_2) & \ldots & \Phi(\mathbf{y}_M) \end{array}\right], \mathbf{y}_i \in \mathbf{Z_B}$ and $\Phi_{\mathbf{Z_T}} = \left[\begin{array}{cccc} \Phi(\mathbf{y}_1) & \Phi(\mathbf{y}_2) & \ldots & \Phi(\mathbf{y}_N) \end{array}\right]$, $\mathbf{y}_i \in \mathbf{Z_T}$; the column vectors of $\mathcal{B}$ and $\mathcal{T}$ represent only the significant normalized eigenvectors (β_1, β_2, $\ldots$, β_{N_b}) and ($\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\ldots$, $\boldsymbol{\alpha}_{N_t}$) 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 the target centered kernel matrix $\mathbf{K}(\mathbf{Z_T}, \mathbf{Z_T}) = (\mathbf{K})_{ij} = k(\mathbf{y}_i, \mathbf{y}_j)$, $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Z_T}$, respectively. Using (10) the projection of $\Phi(\mathbf{y})$ onto $\mathbf{B}_\Phi$ becomes $\mathbf{B}_\Phi^T \Phi(\mathbf{y}) = \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{y})$ and, similarly, using (11) the projection onto $\mathbf{T}_\Phi$ is $\mathbf{T}_\Phi^T \Phi(\mathbf{y}) = \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{y})$ 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 [7], 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

$$\Phi(\mathbf{y})^T \mathbf{B}_\Phi \mathbf{B}_\Phi^T \Phi(\mathbf{y}) = \mathbf{K}(\mathbf{Z_B}, \mathbf{y})^T \mathcal{B}\mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{y}). \quad (12)$$

The projection onto the identity operator $\Phi(\mathbf{y})^T \mathbf{P}_{\mathbf{I}_\Phi} \Phi(\mathbf{y})$ also needs to be kernelized which is given by

$$\begin{aligned} \Phi(\mathbf{y})^T \mathbf{P}_{\mathbf{I}_\Phi} \Phi(\mathbf{y}) &= \Phi(\mathbf{y})^T \Phi_{\mathbf{Z_{TB}}} \Delta\Delta^T \Phi_{\mathbf{Z_{TB}}}^T \Phi(\mathbf{y}) \quad (13) \\ &= \mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y})^T \Delta\Delta^T \mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y}), \end{aligned}$$

where $\Phi_{\mathbf{Z_{TB}}} = \Phi_{\mathbf{Z_T}} \cup \Phi_{\mathbf{Z_B}}$ and Δ is a matrix whose columns are the eigenvectors ($\boldsymbol{\kappa}_1$, $\boldsymbol{\kappa}_2, \ldots, \boldsymbol{\kappa}_{N_{bt}}$) of the centered kernel matrix $\mathbf{K}(\mathbf{Z_{TB}}, \mathbf{Z_{TB}}) = (\mathbf{K})_{ij} = k(\mathbf{y}_i, \mathbf{y}_j)$, $\mathbf{y}_i, \mathbf{y}_j \in \mathbf{Z_T} \cup \mathbf{Z_B}$ with nonzero eigenvalues, normalized by the square root of their associated eigenvalues and $\mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y})$ is a concatenated vector $\left[\begin{array}{cc} \mathbf{K}(\mathbf{Z_T}, \mathbf{y})^T & \mathbf{K}(\mathbf{Z_B}, \mathbf{y})^T \end{array}\right]^T$. To complete the kernelization process the denominator of (7) is given by

$$\begin{aligned} &\Phi(\mathbf{y})^T \mathbf{P}_{\mathbf{T}_\Phi \mathbf{B}_\Phi} \Phi(\mathbf{y}) = \Phi(\mathbf{y})^T \left[\begin{array}{cc} \mathbf{T}_\Phi & \mathbf{B}_\Phi \end{array}\right] \times \quad (14) \\ &\left[\begin{array}{cc} \mathbf{T}_\Phi^T \mathbf{T}_\Phi & \mathbf{T}_\Phi^T \mathbf{B}_\Phi \\ \mathbf{B}_\Phi^T \mathbf{T}_\Phi & \mathbf{B}_\Phi^T \mathbf{B}_\Phi \end{array}\right]^{-1} \left[\begin{array}{c} \mathbf{T}_\Phi^T \\ \mathbf{B}_\Phi^T \end{array}\right] \Phi(\mathbf{y}) \\ &= \left[\begin{array}{cc} \mathbf{K}(\mathbf{Z_T}, \mathbf{y})^T \mathcal{T} & \mathbf{K}(\mathbf{Z_B}, \mathbf{y})^T \mathcal{B} \end{array}\right] \times \\ &\left[\begin{array}{cc} \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{Z_T}) \mathcal{T} & \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{Z_B}) \mathcal{B} \\ \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{Z_T}) \mathcal{T} & \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{Z_B}) \mathcal{B} \end{array}\right]^{-1} \times \\ &\left[\begin{array}{c} \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{y}) \\ \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{y}) \end{array}\right]. \end{aligned}$$

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

$$\begin{aligned} \mathbf{L}_{2\mathbf{K}} = &(\mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y})^T \Delta\Delta^T \mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y}) - \quad (15) \\ &\mathbf{K}(\mathbf{Z_B}, \mathbf{y})^T \mathcal{B}\mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{y})) / \\ &(\mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y})^T \Delta\Delta^T \mathbf{K}(\mathbf{Z_{TB}}, \mathbf{y}) - \\ &\left[\begin{array}{cc} \mathbf{K}(\mathbf{Z_T}, \mathbf{y})^T \mathcal{T} & \mathbf{K}(\mathbf{Z_B}, \mathbf{y})^T \mathcal{B} \end{array}\right] \Lambda_1^{-1} \times \\ &\left[\begin{array}{c} \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{y}) \\ \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{y}) \end{array}\right]), \end{aligned}$$

where $\Lambda_1 = \left[\begin{array}{cc} \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{Z_T}) \mathcal{T} & \mathcal{T}^T \mathbf{K}(\mathbf{Z_T}, \mathbf{Z_B}) \mathcal{B} \\ \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{Z_T}) \mathcal{T} & \mathcal{B}^T \mathbf{K}(\mathbf{Z_B}, \mathbf{Z_B}) \mathcal{B} \end{array}\right]$.

In the above derivation (15) we assumed that the mapped input data was centered in the feature space by removing the sample mean. 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. The resulting centered $\hat{\mathbf{K}}$ is shown in [7] to be given by

$$\hat{\mathbf{K}} = (\mathbf{K} - \mathbf{1}_N \mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N \mathbf{K} \mathbf{1}_N), \quad (16)$$

where the $N \times N$ matrix $(\mathbf{1}_N)_{ij} = 1/N$. 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 also to be centered by removing their corresponding empirical kernel map mean. (e.g. $\hat{\mathbf{K}}(\mathbf{Z_T}, \mathbf{y}) = \mathbf{K}(\mathbf{Z_T}, \mathbf{y}) - \frac{1}{N} \sum_{i=1}^N k(\mathbf{y}_i, \mathbf{y}),\ \mathbf{y}_i \in \mathbf{Z_T}$.)

4 OSP and Kernel OSP Algorithms

4.1 Linear spectral mixture model

The OSP algorithm [2] is based on maximizing the SNR (signal-to-noise ratio) in the subspace orthogonal to the background subspace and only depends on the noise second-order statistics. It also does not provide any estimate of the abundance measure for the desired end member in the mixed pixel. A linear mixture model for pixel $\mathbf{y}$ consisting of p spectral bands is described by

$$\mathbf{y} = \mathbf{M}\alpha + \mathbf{n}, \quad (17)$$

where the $(p \times l)$ matrix $\mathbf{M}$ represent l endmembers spectra, α is a $(l \times 1)$ column vector whose elements are the coefficients that account for the proportions (abundances) of each endmember spectrum contributing to the mixed pixel, and $\mathbf{n}$ is an $(p \times 1)$ vector representing an additive zero-mean Gaussian noise with covariance matrix $\sigma^2 \mathbf{I}$ and $\mathbf{I}$ is the $(p \times p)$ identity matrix.

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 α in partition form as $\mathbf{M} = (\mathbf{U} : \mathbf{d})$ and

$\alpha = \begin{bmatrix} \gamma \\ \alpha_l \end{bmatrix}$ then model (17) can be rewritten as

$$\mathbf{r} = \mathbf{d}\alpha_l + \mathbf{B}\gamma + \mathbf{n}, \tag{18}$$

where the columns of $\mathbf{B}$ represent the undesired spectral signatures (background signatures or eigenvectors) and the column vector γ is the abundance measures for the undesired spectral signatures. The OSP operator that maximizes the signal to noise ratio is given by

$$\mathbf{q}_{OSP}^T = \mathbf{d}^T(\mathbf{I} - \mathbf{B}\mathbf{B}^{\#}) \tag{19}$$

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

$$D_{OSP} = \mathbf{q}_{OSP}^T\mathbf{r} = \mathbf{d}^T(\mathbf{I} - \mathbf{B}\mathbf{B}^{\#})\mathbf{y}. \tag{20}$$

4.2 OSP in feature space and its kernel version

The mixture model in the high dimensional feature space $\mathcal{F}$ is given by

$$\Phi(\mathbf{r}) = \mathbf{M}_\Phi\alpha + \mathbf{n}_\Phi, \tag{21}$$

where $\mathbf{M}_\Phi$ is a matrix whose columns are the endmembers spectra in the feature space and $\mathbf{n}_\Phi$ is an additive zero-mean Gaussian noise with covariance matrix $\sigma^2\mathbf{I}_\Phi$ and $\mathbf{I}_\Phi$ is the identity matrix in the feature space. The model (21) can also be rewritten as

$$\Phi(\mathbf{r}) = \Phi(\mathbf{d})\alpha_p + \mathbf{B}_\Phi\gamma + \mathbf{n}_\Phi, \tag{22}$$

where $\Phi(\mathbf{d})$ represent the spectral signature of the desired target in the feature space and the columns of $\mathbf{B}_\Phi$ represent the undesired background signatures in the feature space.

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

$$D_{OSP_\Phi} = \mathbf{q}_{OSP_\Phi}^T\mathbf{r} = \Phi(\mathbf{d})^T(\mathbf{I}_\Phi - \mathbf{B}_\Phi\mathbf{B}_\Phi^{\#})\Phi(\mathbf{r}). \tag{23}$$

This output (23) is very similar to the numerator of (7). It can easily be shown that the kernelized version of (23) is given by

$$D_{KOSP} = \mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{d})^T\Upsilon\Upsilon^T\mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{y}) - \mathbf{K}(\mathbf{Z}_B, \mathbf{d})^T\mathcal{B}\mathcal{B}^T\mathbf{K}(\mathbf{Z}_B, \mathbf{y}) \tag{24}$$

where $\mathbf{Z}_B = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_N]$ correspond to N input background spectral signatures and $\mathcal{B} = (\beta^1, \beta^2, \ldots, \beta^{N_b})^T$ 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 [8]. $\mathbf{K}(\mathbf{Z}_B, \mathbf{r})$ 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. $\mathbf{Z}_{Bd} = \mathbf{Z}_B \cup \mathbf{d}$ and Υ is a matrix whose columns are the N_{bd} eigenvectors ($\boldsymbol{v}_1,\ \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{N_{bd}}$) of the centered kernel matrix $\mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{Z}_{Bd}) = (\mathbf{K})_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{Z}_{Bd}$ with nonzero eigenvalues, normalized by the square root of their associated eigenvalues. Also $\mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{y})$ is the concatenated vector $\begin{bmatrix} \mathbf{K}(\mathbf{Z}_B, \mathbf{r})^T & \mathbf{K}(\mathbf{d}, \mathbf{y})^T \end{bmatrix}^T$ and $\mathbf{K}(\mathbf{Z}_{Bd}, \mathbf{d})$ is the concatenated vector $\begin{bmatrix} \mathbf{K}(\mathbf{Z}_B, \mathbf{d})^T & \mathbf{K}(\mathbf{d}, \mathbf{d})^T \end{bmatrix}^T$. In the above derivation (24) we assumed that the mapped input data was centered in the feature space. The kernel matrices and the empirical kernel maps have to be properly centered as was shown in subsection 3.2

5 Linear SMF and Kernel Spectral Matched Filter

5.1 Linear Spectral Matched Filter

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 $\mathbf{x} = [x(1), x(2), \ldots, x(p)]^T$ consisting of p spectral bands. We can model each spectral observation as a linear combination of the target spectral signature and noise

$$\mathbf{x} = a\mathbf{s} + \mathbf{n}, \tag{25}$$

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

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

$$\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_N]. \tag{26}$$

We can design a linear matched filter such that the desired target signal $\mathbf{s}$ is passed through while the average filter output energy is minimized. The solution to this minimization problem was shown in [12] and was called Constrained Energy Minimization (CEM) filter. The output of the linear matched filter for a test input $\mathbf{r}$, given the estimated covariance matrix is given by

$$y_\mathbf{r} = \mathbf{w}^T\mathbf{r} = \frac{\mathbf{s}^T\hat{\mathbf{C}}^{-1}\mathbf{r}}{\mathbf{s}^T\hat{\mathbf{C}}^{-1}\mathbf{s}} \tag{27}$$

where $\hat{\mathbf{C}}$ is the estimated covariance matrix. In [4, 5] it was shown that using the GLRT the same expression for the linear matched filter (27) can be obtained.

5.2 SMF in Feature Space and its Kernel Version

Consider the linear model of the input data in a kernel feature space which is equivalent to a non-linear model in the input space

$$\Phi(\mathbf{x}) = a_{\Phi}\Phi(\mathbf{s}) + \mathbf{n}_{\Phi}, \tag{28}$$

where Φ is the non-linear mapping that maps the input data into a kernel feature space, 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 $\mathbf{n}_{\Phi}$ contains the added noise in the feature space.

Using the constrained least squares approach it can easily be shown that the output of the desired matched filter for the input $\Phi(\mathbf{r})$ is given by

$$y_{\Phi(\mathbf{r})} = \frac{\Phi(\mathbf{s})^T\hat{\mathbf{C}}_{\Phi}^{-1}\Phi(\mathbf{r})}{\Phi(\mathbf{s})^T\hat{\mathbf{C}}_{\Phi}^{-1}\Phi(\mathbf{s})} \tag{29}$$

where $\hat{\mathbf{C}}_{\Phi}$ is the estimated covariance of pixels in the feature space.

We now show how to kernelize the matched filter expression (29) where the resulting non-linear matched filter is called the kernel matched filter. The pseudoinverse (inverse) of the estimated background covariance matrix can be written in terms of its eigenvector decomposition as [10]

$$\hat{\mathbf{C}}_{\Phi}^{\#} = \mathbf{X}_{\Phi}\boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{B}}^T\mathbf{X}_{\Phi}^T \tag{30}$$

where $\mathbf{X}_{\Phi}$=$[\Phi(\mathbf{x}_1)\ \Phi(\mathbf{x}_2)\ \ldots\ \Phi(\mathbf{x}_N)]$ is a matrix whose columns are the mapped background reference data in the feature space and $\mathcal{B} = [\beta^1\ \beta^2\ \ldots\ \beta^{N_b}]$ are the nonzero eigenvectors of the centered kernel matrix (Gram matrix) $\mathbf{K}(\mathbf{X},\mathbf{X})$ normalized by the square root of their corresponding eigenvalues.

Inserting Equation (30) into (29) it can be rewritten as

$$y_{\Phi(\mathbf{r})} = \frac{\Phi(\mathbf{s})^T\mathbf{X}_{\Phi}\boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{B}}^T\mathbf{X}_{\Phi}^T\Phi(\mathbf{r})}{\Phi(\mathbf{s})^T\mathbf{X}_{\Phi}\boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{B}}^T\mathbf{X}_{\Phi}^T\Phi(\mathbf{s})}. \tag{31}$$

Also using the properties of the Kernel PCA [7], we have the relationship $\mathbf{K}^{-1} = \boldsymbol{\mathcal{B}}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mathcal{B}}$. 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 $< \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) >$. Finally, the kernelized version of SMF is now given by

$$y_{\mathbf{K_r}} = \frac{\mathbf{K}(\mathbf{X},\mathbf{s})^T\mathbf{K}^{-1}\mathbf{K}(\mathbf{X},\mathbf{r})}{\mathbf{K}(\mathbf{X},\mathbf{s})^T\mathbf{K}^{-1}\mathbf{K}(\mathbf{X},\mathbf{s})} = \frac{\mathbf{K_s}^T\mathbf{K}^{-1}\mathbf{K_r}}{\mathbf{K_s}^T\mathbf{K}^{-1}\mathbf{K_s}} \tag{32}$$

where the empirical kernel maps $\mathbf{K_s} = \mathbf{K}(\mathbf{X},\mathbf{s})$ and $\mathbf{K_r} = \mathbf{K}(\mathbf{X},\mathbf{r})$. As in the previous section the kernel matrix $\mathbf{K}$ as well as the empirical kernel maps need to be properly centered.

6 Adaptive Subspace Detector and Kernel Adaptive Subspace Detector

6.1 Linear ASD

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 $\mathbf{x}$ (a pixel vector) is expressed as

$$\begin{aligned} &\mathbf{H}_0 : \mathbf{x} = \mathbf{n}, && \text{Target absent} \\ &\mathbf{H}_1 : \mathbf{x} = \mathbf{U}\theta + \sigma\mathbf{n}, && \text{Target present} \end{aligned} \tag{33}$$

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

In the model, $\mathbf{x}$ 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 $\mathcal{N}(\mathbf{U}\theta, \sigma^2\mathbf{C})$, 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 as described in [5] (so called ASD) is given by

$$D_{ASD}(\mathbf{x}) = \frac{\mathbf{x}^T\hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{x}}{\mathbf{x}^T\hat{\mathbf{C}}^{-1}\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}, \tag{34}$$

where $\hat{\mathbf{C}}$ is the MLE (maximum likelihood estimate) of the covariance $\mathbf{C}$ and η_{ASD} represents a threshold. Expression (34) has a constant false alarm rate (CFAR) property and is also referred to as the adaptive cosine estimator because (34) measures the angle between $\tilde{\mathbf{x}}$ and $< \tilde{\mathbf{U}} >$ 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 by assuming that the input data has been implicitly mapped by a nonlinear function Φ into a high dimensional feature space $\mathcal{F}$. The model in $\mathcal{F}$ is then given by

$$\begin{aligned} &\mathbf{H}_{0_{\Phi}} : \Phi(\mathbf{x}) = \mathbf{n}_{\Phi}, && \text{Target absent} \\ &\mathbf{H}_{1_{\Phi}} : \Phi(\mathbf{x}) = \mathbf{U}_{\Phi}\theta_{\Phi} + \sigma_{\Phi}\mathbf{n}_{\Phi}, && \text{Target present} \end{aligned} \tag{35}$$

where $\mathbf{U}_{\Phi}$ represents a full-rank matrix whose M_1 column vectors are the eigenvectors that span target subspace

$< \mathbf{U}_\Phi >$ in $\mathcal{F}$; θ_Φ is unknown vectors whose entries are coefficients that account for the abundances of the corresponding column vectors of $\mathbf{U}_\Phi$; $\mathbf{n}_\Phi$ represents Gaussian random noise distributed by $\mathcal{N}(0, \mathbf{C}_\Phi)$; and σ_Φ is the noise variance under $\mathbf{H}_{1_\Phi}$. The GLRT for the model (35) in $\mathcal{F}$ is now given by

$$D(\Phi(\mathbf{x})) = \frac{\Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi (\mathbf{U}_\Phi^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi)^{-1} \mathbf{U}_\Phi^T \hat{\mathbf{C}}_\Phi^{-1} \Phi(\mathbf{x})}{\Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \Phi(\mathbf{x})}, \tag{36}$$

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

The kernelized expression of (36) is given by

$$D_{KASD}(\mathbf{x}) = \frac{\mathbf{K_x}[\mathcal{T}^T \mathbf{K}(\mathbf{X},\mathbf{Y})^T \mathbf{K}(\mathbf{X},\mathbf{X})^{-1} \mathbf{K}(\mathbf{X},\mathbf{Y}) \mathcal{T}]^{-1} \mathbf{K_x}^T}{\mathbf{k}(\mathbf{x},\mathbf{X})^T \mathbf{K}(\mathbf{X},\mathbf{X})^{-1} \mathbf{k}(\mathbf{x},\mathbf{X})} \tag{37}$$

where $\mathbf{K_x} = \mathbf{k}(\mathbf{x},\mathbf{X})^T \mathbf{K}(\mathbf{X},\mathbf{X})^{-1} \mathbf{K}(\mathbf{X},\mathbf{Y}) \mathcal{T}$, background spectral signatures is denoted by $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_N]$, target spectral signatures are denoted by $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_M]$ and $\mathcal{T} = [\boldsymbol{\alpha}^1\ \boldsymbol{\alpha}^2\ \ldots\ \boldsymbol{\alpha}^{M_1}], M_1 < M$, is a matrix consisting of the M_1 eigenvectors of the kernel matrix $\mathbf{K}(\mathbf{Y},\mathbf{Y})$. As in the previous section, all the kernel matrices as well as the empirical kernel maps need to be properly centered [7].

7 Experimental Results

In this section, the kernel-based matched signal detectors, such as 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 a real hyperspectral image that contains military targets. The Gaussian RBF kernel, $k(\mathbf{x},\mathbf{y}) = \exp(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{c})$, was used to implement the kernel-based detectors. c represents the width of the Gaussian distribution and the value of c was chosen such that the overall data variations can be fully exploited by the Gaussian RBF function. In this paper, the values of c were determined experimentally.

A. Illustrative Toy Examples

Figs 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 Fig. 1 and nonlinearly mapped data in Fig. 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 Fig. 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 Fig. 1. All the data points in Fig. 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 (Fig. 1(a) and Fig. 2 (a)) represent only the numerator of Eq. 4 because the denominator becomes unstable for the two-dimensional cases: i.e., for the two-dimensional data $(\mathbf{I} - \mathbf{P_{TB}})$ becomes zero.

B. Hyperspectral Images

In this section, a HYDICE (HYperspectral Digital Imagery Collection Experiment) image from the Desert Radiance II data collection (DR-II) was used to compare detection performance between the kernel-based and conventional methods. The HYDICE imaging sensor generates 210 bands across the whole spectral range (0.4 – 2.5 μm) which includes the visible and short-wave infrared (SWIR) bands. But we only use 150 bands by discarding water absorption and low signal to noise ratio (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, as shown in the sample band images in Fig. 3. The detection performance of the DR-II image 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.

Figs. 4-5 show the detection results including the ROC curves generated by applying the kernel-based and conventional detectors to the DR-II image. 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 Fig. 4. Fig. 5 shows the ROC curve plots for the kernel-based and conventional detectors; the kernel-based detectors clearly outperformed the conventional de-

tectors. In particular, KMSD performed the best of all the 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 background kernel subspaces representing the target and background signals in the feature space, respectively.

8 Conclusions

In this paper, nonlinear 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 a real hyperspectral image. It is shown that the kernel-based nonlinear versions of these detectors outperform the linear versions. If enough target spectral samples are available to build a target subspace the kernel matched subspace detectors (KMSD and KASD) generally provide improved detection performance over KOSP and KSMF. If the target subspace cannot be properly estimated because only a small number of target samples are available KOSP or KSMF can be used instead which uses a single spectral signature as a reference to a target of interest.

References

[1] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[2] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 4, pp. 779–785, July 1994.

[3] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detector," in *Proc. SPIE*, April 2000, vol. 4049, pp. 2–17.

[4] F. C. Robey, D. R. Fuhrmann, and E. J. Kelly, "A CFAR adaptive matched filter detector," *IEEE Trans. on Aerospace and Elect. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[5] S Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.

[6] S Kraut, L. L. Scharf, and T McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[7] B Schökolpf and A. J. Smola, *Learning with Kernels*, The MIT Press, 2002.

[8] B Schökolpf, A. J. Smola, and K.-R. Müller, "Kernel principal component analysis," *Neural Computation*, vol. 10, pp. 1299–1319, 1998.

[9] G. Baudat and F Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 12, pp. 2385–2404, 2000.

[10] A. Ruiz and E. Lopez-de Teruel, "Nonlinear kernel-based statistical patten analysis," *IEEE Trans. on Neural Networks*, vol. 12, pp. 16–32, 2001.

[11] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm : A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[12] J. C. Harsanyi, *Detection and Classification of Subpixel Spectral Signatures in Hyperspectral Image Sequences*, Ph.D. dissertation, Dept. Elect. Eng., Univ. of Maryland, Baltimore County, 1993.

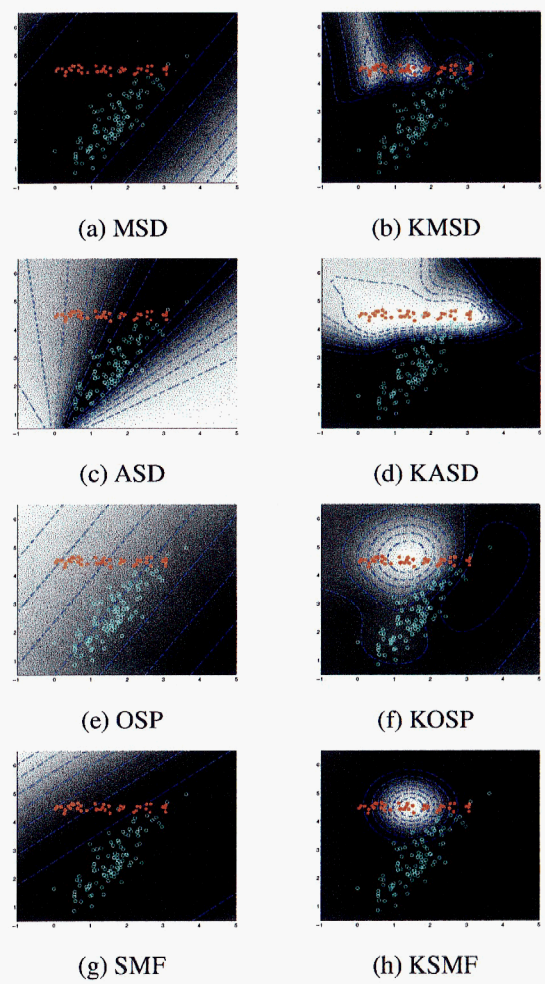

Figure 1: Contour and surface plots of the conventional matched signal detectors and their corresponding kernel versions on a toy dataset (a mixture of Gaussian).

1063-6919/05 $20.00 © 2005 **IEEE**

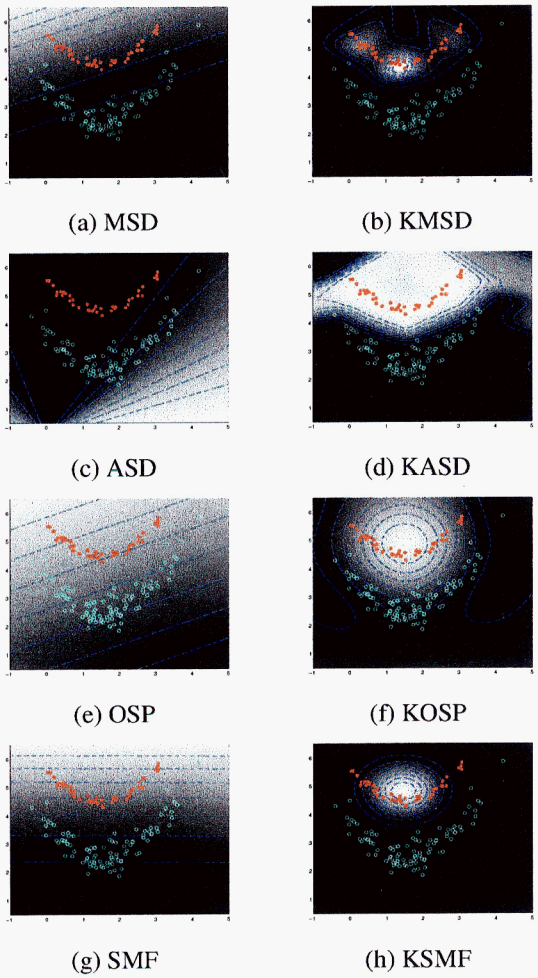

Figure 2: Contour and surface plots of the conventional matched signal detectors and their corresponding kernel versions on a toy dataset: in this toy example, the Gaussian mixture data shown in Fig. 1 was modified to generate nonlinearly mixed data.

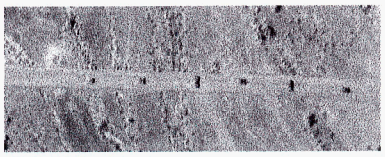

Figure 3: A sample band image from the DR-II data

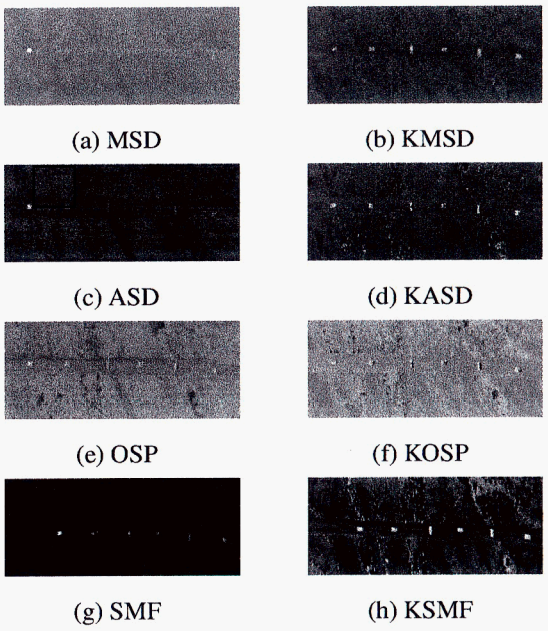

Figure 4: Detection results for the DR-II image using the conventional detectors and the corresponding kernel versions.

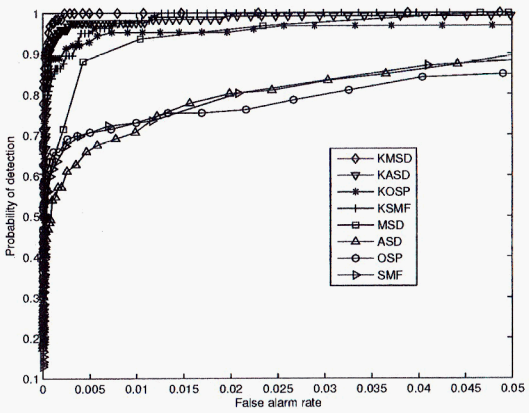

Figure 5: ROC curves obtained by conventional detectors and the corresponding kernel versions for the DR-II image.

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**