KERNEL ADAPTIVE SUBSPACE DETECTOR FOR HYPERSPECTRAL TARGET DETECTION

Heesung Kwon and Nasser M. Nasrabadi

US Army Research Laboratory
2800 Powder Mill Road, Adelphi, MD 20783

ABSTRACT

In this paper, we present a kernel-based nonlinear version of the adaptive subspace detector (ASD) that detects signals of interest in a high dimensional (possibly infinite) feature space associated with a certain nonlinear mapping. In order to address the high dimensionality of the feature space, ASD is first implicitly formulated in the feature space which is then converted into an expression in terms of kernel functions via the *kernel trick* of the Mercer kernels. The proposed kernel-based ASD (KASD) exploits the nonlinear correlations between the spectral bands that is ignored by the conventional ASD. Experimental results based on the given hyperspectral image show that the proposed KASD outperforms the conventional ASD.

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 detectors (MSDs) [1] have been widely used to detect a desired signal (target) represented by a subspace buried within a background noise whose covariance structure is assumed to be known. However, in reality, the noise covariance is normally unknown and has to be estimated based on training samples. In [2], adaptive versions of MSDs, referred to as adaptive subspace detectors (ASDs), have been introduced whose detector statistic uses the maximum likelihood estimate (MLE) of the noise covariance.

Matched signal detectors, such as spectral matched filter and matched subspace detectors (whether adaptive or nonadaptive), 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, we aim to develop a nonlinear version of ASD 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 [3]. A nonlinear mapping of input data into the high dimensional feature space often increases data separability and reduces the complexity of the corresponding data structure [4]. Nonlinear versions of a number of signal processing techniques such as principal component analysis (PCA) [5], clustering [4], matched subspace detectors [6], and anomaly detectors [7] have already been defined in a kernel feature space.

To obtain the derivation for the nonlinear ASD, ASD is first implicitly formulated in the feature space which in turn corresponds to a nonlinear version of ASD in the input space. However, the expression for ASD in the feature space is not tractable because of possibly infinite dimensionality. The ASD expression in the feature space has to be *kernelized* using the kernel trick property of the Mercer kernels [3] that converts the dot product of spectral vectors in the feature space into a kernel function – a nonlinear function defined in the input space. The final expression for the kernelized ASD is referred to as the kernel adaptive subspace detector (KASD), which is equivalent to a nonlinear ASD in the original input space.

This paper is organized as follows. Section 2 introduces ASD defined in the input space. In Section 3 we describe the ASD algorithm in the feature space and reformulate the the expression in terms of the kernel function using the kernel trick. Performance comparison between the KASD and the conventional ASD algorithms is provided in Section 4 and conclusions are given in Section 5.

2. ADAPTIVE SUBSPACE DETECTOR

In this section, the general likelihood ratio test (GLRT) under the two competing hypotheses ($\mathbf{H}_0$ and $\mathbf{H}_1$) for a certain mixed pixel problem is described. The subpixel detection model for a measurement $\mathbf{x}$ (a pixel vector) is expressed as

$$\begin{aligned} &\mathbf{H}_0 : \mathbf{x} = \mathbf{n}, && \text{Target absent} \\ &\mathbf{H}_1 : \mathbf{x} = \mathbf{U}\theta + \sigma\mathbf{n}, && \text{Target present} \end{aligned} \tag{1}$$

where $\mathbf{U}$ represents orthogonal matrices 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 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 described in [2] (so called ASD) is given by

$$D_{ASD}(\mathbf{x}) = \frac{\mathbf{x}^T\hat{\mathbf{C}}^{-1}\mathbf{U}(\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{U})^{-1}\mathbf{U}^T\hat{\mathbf{C}}^{-1}\mathbf{x}}{\mathbf{x}^T\hat{\mathbf{C}}^{-1}\mathbf{x}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{ASD}, \tag{2}$$

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

3. KERNEL ADAPTIVE SUBSPACE DETECTOR

3.1. Adaptive Subspace Detector in the Feature Space

We now define a new subpixel model by assuming the input data has been implicitly mapped by a nonlinear function Φ into a high dimensional feature space $\mathcal{F}$. The subpixel model in $\mathcal{F}$ is then given by

$$\begin{aligned} \mathbf{H}_{0_\Phi} &: \Phi(\mathbf{x}) = \mathbf{n}_\Phi, && \text{Target absent} \\ \mathbf{H}_{1_\Phi} &: \Phi(\mathbf{x}) = \mathbf{U}_\Phi \theta_\Phi + \sigma_\Phi \mathbf{n}_\Phi, && \text{Target present} \end{aligned} \tag{3}$$

where $\mathbf{U}_\Phi$ represents full-rank matrices 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 (3) in $\mathcal{F}$ is now given by

$$D(\Phi(\mathbf{x})) = \frac{\Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi (\mathbf{U}_\Phi^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi)^{-1} \mathbf{U}_\Phi^T \hat{\mathbf{C}}_\Phi^{-1} \Phi(\mathbf{x})}{\Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \Phi(\mathbf{x})}, \tag{4}$$

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

Every term in (4) is in $\mathcal{F}$, meaning that the calculation of (4) is not feasible. The expression (4) has to be converted in term of kernels using the kernel trick to obtain an expression that can easily be calculated.

3.2. Kernel Methods 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

$$\Phi : \mathcal{X} \to \mathcal{F}, \mathbf{x} \mapsto \Phi(\mathbf{x}), \tag{5}$$

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 directly in the feature space. However, kernel-based learning algorithms use an effective kernel trick given by Eq. (6) to implement dot products in feature space by employing kernel functions [3]. 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

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j), \tag{6}$$

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 [3] 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.

3.3. Kernelization of Adaptive Subspace Detector in the Feature Space

In this subsection we show how to kernelize the ASD expression (4) in the feature space. The MLE of the background covariance matrix can be represented by its eigenvector decomposition or so called spectral decomposition [8] given by

$$\hat{\mathbf{C}}_\Phi = \mathbf{V}_\Phi \Lambda \mathbf{V}_\Phi{}^T, \tag{7}$$

where Λ is a diagonal matrix consisting of the eigenvalues and $\mathbf{V}_\Phi$ is a matrix whose columns are the eigenvectors of $\hat{\mathbf{C}}_\Phi$ in the feature space. The eigenvector matrix is represented by

$$\mathbf{V}_\Phi = [\mathbf{v}_\Phi^1 \ \mathbf{v}_\Phi^2 \ \dots \ \mathbf{v}_\Phi^N], \tag{8}$$

where N is the maximum number of eigenvectors with non-zero eigenvalue.

The pseudoinverse of the estimated background covariance matrix can also be written in terms of its eigenvector decomposition [8] as

$$\hat{\mathbf{C}}_\Phi^\# = \mathbf{V}_\Phi \Lambda^{-1} \mathbf{V}_\Phi{}^T. \tag{9}$$

Each eigenvector $\mathbf{v}_\Phi^j$ in the feature space, as shown in [5], can be expressed as a linear combination of the input reference vectors in the feature space as shown by

$$\mathbf{v}_\Phi^j = \sum_{i=1}^{N} \beta_i^j \Phi(\mathbf{x}_i) = \mathbf{X}_\Phi \boldsymbol{\beta}^{\boldsymbol{j}}, \tag{10}$$

where $\mathbf{X}_\Phi = [\Phi_c(\mathbf{x}_1) \ \Phi_c(\mathbf{x}_2) \dots \Phi_c(\mathbf{x}_N)]$ are the mean-removed (centered) vectors in the feature space corresponding to the N independent background spectral signatures $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]$ and $\boldsymbol{\beta}^{\boldsymbol{j}} = (\beta_1^j, \beta_2^j, \dots, \beta_N^j)^T$, $j = 1, \dots, N_1$, $N_1 \leq N$, the expansion vector coefficients, are the nonzero eigenvectors of the centered kernel matrix (Gram matrix) $\mathbf{K}_b(\mathbf{X}, \mathbf{X}) = (\mathbf{K}_b)_{ij}$, an $N \times N$ matrix whose entries are the dot products $k(\mathbf{x}_i, \mathbf{x}_j) =< \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) >$ for $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, normalized by the square root of their corresponding eigenvalues [5]. For all the eigenvectors $\mathbf{V}_\Phi$ in the feature space

$$\mathbf{V}_\Phi = \mathbf{X}_\Phi \mathcal{B}, \tag{11}$$

where $\mathcal{B} = [\boldsymbol{\beta}^1 \ \boldsymbol{\beta}^2 \ \dots \ \boldsymbol{\beta}^{N_1}]$. Similarly, $\mathbf{U}_\Phi$ is given by

$$\mathbf{U}_\Phi = \mathbf{Y}_\Phi \mathcal{T}, \tag{12}$$

where $\mathbf{Y}_\Phi = [\Phi_c(\mathbf{y}_1) \ \Phi_c(\mathbf{y}_2) \dots \Phi_c(\mathbf{y}_M)]$ are the mean-removed (centered) vectors in the feature space corresponding to the M independent target spectral signatures $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_M]$ and $\mathcal{T} = [\boldsymbol{\alpha}^1 \ \boldsymbol{\alpha}^2 \ \dots \ \boldsymbol{\alpha}^{M_1}]$, $M_1 < M$, is a matrix consisting of the M_1 eigenvectors of the kernel matrix $\mathbf{K}(\mathbf{Y}, \mathbf{Y})$.

Now, the term $\Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi$ in the numerator of (4) becomes

$$\begin{aligned} \Phi(\mathbf{x})^T \hat{\mathbf{C}}_\Phi^{-1} \mathbf{U}_\Phi &= \Phi(\mathbf{x})^T \mathbf{V}_\Phi \Lambda^{-1} \mathbf{V}_\Phi{}^T \mathbf{Y}_\Phi \mathcal{T} \\ &= \Phi(\mathbf{x})^T \mathbf{X}_\Phi \mathcal{B} \Lambda^{-1} \mathcal{B}^T \mathbf{X}_\Phi{}^T \mathbf{Y}_\Phi \mathcal{T} \\ &= \mathbf{k}(\mathbf{x}, \mathbf{X})^T \mathbf{K}_b^{-1} \mathbf{K}(\mathbf{X}, \mathbf{Y}) \mathcal{T} \equiv \mathbf{K}_\mathbf{x}, \end{aligned} \tag{13}$$

where $\mathcal{B}\Lambda^{-1}\mathcal{B}^T$ is substituted by $\mathbf{K}_b^{-1}$ using (29), as shown in Appendix I. The dot product term $\Phi(\mathbf{x})^T \mathbf{X}_\Phi$ in (13) (referred

to as an empirical kernel map [3]) is expressed as $\Phi(\mathbf{x})^T\mathbf{X}_\Phi = (\mathbf{k}(\mathbf{x},\mathbf{x}_1),\mathbf{k}(\mathbf{x},\mathbf{x}_2),\ldots,\mathbf{k}(\mathbf{x},\mathbf{x}_N) = \mathbf{k}(\mathbf{x},\mathbf{X})^T$. Similarly,

$$\mathbf{U}_\Phi{}^T\hat{\mathbf{C}}_\Phi^{-1}\Phi(\mathbf{x}) = \mathcal{T}^T\mathbf{K}(\mathbf{X},\mathbf{Y})^T\mathbf{K}_b^{-1}\mathbf{k}(\mathbf{x},\mathbf{X}) = \mathbf{K}_\mathbf{x}^T. \quad (14)$$

Also,

$$\mathbf{U}_\Phi{}^T\hat{\mathbf{C}}_\Phi^{-1}\mathbf{U}_\Phi = \mathcal{T}^T\mathbf{K}(\mathbf{X},\mathbf{Y})^T\mathbf{K}_b^{-1}\mathbf{K}(\mathbf{X},\mathbf{Y})\mathcal{T}. \quad (15)$$

The denominator of (4) is expressed as

$$\Phi(\mathbf{x})^T\hat{\mathbf{C}}_\Phi^{-1}\Phi(\mathbf{x}) = \mathbf{k}(\mathbf{x},\mathbf{X})^T\mathbf{K}_b^{-1}\mathbf{k}(\mathbf{x},\mathbf{X}). \quad (16)$$

Finally, the kernelized expression of (4) is given by

$$D_{KASD}(\mathbf{x}) = \frac{\mathbf{K}_\mathbf{x}[\mathcal{T}^T\mathbf{K}(\mathbf{X},\mathbf{Y})^T\mathbf{K}_b^{-1}\mathbf{K}(\mathbf{X},\mathbf{Y})\mathcal{T}]^{-1}\mathbf{K}_\mathbf{x}^T}{\mathbf{k}(\mathbf{x},\mathbf{X})^T\mathbf{K}_b^{-1}\mathbf{k}(\mathbf{x},\mathbf{X})}. \quad (17)$$

If the original data is not centered, then the estimated mean in the feature space can not be explicitly computed, therefore, the kernel matrices have to be properly centered. As shown in [3], the centered Gram matrix $\hat{\mathbf{K}}_b$ can be obtained from the uncentered Gram Matrix $\mathbf{K}_b$ by

$$\hat{\mathbf{K}}_b = (\mathbf{K}_b - \mathbf{1}_N\mathbf{K}_b - \mathbf{K}_b\mathbf{1}_N + \mathbf{1}_N\mathbf{K}_b\mathbf{1}_N), \quad (18)$$

where $(\mathbf{1}_N)_{ij} = 1/N$ is an $N \times N$ matrix. The Gram matrix $\mathbf{K}(\mathbf{X},\mathbf{Y})$ needs also to be centered according to (18). Similarly, the vector $\mathbf{k}(\mathbf{x},\mathbf{X})$ is centered by removing its corresponding empirical kernel map mean ($\hat{\mathbf{k}}(\mathbf{x},\mathbf{X}) = \mathbf{k}(\mathbf{x},\mathbf{X}) - \frac{1}{N}\sum_{i=1}^N k(\mathbf{x},\mathbf{x}_i)$, $\mathbf{x}_i \in \mathbf{X}$).

4. EXPERIMENTAL RESULTS

In this section, we implemented both the proposed kernel ASD detector (KASD) described by (17) and the conventional ASD detector (ASD) described by (2) to detect targets of interest (military vehicles) in the HYDICE (HYperspectral Digital Imagery Collection Experiment) images. 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.

(a)

Fig. 1. Sample band (48th) from the Desert Radiance II image.

The HYDICE image from the Desert Radiance II data collection was used to test both the kernel-based and conventional ASD detectors. The Desert Radiance II (DR-II) image contains 6 targets located in the dirt road, as shown in the sample band in Fig. 1.

The Gaussian RBF kernel, $k(\mathbf{x},\mathbf{y}) = \exp(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{c})$, was used to implement KASD. 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 value of c was determined experimentally and was set to 5.

The Gaussian RBF kernel was chosen based on two particular reasons; first, it is a *translation invariant* kernel and second, its associated non-linear map is smooth. Translation invariant kernels can normally provide robust detection performance even when spectral signatures in a given hyperspectral data set were subject to irregular illumination, because it only depends on the difference between $\mathbf{x}$ and $\mathbf{y}$, not the absolute positions of individual spectral vectors. The smooth non-linear mapping Φ associated with the Gaussian RBF kernel implies that the topographic ordering of the data in the input space is preserved in the feature space after the nonlinear mapping [4]. The mapped data in the feature space also occupies a small subspace of the feature space where data belonging to different classes may be separated by a larger degree than in the input space [3], thus making a target detection task in the feature space associated with the Gaussian RBF kernel more effective.

(a)

(b)

Fig. 2. Detector output for the Desert Radiance II image using KASD and ASD. (a) KASD and (b) ASD.

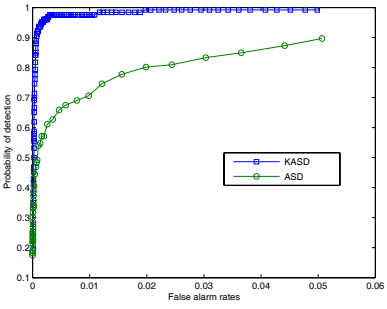

Fig. 3. ROC curves obtained by KASD and ASD for the Desert Radiance II image.

In implementing KASD and ASD the background samples were obtained from outside the test images to estimate the background subspace. Due to a lack of available target samples, the target samples from the HYDICE test set were used: the right most target in the DR-II image, as shown in Fig. 1, was used to generate the target subspace for all the targets in the test image. Figs.

2-3 show the detection results including the ROC curves generated by applying KASD and ASD to the DR-II image. KASD showed significantly improved target detection performance over ASD, as shown by the ROC curves in Fig 3.

5. CONCLUSIONS

We have presented a kernel realization of the nonlinear adaptive subspace detector by kernelizing the corresponding ASD expression defined in a high dimensional feature space. The complete kernelization procedure for the ASD expression was derived. Kernel ASD, was applied to hyperspectral subpixel target detection and its performance was compared to the conventional ASD. The detection results based on the given HYDICE image confirmed that the kernel ASD outperforms the conventional ASD.

6. APPENDIX I. (KERNEL PCA)

In this Appendix, we present derivation of Kernel PCA and its properties providing the relationship between the covariance matrix and the corresponding Gram matrix. Our goal is to prove $\mathbf{K}_b^{-1} = \mathcal{B}\Lambda^{-1}\mathcal{B}^T$. 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

$$\hat{\mathbf{C}}_\Phi = \mathbf{X}_\Phi \mathbf{X}_\Phi^T. \tag{19}$$

The PCA eigenvectors are computed by solving the eigenvalue problem

$$\begin{aligned} \lambda \mathbf{v}_\Phi &= \hat{\mathbf{C}}_\Phi \mathbf{v}_\Phi \\ &= \frac{1}{N}\sum_{i=1}^{N} \Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^T \mathbf{v}_\Phi \\ &= \frac{1}{N}\sum_{i=1}^{N} < \Phi(\mathbf{x}_i), \mathbf{v}_\Phi > \Phi(\mathbf{x}_i), \end{aligned} \tag{20}$$

where $\mathbf{v}_\Phi$ is an eigenvector in $\mathcal{F}$ with a corresponding nonzero eigenvalue λ. Substituting (10) into (20) and multiplying with $\Phi(\mathbf{x}_n)^T$, $n = 1, \ldots, N$, yields

$$\begin{aligned} &\lambda \sum_{i=1}^{N} \beta_i < \Phi(\mathbf{x}_n), \Phi(\mathbf{x}_i) > \\ &= \frac{1}{N}\sum_{i=1}^{N} \beta_i \Phi(\mathbf{x}_n)\Phi(\mathbf{x}_i)\Phi(\mathbf{x}_i)^T \sum_{i=1}^{N} \Phi(\mathbf{x}_i) \\ &= \frac{1}{N}\sum_{i=1}^{N} \beta_i < \Phi(\mathbf{x}_n), \sum_{j=1}^{N} \Phi(\mathbf{x}_j) < \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i) >>, \end{aligned} \tag{21}$$

for all $n = 1, \ldots, N$.

Now (21) can be written as

$$N\lambda \mathbf{K}_b \boldsymbol{\beta} = \mathbf{K}_b^2 \boldsymbol{\beta}, \tag{22}$$

where $\mathbf{K}_b = (\mathbf{K}_b)_{ij}$ the $N \times N$ kernel matrix (Gram matrix) whose entries are the dot products $< \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) >$. The solutions of (22) are obtained by solving the following eigenvalue problem

$$N\lambda \boldsymbol{\beta} = \mathbf{K}_b \boldsymbol{\beta}, \tag{23}$$

where $\boldsymbol{\beta}$ are the eigenvectors with nonzero eigenvalues of the kernel matrix $\mathbf{K}_b$. Note that $\boldsymbol{\beta}$ need to be normalized by the square root of their corresponding eigenvalues.

From the definition of PCA in the feature space (20) the estimated background covariance matrix is decomposed as

$$\hat{\mathbf{C}}_\Phi = \mathbf{V}_\Phi \Lambda {\mathbf{V}_\Phi}^T, \tag{24}$$

where $\mathbf{V}_\Phi = [\mathbf{v}_\Phi^1\ \mathbf{v}_\Phi^2\ \ldots\ \mathbf{v}_\Phi^N]$ and Λ is a diagonal matrix with its diagonal elements being the eigenvalues of $\hat{\mathbf{C}}_\Phi$. Similarly, from the the Kernel PCA (23) the kernel matrix eigen decomposition is given by

$$\mathbf{K}_b = \mathcal{B}\Omega\mathcal{B}^T, \tag{25}$$

where $\mathcal{B} = [\boldsymbol{\beta}^1\ \boldsymbol{\beta}^2\ \ldots\ \boldsymbol{\beta}^N]$ are the eigenvectors of the kernel matrix and Ω is a diagonal matrix with diagonal values equal to the eigenvalues of the kernel matrix $\mathbf{K}_b$. Then, the pseudoinverse background covariance matrix $\hat{\mathbf{C}}^\#$ and inverse Gram matrix $\mathbf{K}_b^{-1}$ can also be written as

$$\hat{\mathbf{C}}_\Phi^\# = \mathbf{V}_\Phi \Lambda^{-1} {\mathbf{V}_\Phi}^T \tag{26}$$

and

$$\mathbf{K}_b^{-1} = \mathcal{B}\Omega^{-1}\mathcal{B}^T, \tag{27}$$

respectively. The eigenvalues of the covariance matrix in the feature space and the eigenvalues of the kernel matrix are related by

$$\Lambda = \frac{1}{N}\Omega. \tag{28}$$

Substituting (28) into (27) we obtain the relationship

$$\mathbf{K}_b^{-1} = \frac{1}{N}\mathcal{B}\Lambda^{-1}\mathcal{B}^T, \tag{29}$$

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

7. 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] 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.

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

[4] Mark Girolami, "Mercer kernel-based clustering in feature space," *IEEE Trans. on Neural Networks*, vol. 13, no. 3, pp. 780–784, 2002.

[5] B Schökolpf, A. J. Smola, and K.-R. Müller, "Kernel principal component analysis," *Neural Computation*, , no. 10, pp. 1299–1319, 1999.

[6] H. Kwon and N. M. Nasrabadi, "Kernel-based subpixel target detection in hyperspectral images," in *In Proc. of IEEE Joint Conference on Neural Networks*, Budapest, Hungary, July 2004, pp. 717–722.

[7] 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, Feb. 2005.

[8] G. Strang, *Linear algebra and its applications*, Harcourt Brace & Company, 1986.