2009 Ninth International Conference on Intelligent Systems Design and Applications

Semi-supervised kernel target detection in hyperspectral images

Luca Capobianco, Andrea Garzelli
Dipartimento di Ingegneria dell'Informazione
Università di Siena
Siena, Italy
andrea.garzelli@unisi.it

Gustavo Camps-Valls
Departament d'Enginyeria Electrònica
Universitat de València
Burjassot (València) Spain
gustavo.camps@uv.es

***Abstract*—A semi-supervised graph-based approach to target detection is presented. The proposed method improves the Kernel Orthogonal Subspace Projection (KOSP) by deforming the kernel through the approximation of the marginal distribution using the unlabeled samples. The good performance of the proposed method is illustrated in a hyperspectral image target detection application for thermal hot spot detection. An improvement is observed with respect to the linear and the non-linear kernel-based OSP, demonstrating good generalization capabilities when low number of labeled samples are available, which is usually the case in target detection problems.**

***Keywords*-Machine learning; hyperspectral images; target detection**

I. Introduction

Target detection from hyperspectral data is of great interest in many applications. The goal of target detection, which generally assumes that the target spectral signature is known (or available from spectral libraries), is to detect pixels that match the target. Detecting targets in remote sensing images is typically described as a two-steps methodology, in which first an anomaly detector identifies spectral anomalies, and second a classifier is aimed at identifying whether or not the anomaly is a target or natural clutter. This step is only possible if the target spectral signature is known, which can be obtained from a spectral library or by using a spectral subspace matched filter learnt from the data. Several techniques have been proposed in the literature, such as the Reed-Xiaoli anomaly detector [1], the orthogonal subspace projection (OSP) [2], the Gaussian mixture model [3], the cluster-based detector [3], or the signal subspace processor [4].

All these techniques assume a parametric (linear or Gaussian mixture) model. Even though linearity is a convenient assumption, it is far from being realistic. Nonlinearity appears in different forms in remote sensing data [5], [6]: 1) the nonlinear scatter described in the bidirectional reflectance distribution function (BRDF); 2) the variable presence of water in pixels as a function of position in the landscape; 3) multiscattering and heterogeneities at subpixel level; and 4) atmospheric and geometric corrections that need to be done to deliver useful image products. These facts need to be encoded in the method either by allowing flexible non-linear mappings or by accurate description of the data manifold coordinates. In fact, the nonlinear relationships between different spectral bands within the target or clutter spectral signature need to be exploited in order to better distinguish between target and background. However, most of the target detection algorithms are based on linear matched (subspace) filters where the spectral characteristics of a target or a target subspace representing target information is assumed to be known.

Practical experience has shown that the design of a good classifier requires a sufficient amount of training data for each background class and that, usually, in operational remote sensing applications, a small number of labeled samples is typically available. Besides, in target detection applications, the main objective is to search a specific material (*target*), usually within very small number of pixels.

Conceptually, target detection can be viewed as a binary hypothesis testing problem, where each pixel is assigned a target or non-target label. In this way, target detection problems can be casted as *ill-posed* classification problems, thus being in the middle between canonical classification and target detection framework. For multiple targets, detection can be approached as a multiclass pattern recognition problems in which each class appears as a target 'lying' in the background. In the last years, many detection algorithms based on spectral matched (subspace) filters have been reformulated under the kernel methods framework: matched subspace detector (MSD), orthogonal subspace detector (OSD), spectral matched filter (SMF), and adaptive subspace detectors (ASD) [7]. Certainly, the use of kernel methods offers many advantages: they combat the high dimensionality problem in hyperspectral images, make the method robust to noise, and allow for flexible non-linear mappings with controlled (regularized) complexity [8]. Kernel methods in general, and kernel detectors in particular, rely on the proper definition of a kernel (or similarity) matrix among samples. So far, however, standard 'ad hoc' RBF or polynomial kernels have been used, and no attention has been payed to model the data marginal distribution that would be potentially helpful to design the

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.121

kernel structural form. In addition, kernel detectors have only considered labeled information, and unlabeled samples from guard windows are merely used as contrast density in hypothesis testing.

In this paper, we present a semi-supervised technique that collectively incorporates labeled and unlabeled data in the target detection framework. In semi-supervised learning (SSL) [9], [10], the algorithm is provided with some available *labeled* information in addition to the *unlabeled* information, thus allowing to encode some knowledge about the geometry and the shape of the dataset. This idea of exploring the shape of the marginal distribution in the dataset can be applied in kernel target detection in order to deform the 'measure' of distance in the kernel space according to the geometry of the neighboring pixels. We propose a Semi-Supervised Kernel Orthogonal Subspace Projection (S^2KOSP), which introduces an additional regularization term on the geometry of both labeled and unlabeled samples by using the *graph Laplacian* [9].

II. Kernel Orthogonal Subspace Projection Algorithm

This section reviews the orthogonal subspace projection (OSP) method and the non-linear kernel-based version proposed in [7].

In the standard formulation of the OSP algorithm [2], a linear mixing is assumed to model each B-bands pixel $\mathbf{r}$ as follows:

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}, \tag{1}$$

where $\mathbf{M}$ is the matrix of size $(B \times p)$ containing the p endmembers contributing to the mixed pixel $\mathbf{r}$, $\boldsymbol{\alpha}$ is a $(p \times 1)$ column vector of the coefficients that account for the spectral abundance of each endmember, and $\mathbf{n}$ stands for an additive zero mean Gaussian noise vector.

In order to identify one particular signature in the images, and given its spectral signature $\mathbf{d}$ with corresponding abundance measurements α_p, the above expression can be organized by rewriting the $\mathbf{M}$ matrix in two submatrices $\mathbf{M} = (\mathbf{U} : \mathbf{d})$, so that

$$\mathbf{r} = \mathbf{d}\alpha_p + \mathbf{U}\boldsymbol{\gamma} + \mathbf{n}. \tag{2}$$

The columns of $\mathbf{U}$ represent the undesired spectral signatures (background), while the $\boldsymbol{\gamma}$ represents the abundance for the undesired spectral signatures.

The effect of the OSP algorithm on the data set can be summarized in two steps. First, an annihilating operator rejects the background signatures for each pixel, so that only the desired signature should remain in the spectral component of the data. This operator is given by the $(B \times B)$ matrix $\mathbf{P}_U^{\perp} = \mathbf{I} - \mathbf{U}\mathbf{U}^{\#}$, where $\mathbf{U}^{\#}$ is the right Moore-Penrose pseudoinverse of $\mathbf{U}$. The second step of the OSP algorithm is represented by the filter $\mathbf{w}$ that maximizes the SNR of the filter output, that is the matched filter $\mathbf{w} = k\mathbf{d}$ where k is a constant [11], [12].

The OSP operator is given by $\mathbf{q}_{OSP}^{\top} = \mathbf{d}^{\top}\mathbf{P}_U^{\perp}$, and the output of the OSP classifier is:

$$\mathbf{D}_{OSP} = \mathbf{q}_{OSP}^{\top}\mathbf{r} = \mathbf{d}^{\top}\mathbf{P}_U^{\perp}\mathbf{r}. \tag{3}$$

By using the singular value decomposition (SVD) of $\mathbf{U} = \mathbf{B}\boldsymbol{\Sigma}\mathbf{A}^{\top}$, the annihilating operator becomes $\mathbf{P}_U^{\perp} = \mathbf{I} - \mathbf{B}\mathbf{B}^{\top}$, where the columns of $\mathbf{B}$ are obtained from the eigenvectors of the covariance matrix of the background spectral samples. If a given algorithm can be expressed in the form of dot products in the input space, its kernel version only needs the dot products among mapped samples.

Kernel methods compute the similarity between training samples $S = \{\mathbf{x}_i\}_{i=1}^{n}$ using pair-wise inner products between mapped samples, and thus the so-called kernel matrix $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \langle \boldsymbol{\Phi}(\mathbf{x}_i), \boldsymbol{\Phi}(\mathbf{x}_j) \rangle$ contains all the necessary information to perform many classical linear algorithms in the feature space.

The *kernelized* version of the output of the OSP classifier is given by [12]

$$\begin{aligned} \mathbf{D}_{KOSP} \;=\; & \mathbf{K}(\mathbf{X}_{bd}, \mathbf{d})^{\top}\Upsilon\Upsilon^{\top}\mathbf{K}(\mathbf{X}_{bd}, \mathbf{r}) \\ & -\mathbf{K}(\mathbf{X}_b, \mathbf{d})^{\top}\mathcal{B}\mathcal{B}^{\top}\mathbf{K}(\mathbf{X}_b, \mathbf{r}), \end{aligned} \tag{4}$$

where $\mathbf{K}(\mathbf{X}_b, \mathbf{r})$ and $\mathbf{K}(\mathbf{X}_b, \mathbf{d})$, referred to as the empirical kernel maps in the machine learning literature, are column vectors whose entries are $K(\mathbf{x}_i, \mathbf{r})$ and $K(\mathbf{x}_i, \mathbf{d})$ for $\mathbf{x}_i \in \mathbf{X}_b$ ($\mathbf{x}_i \in \mathbb{R}^B$), $i = 1, ..., l$, being l the number of labeled samples; $\mathcal{B}$ is the matrix containing the eigenvectors $\boldsymbol{\beta}^j$ described above; and Υ is the matrix containing the eigenvectors $\boldsymbol{\upsilon}^j$, similar to $\boldsymbol{\beta}^j$, but obtained from the centered kernel matrix $\mathbf{K}(\mathbf{X}_{bd}, \mathbf{X}_{bd})$, where $\mathbf{X}_{bd} = \mathbf{X}_b \cup \mathbf{d}$.

III. Proposed Semi-Supervised KOSP

In this section, we pay attention to the appropriate definition of the kernel under semi-supervised criteria. Essentially, we propose to deform the kernel using the graph Laplacian. This idea, which was originally presented in [13] for inductive SVM, has been recently presented for one-class classification [14] and support vector regression [15]. Here, it is extended to the kernel OSP.

Among SSL algorithms, we focus on *Graph-based methods*, in which each sample spreads its label information to its neighbors until a stable state is achieved on the whole dataset [16], [17]. In this paper, we propose a graph-based method with a kernel adapted to the geometry of the data marginal distribution. Essentially, we aim at exploiting the relation between labeled and unlabeled samples through the construction of a graph representation, where the vertices are the (labeled and unlabeled) samples, and the edges represent the similarity among samples in the dataset.

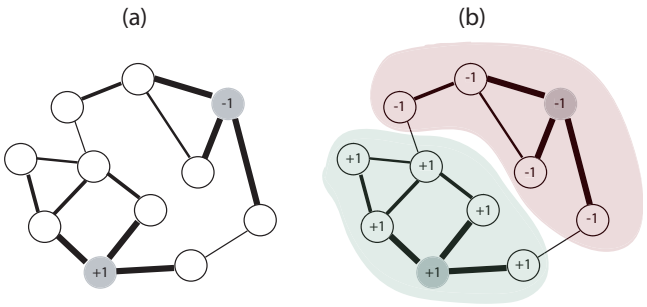

Figure 1. Graph classification on a toy graph.

A. Deforming the Kernel with the Graph Laplacian

In order to include the geometry of the data distribution as a similarity measure in the model, let us first define a linear space $\mathcal{V}$ with a positive semi-definite inner product, and let

$$\mathcal{S} : \mathcal{H} \rightarrow \mathcal{V}$$

be a bounded linear operator. We can now define $\widetilde{\mathcal{H}}$ to be the space of functions from $\mathcal{H}$ with the modified inner product $\langle f, g\rangle_{\widetilde{\mathcal{H}}} = \langle f, g\rangle_{\mathcal{H}} + \langle Sf, Sg\rangle_{\mathcal{V}}$. Semi-norm $\|S(\mathbf{f})\|^2 = \mathbf{f}^\top \mathbf{N}\mathbf{f}$ is given by a symmetric semi-definite matrix $\mathbf{N}$ and the decision function $\mathbf{f}$ is given by

$$\mathbf{f} = sgn(\mathbf{D}_{KOSP} - \theta),$$

where θ is the decision threshold. If u is the number of unlabeled samples available, the explicit form of the corresponding reproducing kernel function $\tilde{K}(\mathbf{x}_i, \mathbf{x}_j)$ can be explicitly defined over labeled samples as [13]:

$$\tilde{K}(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) - \mathbf{K}_{\mathbf{x}_i}^\top (\mathbf{I} + \mathbf{N}\mathbf{K})^{-1}\mathbf{N}\mathbf{K}_{\mathbf{x}_j}, \quad (5)$$

where $i, j \in \{1, ..., l + u\}$; $\mathbf{K}$ is the (complete) kernel matrix; $\mathbf{I}$ is the identity matrix; and $\mathbf{K}_{\mathbf{x}_i} = [K(\mathbf{x}_1, \mathbf{x}_i), \ldots, K(\mathbf{x}_{l+u}, \mathbf{x}_i)]^\top$.

In Fig.1(a), the two shaded circles are the initially labeled vertices (± 1), while the white nodes represent unlabeled samples. The thickness of the edges represent the similarity among samples, easily computable with a proper distance measure. In Fig.1(b), undirected graph methods classify the unlabeled samples according to the weighted distance, not just to the shortest path lengths, the latter leading to incorrectly classified samples. The two clusters (shaded) are intuitively correct, even being connected by (thin weak) edges.

In this formulation the geometry of the data is included through $\mathbf{N}$, usually defined proportional to the graph Laplacian [9], that is $\mathbf{N} = \gamma\mathbf{L}$, where $\gamma \in [0, \infty)$ is a free parameter that controls the 'deformation' of the kernel. To define L, let's first define a graph $G(V, E)$ with a set of n nodes, V, connected by a set of edges, E. The edge connecting nodes i and j has an associated weight, W_{ij} [9]. In this framework, the nodes are the samples, and the edges represent the similarity among samples in the dataset (see Fig. 1). A proper definition of the graph is the key to accurately introduce data structure in the machine.
Two mathematical tools have to be introduced to understand how matrix L is constructed [11], [16], [17]:

- D is the *degree* matrix of size $n \times n$. Basically, D is a diagonal matrix $D = [d_1, \ldots, d_n]$ containing the number of connections to a node (degree);
- A is the adjacency matrix of size $n \times n$, where the nondiagonal entry is the number of connection from node i to node j, and the diagonal entry is either twice the number of loops at vertex i or just the number of loops. In our case, it is a matrix containing only $(0, 1)$.

Finally, the Laplacian matrix L is defined as $L = D - W$, where W is obtained from A, the adjacency matrix, by assigning weights to each connection. Also, a normalized version of L can be obtained as

$$L_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } d_j \neq 0 \\ -\frac{1}{\sqrt{d_i d_j}} & \text{if } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases}$$

where subscripts i and j stand for the row and column indexes as well as the edges as defined before.
Intuitively, $\mathbf{L}$ measures the variation of the decision function $\mathbf{f}$ along the graph built upon all (labeled and unlabeled) samples [9]. Note that, by fixing $\gamma = 0$, the original (undeformed) kernel is obtained. Therefore, a proper selection of this free parameter theoretically leads to better results than the pure supervised approach (the KOSP method in our case). By plugging (5) into (4), the semi-supervised KOSP, called S^{2}KOSP, can be written as

$$\mathbf{D}_{KOSP}^{ss} = \mathbf{D}_{KOSP} - [\mathbf{K}_{\mathbf{x}_i}^\top (\mathbf{I} + \mathbf{N}\mathbf{K})^{-1}\mathbf{N}\mathbf{K}_{\mathbf{x}_j}]_{\mathbf{d}}, \quad (6)$$

where the subscript $\mathbf{d}$ indicates the operation of extracting information relative to the desired target signature $\mathbf{d}$.

IV. Experiments on thermal hot spot detection

The experiment deals with the hot spot detection using an hyperspectral image. The used dataset comes from the AVIRIS instrument that acquired data over the World Trade Center (WTC), New York, few days later the well-known collapse of the towers of September 11th, 2001. The data was acquired from an altitude of $\sim$ 2 km and has a spatial resolution of about 2 m. The accompanying maps available are false color images acquired over the same site on September 16th and 21st, which show the core affected area around the WTC (Fig. 2). Visual analysis of these data revealed a number of thermal hot spots on September 16th in the region where the buildings collapsed 5 days earlier. Analysis of the data indicates temperatures higher than 400°C. Different hot spots appear in the core zone, and those will be considered in the following as being the target for the detection.

There is no ground truth available (from *in situ* data collection) for the World Trade Center dataset, so we carried out a pixel selection through visual photointerpretation of many false RGB compositions (Figure 2). The WTC dataset is a urban area showing vegetated areas, roofs, streets and shadows but, however, it also shows presence of dust from different materials and debris due to the towers collapse. In order to keep the problem of detection as real as possible, the *a priori* information about the pixel components of the scene are not used, since it is supposed to be known only after performing a 'in loco' analysis. Therefore, in order to define a subspace as simple as possible –as an end-user without prior knowledge would do– we selected only a few pixels from roofs, trees, shadows, water and fire. As already explained, 'fire' is considered the target class in this experiment (see Fig. 2). The pixel closest to the center of mass of each class was selected as the prototype of that class. Results for KOSP and S^{2}KOSP are shown by means of gray level images containing the outputs of the algorithms, and coloured images containing the thresholded outputs. It is worth noting that the selection of background prototypes (from a small number of non-fire pixels) is required by the OSP algorithm in (2) to define a proper background subspace $\mathbf{U}$, and, as a consequence, by the KOSP algorithm (4).

There are several parameters to be tuned for the model. Unfortunately, the reduced number of labeled samples available cannot be used to properly maximize a performance score, such as the area under the ROC curve. Instead, the parameters are tuned according to a 'visual' interpretation of the detection, as it will be better explained in the following subsection.

In the case of KOSP, the algorithm is performed by using windows of size $K_c \times K_c$, and the prototypes extracted from few manually-selected pixels, as explained before. This approach is motivated by the fact that a few spatial coverage of the target class is expected. Therefore, it makes sense to define a maximum window where the target pixels may lie. This procedure not only reduces the computational cost of the algorithm but also improves the detection relying on the smooth spatial variation of the spectral signature.

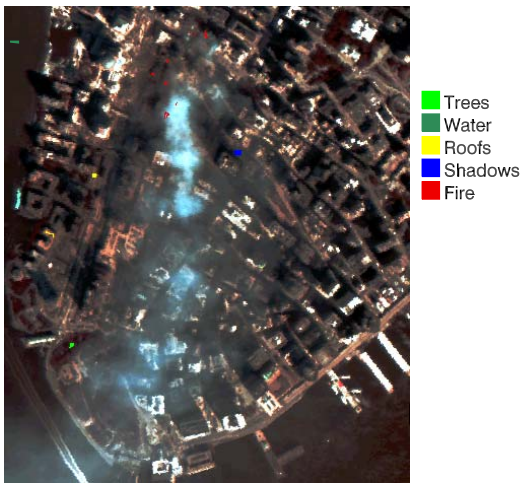

Figure 2. Color composite of WTC Aviris data.

In the semi-supervised version of the algorithm instead, the deformed kernel strictly depends on the choice of unlabeled samples and the distance between the samples of the image. Hence, to explore as much useful information as possible, the best choice would be performing the semi-supervised KOSP by processing the whole dataset at the same time. This would involve an inversion of a huge matrix that is not computationally possible at the moment, so one have to choose a strategy to select the unlabeled samples to build the model. For this experiment, we consider the same block of $K_c \times K_c$ test samples used for the KOSP but, in addition, the deformed kernel is evaluated by considering also an overlapping of K_u pixels between adjacent windows, in order to use 'cross-contextual' information between pixels belonging to neighbor windows (Fig. 3), .

In the following we compare the results of the canonical OSP and KOSP methods, and the proposed semi-supervised KOSP. Two versions of the proposed semi-supervised KOSP are presented: (1) the windowed approach explained before, and (2) a technique to select the unlabeled samples relying on the results obtained with the standard KOSP.

Figure 4 shows the (raw and thresholded) outputs obtained with KOSP and the two versions of S^{2}KOSP presented. From the KOSP predictions (top row), it can be seen that class 'fire' presents a variability that cannot be captured by the KOSP algorithm which offers poor results. Attending to the thresholded outputs (bottom row), there is only one pixel correctly classified by the KOSP (blue), but this is the selected prototype pixel. Pixels marked in yellow are all misclassified pixels with KOSP method, while red pixels are 'false alarms', or pixels erroneously detected by KOSP as belonging to class 'fire'. Actually, a spectral inspection reveals that most of these pixels are very similar to the pixel correctly detected, and moreover, they are so spatially connected to the manually-selected pixels containing fire that they might contain fire as well.

The proposed S^{2}KOSP performs well in this case: the visually-built ground truth is completely detected (blue) and, in addition, all red pixels are not 'false alarms', except two in the lower part of the image. It has to be remarked that S^{2}KOSP is very sensitive to the choice of pixels used as unlabeled samples, so a smarter selection of unlabeled samples should be carried out. Two clear shortcomings of the proposed method are: (i) one can easily note the presence of a highly patched (and annoying) detection map, mainly due to the windowed technique used for detection; and (ii) by using the window technique, the computational time for a given image increases linearly with the number of windows, since for each block the evaluation of the inverse

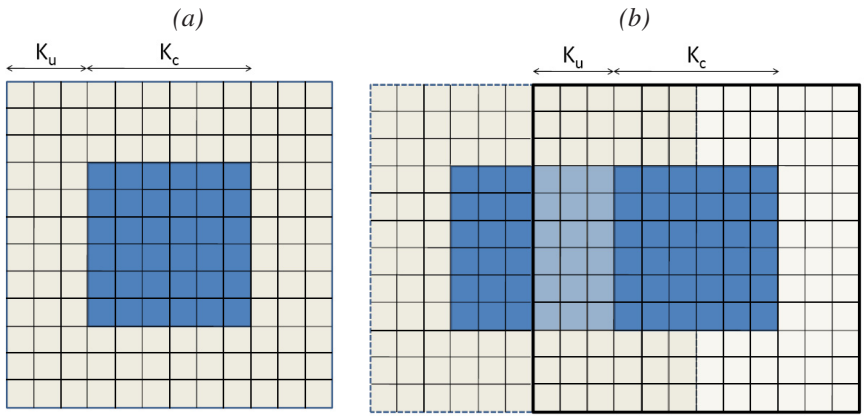

Figure 3. Window Technique used to compute KOSP and S^2KOSP.

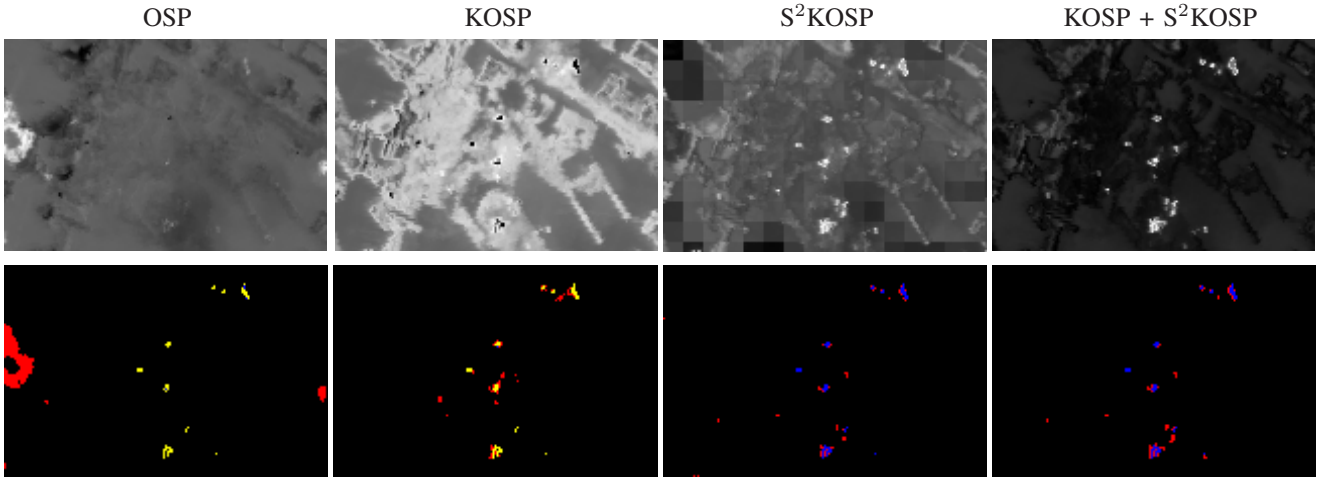

Figure 4. Raw detection (top) and thresholded outputs (bottom) for OSP, KOSP, S^2KOSP, and S^2KOSP with unlabeled samples from the umbralized KOSP in a WTC image detail.

of the matrix $(\mathbf{I}+\mathbf{NK})$ (see Eq. 5) is needed. To solve these problems, an alternative semi-supervised version of the algorithm is introduced here. Note that, even though KOSP cannot capture the variability of the class, it detects pixels that are spectrally very similar to the target and, besides, it is fast and generally reliable. Therefore, we advocate that a proper selection of the unlabeled samples for the S^2KOSP can be accomplished in an easy way by performing a preliminary detection applying KOSP. In this way, the detected pixels by KOSP can be used as unlabeled samples for S^2KOSP. We want to remark that, even if there were pixels identifying false alarms, with a correct tuning of the S^2KOSP parameters, they would be moved away from the principal coordinates of the data manifold to not affecting the detection. Note that, with this alternative method, the inverse of the matrix $(\mathbf{I}+\mathbf{NK})$ has to be evaluated only one time in the complete routine, making the computational time of KOSP and S^2KOSP similar. Detection maps offered by this method are now smoother (see last column in Fig. 4) while performance of the KOSP algorithm is improved: the patches due to the use of different sets of unlabeled samples *per* window have disappeared and the false alarm rate have been reduced.

V. Conclusions

A semi-supervised version of the KOSP algorithm was proposed for target detection applications. The information from unlabeled samples is included in the standard KOSP by means of the graph Laplacian with a contextual unlabeled sample selection mechanism.

The good results obtained suggest that unlabeled information is properly exploited, and that the data manifold can be modeled with the unlabeled surrounding samples. Additionally, the proposed methodology can be useful to easily extend other kernel methods in general, and for target detection in particular.

REFERENCES

[1] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Signal Processing*, vol. 38, no. 10, pp. 1760–1770, Oct 1990.

[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, pp. 779–785, 1994.

[3] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *Signal Processing Magazine, IEEE*, vol. 19, no. 1, pp. 58–69, Jan 2002.

[4] K. I. Ranney and M. Soumekh, "Hyperspectral Anomaly Detection Within the Signal Subspace," *IEEE Geoscience and Remote Sensing Letters, vol. 3, issue 3, pp. 312-316*, vol. 3, pp. 312–316, Jul. 2006.

[5] C. Bachmann, T. Ainsworth, and R. Fusina, "Exploiting manifold geometry in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 441–454, Mar 2005.

[6] ——, "Improved manifold coordinate representations of large-scale hyperspectral scenes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2786–2803, Oct. 2006.

[7] H. Kwon and N. Nasrabadi, "Kernel matched signal detectors for hyperspectral target detection," in *Proc. of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*. Washington, DC, USA: IEEE Computer Society, 2005, p. 6.

[8] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.

[9] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*, 1st ed. Cambridge, MA and London, England: MIT Press, 2006.

[10] X. Zhu, "Semi-supervised learning literature survey," Computer Sciences, University of Wisconsin-Madison, USA, Tech. Rep. 1530, 2005, online document: http://www.cs.wisc.edu/∼jerryzhu/pub/ssl_survey.pdf. Last modified on June 24, 2007.

[11] 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, pp. 779–785, Jul. 1994.

[12] H. Kwon and N. Nasrabadi, "Kernel orthogonal subspace projection for hyperspectral signal classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 12, pp. 2952–2962, Dec. 2005.

[13] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: from transductive to semi-supervised learning," in *International Conference on Machine Learning (ICML)*, ser. ACM International Conference Proceeding Series, L. De Raedt and S. Wrobel, Eds., vol. 119. ACM, 2005, pp. 824–831. [Online]. Available: http://dx.doi.org/10.1145/1102351.1102455

[14] J. Muñoz-Marí, L. Gómez-Chova, G. Camps-Valls, and J. Calpe-Maravilla, "Image classification with semi-supervised one-class support vector machine," in *SPIE International Symposium Remote Sensing XIV*, Cardiff, UK, Set 2008, p. 71090B.

[15] G. Camps-Valls, J. Muñoz-Marí, L. Gómez-Chova, and J. Calpe-Maravilla, "Semi-supervised support vector biophysical parameter estimation," in *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'2008*, Boston, USA, Jul 2008.

[16] F. Chung, *Spectral Graph Theory*, 1st ed., ser. CBMS Regional Conference Series in Mathematics. Providence, RI: American Mathematical Society, February 1997, no. 92.

[17] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 2044–3054, 2007.