IMPROVING RESOLUTION IN SUPERVISED PATCH-BASED TARGET DETECTION

Ron Amit, Gal Mishne and Ronen Talmon

Technion - Israel Institute of Technology, Haifa 32000, Israel

{ronamit@campus, galga@tx, ronen@ee}.technion.ac.il

ABSTRACT

Recently, a supervised graph-based target detection method was proposed based on a new affinity measure between a set of target training patches and a test image. In this paper, we propose a new high-resolution detection score, which enhances the performance of the previous method by utilizing the known locations of the targets in the training images. We show that our new score is more reliable and spatially accurate, not only improving the detection resolution of true targets, but also reducing the number of false alarms. The method is successfully tested on side-scan sonar images of sea-mines, demonstrating an improved true detection rate. Our approach is general and can improve the detection resolution of the target in other patch-based detection algorithms for various signals and applications.

Index Terms— automatic target detection, object localization, nonlinear-dimensionality reduction, mine detection, side-scan sonar

1. INTRODUCTION

Automated target detection in images is a common problem in many civilian and military applications, where the images can be obtained from a large variety of sensors such as radar, sonar or optical sensors [1–6]. The goal is to automatically detect a target in a cluttered background, or at least indicate suspicious regions with high probability of target presence for further inspection by an expert. This has great practical importance, since typically a large number of images are collected in such applications, and manual inspection is costly. A supervised approach is useful in target detection, assuming a training set of target image patches is available to the algorithm prior to acquiring the test image. This prior knowledge can be used for modeling the target, feature selection, and training a classifier [1–3, 7, 8]. Target detection can be challenging due to noise, cluttered background, large variations in the target appearance, the large amount of data to be processed, and limited labeled training data.

Supervised manifold learning methods, which construct a new representation of test data using a training set, have become popular in recent years for different applications, such as audio enhancement, classification of electromagnetic measurements, texture extraction and dynamical systems [9–12]. Recently, Mishne et al. [13] presented a new supervised target detection method. The method is based on calculating a local model for each training patch using its neighborhood within the training set. By controlling the measure of locality that defines these neighborhoods, one can construct an affinity measure that is invariant to perturbations in the appearance of the target, as defined by the variability of the target in the neighborhood. This new affinity is integrated in a graph-based framework to embed the patches extracted from a test image in a low-dimensional space. The detection score is then calculated in this space to determine the presence of a target.

Obtaining an accurate location of the target is important in both military and medical applications, for example to ensure accurate extraction of the object for classification, or precise acquisition of the target for tracking [2, 14–17]. However, in [13] the detection score for each test patch is associated with the center pixel of the patch, and does not reflect the true target location within the patch.

In this paper, we introduce a method to enhance the spatial resolution of the detection score. By using prior information regarding the exact location of the targets in the training images, we can calculate a new high-resolution detection score which is more reliable and spatially accurate. We show that this method improves the spatial accuracy of the true detections, and reduces the number of false alarms. We demonstrate our approach in the setting of sea-mine detection in side-scan sonar images [1, 2, 13, 18–20]. However, the presented method is not application-specific, and can be used to improve the detection score of other patch-based supervised detection algorithms, with minimal algorithmic changes.

The remainder of the paper is organized as follows. In Sec. 2 we describe the supervised target detection problem setting, and briefly review the method introduced in [13] that uses local models within a given training set to define an affinity measure between training and test patches. Section 3 presents a low-dimensional embedding of the data based on the affinity measure. In Sec. 4, we present the embedding-based detection score and a novel method for enhancing the spatial resolution of the score. Finally, Sec. 5 presents experimental results on real-world data. In particular, we focus on demonstrating the improvement in detection performance attained by using our method.

2. LOCAL NEIGHBORHOOD MODELING

2.1. Problem Setting

In supervised target detection, a set of training images (either real or simulated) is typically used as a prior knowledge on the target appearance. Once a test image is acquired, the task is to detect whether it contains a target and at which location. In our approach, the test image and the set of training images are treated as sets of image patches sized $\sqrt{N} \times \sqrt{N}$. We consider two sets of patches. The first is a set of $\overline{M}$ training patches, denoted by $\{\overline{Z}_i\}_{i=1}^{\overline{M}}$, which is obtained by extracting all the overlapping patches in the training images that contain sufficient target content. Patches in which the target center is too close to the border or not even contained in the patch, i.e. they contain mostly background pixels, are discarded. The second set consists of $M \gg \overline{M}$ test patches, denoted by $\{Z_i\}_{i=1}^{M}$, and contains all the overlapping patches from the test image.

We assume that each training patch contains target content with

The work was supported by the EU Seventh Framework Programme (FP7) under Marie Curie Grant No. 630657.

added independent measurement noise:

$$\overline{Z}_i(x) = f(x;\boldsymbol{\theta}_i) + \eta_i(x), \quad x \in \{1,...,N\}, \tag{1}$$

where $\boldsymbol{\theta}_i$ are the target appearance parameters, $f(x;\boldsymbol{\theta}_i)$ is a smooth nonlinear function of $\boldsymbol{\theta}_i$ at pixel x, and $\eta_i(x)$ is measurement noise. More specifically, the parameters $\boldsymbol{\theta}_i$ represent the degrees of freedom in the target appearance, such as width, length, and orientation, and are implicitly defined based on the target examples in the given task. Both $\boldsymbol{\theta}_i$ and $f(x;\boldsymbol{\theta}_i)$ are assumed to be unknown. This model neglects explicit interactions between pixels in the patch, however, since all pixels within the patch share the same mapping f and the same parameter vector $\boldsymbol{\theta}$, this provides an implicit conditional dependency between pixels. Test patches are modeled in an analogous way, but they may contain background and measurement noise with no target content.

2.2. Affinity Measure

Mishne et al. [13] recently presented a supervised target detection method based on a new affinity measure between patches. The affinity measure is constructed in a data-driven manner that encompasses the intrinsic target parameters. It aims to indicate whether two patches contain similar target content, while being invariant to perturbations in the target appearance parameters. Choosing a statistical distribution to model the target and background, or explicit estimation of shape parameters are not required. Instead, the variability in the target appearance is learned from the relations within local neighborhoods by calculating the statistics of each patch based on its neighbors.

The neighborhood of a training patch $\overline{Z}_i$ is denoted by $\mathcal{N}_i$, where the neighboring relations grouping similar patches together are defined by the user; for example a naïve grouping can be attained by using k-nearest-neighbors (of patches). For every training patch, its empirical local pixel-wise mean and variance are estimated based on its neighbors in the training set, defining a model for each patch:

$$\hat{\mu}_i(x) = \frac{1}{k}\sum_{\overline{Z}_j \in \mathcal{N}_i} \overline{Z}_j(x), \quad \hat{\sigma}_i^2(x) = \frac{1}{k}\sum_{\overline{Z}_j \in \mathcal{N}_i} \left(\overline{Z}_j(x) - \hat{\mu}_i(x)\right)^2. \tag{2}$$

Note that the statistics are local in the sense that they are calculated for a given pixel x in patch i, i.e., $\overline{Z}_i(x)$, based on the corresponding pixels $\overline{Z}_j(x)$ in its patch neighborhood, i.e. for $\overline{Z}_j \in \mathcal{N}_i$. These are not spatial statistics within the patch $\overline{Z}_i$ itself.

A non-symmetric affinity measure between training patch $\overline{Z}_j$, modeled using $\hat{\mu}_j$ and $\hat{\sigma}_j$, and test patch Z_i is defined as

$$a^2(Z_i, \overline{Z}_j) = \sum_{x=1}^{N} \left(Z_i(x) - \hat{\mu}_j(x)\right)^2 / \hat{\sigma}_j^2(x). \tag{3}$$

Pixels with low variance are more likely to contain target content, so they should contribute more to the affinity measure, as opposed to pixels with high variability. Pixels with high variance either indicate high variability in the appearance of the target, or that they contain only the background and do not contribute to determining the presence of a target. Therefore, high variance pixels should be assigned a low weight. Hence, the contribution of each pixel to the affinity measure is weighted according to the inverse of the local variance.

The local variance determines the weight of each pixel in (3), and therefore, it implicitly defines an invariance to slight differences in the appearance of the target. This property can be employed by the user to define the neighborhoods $\mathcal{N}_i$ so as to achieve a desired invariance. For example, to construct an invariance to slight rotations, patches containing the target at an identical location, but slightly different orientations are grouped together. This can be done by generating training images in simulation, thus having direct control over the target parameters [14, 15, 21]. If the training set is extracted from real data, the parameters are unknown, so choosing neighbors is more challenging. One option is choosing k-nearest neighbors using the Euclidean norm. Another option is to use the Hamming norm between the binary target indicator patches. The binary indicator patch has a value of '1' for pixels in the training patch with target content and '0' for the rest (pixels with target content can be marked in a supervised setting).

In terms of the intrinsic target parameters $\boldsymbol{\theta}$, neighboring patches are consistent in some of the parameters, while having slight variations in others. For a given neighborhood, the parameter vector can be conceptually separated into two sets: $\boldsymbol{\theta} = (\boldsymbol{\theta}_c, \boldsymbol{\theta}_v)$, where $\boldsymbol{\theta}_c$ are the fixed parameters within the neighborhood, while the parameters in $\boldsymbol{\theta}_v$ have variability. The fixed parameters are assumed to describe the target, whereas the varying parameters describe slight perturbations in the target appearance to which we want to be invariant. By learning the variance of each pixel in a patch neighborhood, the contribution of each pixel to the affinity measure can be weighted, where pixels with higher sensitivity to variations in $\boldsymbol{\theta}_v$ are assigned a lower weight. In this way, the constructed affinity measure is more invariant to perturbations in the target appearance compared to the Euclidean distance, which equally weights all the pixels.

Based on the Taylor expansion of Z_i with respect to $\boldsymbol{\theta}$, it can be shown [13] that the local empirical variance relates to the perturbations in the parameters vector via:

$$\hat{\sigma}_i^2(x) \approx \nabla_{\boldsymbol{\theta}^v} f^T(x;\boldsymbol{\theta}_i)\Omega_i^v \nabla_{\boldsymbol{\theta}^v} f(x;\boldsymbol{\theta}_i) + \sigma_\eta^2(x), \tag{4}$$

where $\nabla_{\boldsymbol{\theta}^v} f^T(x;\boldsymbol{\theta}_i)$ is the gradient of $f(x;\boldsymbol{\theta}_i)$ with respect to the intrinsic parameters $\boldsymbol{\theta}^v$, Ω_i^v is a diagonal matrix with the empirical variances of the parameters in $\boldsymbol{\theta}_i^v$ as its diagonal, and $\sigma_\eta^2(x)$ is the variance of the measurement noise. Thus, the local empirical variance of the pixel x depends on the perturbations in the parameter vector within the local neighborhood of the training patch.

3. GRAPH-BASED EMBEDDING

We define an $M \times \overline{M}$ affinity matrix based on the affinity measure (3) between the training set $\{\overline{Z}_j\}_{j=1}^{\overline{M}}$ and test set $\{Z_i\}_{i=1}^{M}$:

$$\mathbf{A}[i,j] = \exp\{-a^2(Z_i, \overline{Z}_j)/\epsilon^2\}, \tag{5}$$

where ϵ is a scale factor, commonly set as the median of the distances within the training set. The Gaussian function enhances locality, as patches with distance larger than ϵ have negligible affinity.

This matrix is used to calculate a low-dimensional embedding of the training and test patches. The singular value decomposition (SVD) of the matrix $\mathbf{A}$ yields a set of decreasing singular values $\{\sqrt{\lambda_l}\}_{l=1}^{\overline{M}}$ ($\overline{M} < M$), left singular vectors $\{\psi_i\}_{l=1}^{M} \subseteq \mathbb{R}^{\mathbf{M}}$ and right singular vectors $\{\phi_l\}_{l=1}^{\overline{M}} \subseteq \mathbb{R}^{\overline{\mathbf{M}}}$. The relationship between the first $\overline{M}$ singular left and right eigenvectors of $\mathbf{A}$ can be expressed as [10, 22]:

$$\psi_l = \frac{1}{\sqrt{\lambda_l}}\mathbf{A}\phi_l. \tag{6}$$

The squared singular values $\{\lambda_l\}_l$ and right singular vectors $\{\phi_l\}_{i=l}^{\overline{M}}$ are the eigenvalues and eigenvectors of the symmetric $\overline{M} \times \overline{M}$ matrix: $\mathbf{W} = \mathbf{A}^T\mathbf{A}$. The matrix element $\mathbf{W}[i.j]$ can be interpreted as an affinity measure between two training patches $\overline{Z}_i$ and

Algorithm 1 Supervised Target Detection Algorithm

Training phase: Given training set $\{\overline{Z}_j\}_j$

Define neighborhoods $\{\mathcal{N}_j\}_j$

For each $\overline{Z}_j, \mathcal{N}_j$ calculate $\hat{\mu}_j(x)$ and $\hat{\sigma}_j(x)$ (2)

Optional: Calculate $\mathbf{W}$, $\{\lambda_l\}_l$, $\{\phi_l\}_{l=1}^d$

Test phase: Given test patches $\{Z_i\}_{i=1}^M$, training shifts $\{\Delta(j)\}_j$

Calculate affinity matrix $\mathbf{A}$ (5)

Calculate test set embedding $\{\psi_l\}_{l=1}^d$ using SVD of $\mathbf{A}$ or (6)

Detection score: **for** $i = 1$ to M **do**

    Calculate the embedding norm $s(i)$ (8)

    Initialize new detection score to zero: $\widetilde{s}(i) \longleftarrow 0$

    for $j = 1$ to $\overline{M}$ **do**

        Calculate w_{ij} (10)

        $\widetilde{s}(i + \Delta(j)) \longleftarrow \widetilde{s}(i + \Delta(j)) + w_{ij} s(i)$

$\overline{Z}_j$ via all the patches in the test set [9,11]. The spectrum of the symmetric matrix $\mathbf{W}$ typically exhibits a spectral gap, i.e. only a few eigenvalues have significant value, while the rest are close to zero. Thus, by retaining only the d most significant spectral components, the leading d left singular vectors are used for the low-dimensional embedding of the test patches, $\{Z_i\}_{i=1}^M$:

$$\mathbf{\Psi}_d(i) : Z_i \rightarrow \big(\psi_1(i), \psi_2(i), ..., \psi_d(i)\big). \tag{7}$$

The embedding maps each patch Z_i to a d-dimensional vector, and is expected to reveal which patches in the image are similar to the reference set. In similar fashion, the right singular vectors provide a low-dimensional embedding of the training set $\{\overline{Z}_i\}_{i=1}^{\overline{M}}$ [10,12,22].

The matrix $\mathbf{W}$ can be approximated in the training phase solely using the training data, as in [13]. Its eigen-decomposition yields $\{\lambda_l\}_l$ and $\{\phi_l\}_l$. Then, for each test image, the embedding of the test patches $\{\psi_l\}_l$ can be efficiently calculated using simple matrix multiplication by $\mathbf{A}$ (6), instead of performing SVD for each image separately. This is a significant advantage of the supervised graph, as opposed to unsupervised manifold learning [23] that requires performing a costly calculation and eigen-decomposition of an $M \times M$ affinity matrix on the test image. In target detection, the test image typically exhibits non-uniform sampling, and it is advised to use the Laplace-Beltrami normalization as in [10, 13, 22, 24], so the embedding does not depend on the density of the data points.

4. TARGET DETECTION

4.1. Embedding-based Detection Score

In the low-dimensional embedding of the test patches, the target can be separated from the background. According to (6), the coordinates of the embedding are a weighted sum of the affinities of the test patch to all training set patches. Background patches have low affinity to all the training patches, since they do not contain the target, i.e. for a background patch Z_i, $\mathbf{A}[i,j] \rightarrow 0 \quad \forall j \in 1, ..., \overline{M}$. Hence, their embedding will have low values. However, test patches with target content have high affinity to some of the training patches, hence their embedding has higher values. Therefore, background patches are expected to cluster near the origin in the embedding space, while target patches are embedded far from the origin.

Thus, an embedding-based detection score $s(i)$ can be defined for every test patch i [13]:

$$s(i) = \|\mathbf{\Psi}_d(i)\|^2 = \sum_{l=1}^{d} \psi_l^2(i). \tag{8}$$

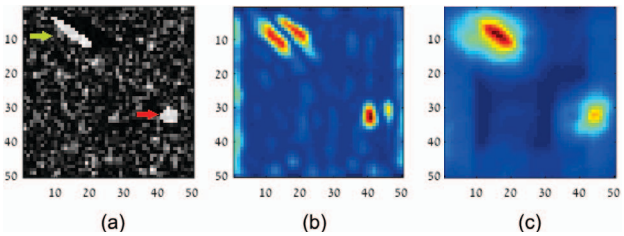

Fig. 1. (a) Original image: target indicated by green arrow, false alarm indicated by red arrow. (b) Low-resolution detection score: both target and false alarm receive a high score and the score is not centered on the object. (c) High-resolution score is concentrated at the target location and with significantly higher values than at the false alarm location.

A higher detection score indicates higher likelihood for target presence. Note that another reasonable detection score can be the sum of the affinity of a test patch to all the training set. However, experimental results show that the affinity sum is a less reliable target indicator compared to the embedding norm [13]. Thus, the affinity in itself is insufficient to determine the presence of the target, and the embedding is an essential component of the proposed approach. The SVD provides a global representation, integrating all the affinities of all patches, instead of treating each test patch on its own.

4.2. High-resolution Detection Score

The training set is composed of all overlapping patches extracted from training images of the target, containing sufficient target content. Therefore, in most training patches, the target center is not located at the center pixel of the patch. Yet, when comparing it to a test patch, the detection score is (arbitrarily) associated with the center pixel of the test patch, and not the true target location. Thus, a pixel with a high detection score indicates the presence of a target with a location uncertainty of half the patch length. This results in the detection score image having low spatial resolution, limited by the patch size. In this section, we present a new method to enhance the spatial resolution of the detection image by using the target location in each training patch. As this is a supervised approach, we assume that the exact locations in the training images are available.

For every patch that is extracted from a training image, we calculate the shift of the target center relative to the patch center using the known target location in the training image. The shift in training patch j is defined as:

$$\Delta(j) = (\Delta_x(j), \Delta_y(j)),\ j = 1, ..., \overline{M}, \tag{9}$$

where $\Delta_x(j)$ and $\Delta_y(j)$ are the horizontal and vertical shifts respectively.

We will use these shifts to calculate a new high-resolution score for each test patch. The affinity matrix (5) indicates the contribution of each training patch to the score of a test patch. If a given training patch has a certain shift and this patch has high contribution to the score of a test patch, then we want to "move" the score using the same shift to the appropriate pixel in the test image. In general, a test patch can have high affinity to several training patches. We want to distribute the score according to the shift of each training patch in a weighted manner. We choose the weight w_{ij} to be the affinity (3) between the test patch Z_i and the specific training patch $\overline{Z}_j$, relative

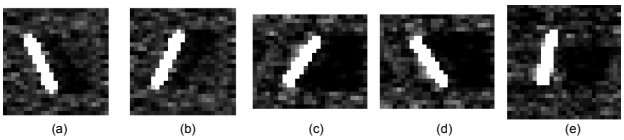

Fig. 2. Training images of sea mines in side-scan sonar. The images (b) and (d) are vertical reflections of (a) and (c), respectively, added to the set to increase variability. The sea-mine highlights were saturated in order to diminish variability of target intensity in the training set, which is due to noisy acquisition. The training set is composed of overlapping patches extracted from the five training images.

to the affinity to all the training set:

$$w_{ij} = \frac{A[i,j]}{\sum_{l=1}^{\overline{M}} A[i,l]}. \tag{10}$$

Thus, the affinity (3) is integrated in both the embedding and the detection score. Our method is summarized in Algorithm 1, where the new detection score is denoted by $\widetilde{s}(i)$.

The algorithm "localizes" the score in the detection image at the target center location. Thus, we get a more reliable and spatially accurate detection. Note that steps 8-11 in Alg. 1 are general and can be applied to other patch-based detection methods, and also used for target detection and tracking in 1D and 3D. The method relies on an available training set of target patches, in which the relative target location inside each patch varies and is known. An affinity measure between training and test patches is also required, but is not limited to the proposed affinity (3).

Figure 1 illustrates the benefits of the new resolution enhancement method. The original image (a) contains both a target and a false alarm, indicated by the green arrow and a false alarm indicated by the red arrow. Using the embedding norm score, high values are assigned at the edges of the target, but low values at its center (b). Moreover, the false alarm gets a higher score than the target. The two problems are solved when applying the high-resolution score (c): the detection score is concentrated at the true target location and with significantly higher values than for the false alarm.

5. EXPERIMENTAL RESULTS

In this section the proposed method is tested in a sea-mine detection problem with real-world side-scan sonar images. The images contain sea-mines (the desired target) as well as reflections from the seabed (which are considered background clutter). Sea-mines appear as bright white objects (highlights) accompanied by a dark region (shadow) to their right, which is the result of the sea-mine blocking the sonar waves from reaching the sea-bed [18, 19]. The method is tested on a set of 46 side-scan sonar images of sea-mines collected by the Naval Surface Warfare Center Coastal System Station (Panama City, FL). Each image is cropped to a region sized 200 (range)$\times$ 200 (cross-range) cells which contains a sea-mine and various background types. The typical size of a sea-mine in the images is about 15 by 3 pixels for the highlight and shadow length is about 15 pixels. The patch size N is therefore chosen based on the prior knowledge on the typical size of the target and the image resolution, such that it covers a significant portion of the target but not necessarily its entire body. Based on experiments with the data set, we chose patches of size 10×10 pixels.

Training images of the target used for our training set are shown in Figure 2. For more variation in the possible orientations of the target, we added vertical reflections of two images (b),(d) to three existing images (a),(c) and (d). The training images exhibit variations in orientation, size and shadow shape. The size of each training image is about 25$\times$ 25 pixels. All overlapping patches are extracted from the images, but patches with low target content are discarded. After obtaining only the relevant patches we get a training set $\{\overline{Z}_i\}$ of $\overline{M} = 277$ patches sized 10×10 pixels. From each test image, we extract a set of $M = 36481$ overlapping patches. The embedding dimension was set to $d = 9$.

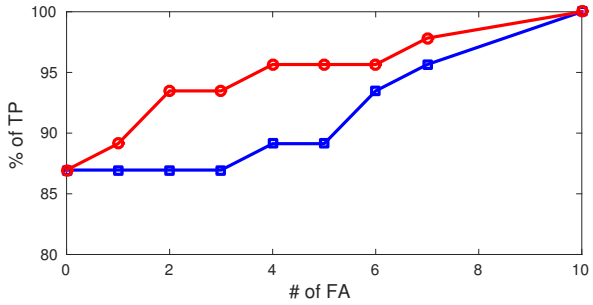

Fig. 3. TP percentage versus FA rate. (Blue, "square") low-resolution score [13]. (Red "circle") High-resolution score.

In Fig. 3, we demonstrate the improvement gained by our high-resolution score. Our high-resolution score image is calculated using Algorithm 1. We compare our score to the method proposed in [13], in which each pixel is assigned the norm of its embedding coordinates (8). The score images are then spatially smoothed to repress small detections which are due to noise using a narrow Gaussian filter of size 3×3 and standard deviation of 0.5. We calculate a ROC curve by applying a threshold to the detection score, resulting in a binary image. A detection on the sea-mine is considered to be a true positive (TP) for a given image, and any other detections are false alarms (FA). Thus, there may be more than one FA per image, but only one TP. Each threshold gives us a (TP,FA) pair plotted in the ROC curve. For each method, we plot the percentage of TPs per overall number of FAs in the test set. The blue (squares) plot is the ROC curve for the original algorithm, whereas the red (circles) plot is the curve for the improved detection score. Our high-resolution score yields a significant gain of approximately 7 percent for low number of FAs. For number of FAs greater than 1, we achieve a TP percentage higher than 90%.

6. CONCLUSIONS

We have presented a data-driven target detection method in a supervised setting based on a new affinity measure. Our method embeds the test image patches in a low-dimensional space, enabling simple calculation of a detection score. We introduced a new method to enhance the spatial resolution of the detection score based on the affinity measure and the known target locations in the training set. In contrast to the common practice of associating the detection score with the center pixel of a test patch, we distributed the score across the test patch, thereby concentrating the score at the correct location of the target. Our method is general and can improve the detection resolution of supervised patch-based algorithms, and as a result, enhance their reliability and spatial accuracy. The new method has been tested in the task of sea-mine detection in side-scan sonar images and significantly improved the detection results.

7. REFERENCES

[1] G. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proc. MTS/IEEE Oceans Conf. and Exhibition*, vol. 1. Marine Technol. Soc, 2001, pp. 130–134.

[2] E. Coiras, P.-Y. Mignotte, Y. Petillot, J. Bell, and K. Lebart, "Supervised target detection and classification by training on augmented reality data," *IET Radar Sonar Navig.*, vol. 1, no. 1, pp. 83–90, 2007.

[3] Y. Chen, N. Nasrabadi, and T. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, 2011.

[4] H. Kwon and N. Nasrabadi, "Kernel matched subspace detectors for hyperspectral target detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 178–194, Feb 2006.

[5] P. Torrione, K. Morton, R. Sakaguchi, and L. Collins, "Histograms of oriented gradients for landmine detection in ground-penetrating radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1539–1550, Mar 2014.

[6] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 24–33, Jan 2014.

[7] C. Spence, L. Parra, and P. Sajda, "Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model," in *Proc. of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, 2001, pp. 3–10.

[8] A. Noiboar and I. Cohen, "Anomaly detection based on wavelet domain GARCH random field modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1361–1373, 2007.

[9] R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman, "Supervised graph-based processing for sequential transient interference suppression," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, no. 9, pp. 2528–2538, 2012.

[10] D. Kushnir, A. Haddad, and R. R. Coifman, "Anisotropic diffusion on sub-manifolds with application to earth structure classification," *Appl. Comput. Harmon. Anal.*, vol. 32, no. 2, pp. 280 – 294, 2012.

[11] A. Haddad, D. Kushnir, and R. R. Coifman, "Texture separation via a reference set," *Appl. Comput. Harmon. Anal.*, vol. 36, no. 2, pp. 335 – 347, Mar. 2014.

[12] R. Talmon and R. R. Coifman, "Intrinsic modeling of stochastic dynamical systems using empirical geometry," *Appl. Comput. Harmon. Anal.*, vol. 39, no. 1, pp. 138 – 160, 2015.

[13] G. Mishne, R. Talmon, and I. Cohen, "Graph-based supervised automatic target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2738–2754, May 2015.

[14] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *IEE Proc. Radar, Sonar and Navigation*, vol. 151, no. 1, pp. 48–56, Feb. 2004.

[15] Y. Petillot, Y. Pailhas, and J. Sawas, "Target recognition in synthetic aperture and high resolution side-scan sonar," in *Proc. European Conference on Underwater Acoustics*, 2010, pp. 99–106.

[16] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, 2003.

[17] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International Journal of Computer Vision*, vol. 73, no. 1, pp. 41–59, 2007.

[18] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *IEEE J. Ocean. Eng*, vol. 28, no. 1, pp. 90–105, Jan. 2003.

[19] E. Dura, Y. Zhang, X. Liao, G. Dobeck, and L. Carin, "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE J. Ocean. Eng*, vol. 30, no. 2, pp. 360–371, 2005.

[20] G. Mishne and I. Cohen, "Multiscale anomaly detection using diffusion maps," *IEEE J. Sel. Topics Signal Process.*, vol. 7, pp. 111 – 123, Feb. 2013.

[21] E. Coiras, Y. Petillot, and D. Lane, "Multiresolution 3-D reconstruction from side-scan sonar images," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 382–390, 2007.

[22] R. Talmon and R. R. Coifman, "Empirical intrinsic geometry for nonlinear modeling and time series filtering," *Proceedings of the National Academy of Sciences*, vol. 110, no. 31, pp. 12 535–12 540, 2013.

[23] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, July 2006.

[24] A. Singer and R. R. Coifman, "Non-linear independent component analysis with diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 25, no. 2, pp. 226 – 239, 2008.