

Detection of anomalies in texture images using multi-resolution random field models

Lior Shadhan, Israel Cohen*

Department of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel

Received 26 March 2007; received in revised form 24 May 2007; accepted 31 May 2007

Available online 7 June 2007

Abstract

In this paper, we present a multi-resolution random field model (RFM) and a corresponding algorithm for anomaly subspace detection. We utilize the redundant discrete wavelet transform (RDWT) for generating a multi-resolution feature space, and model each layer by a non-casual RFM with different sets of parameters. A multi-resolution matched subspace detector (MSD) is designed for detecting targets in the background multi-resolution RFM noise environment. The improved performance of the proposed algorithm is demonstrated compared to using an MSD-based anomaly detector and multi-resolution Gaussian Markov random field (GMRF) model.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Anomaly detection; Matched subspace detector; Texture analysis; Texture classification; Texture segmentation

1. Introduction

Anomaly detection techniques are useful in numerous applications, both civilian and military, such as detection of targets in multi-spectral and hyper-spectral images [1–11], detection of sea-mine targets in side-scan sonar images [12–15], detection of land-mine targets in ground penetrating radar images [16–19], detection of surface targets in synthetic aperture radar images [20,21] and other signal processing and image analysis applications. The type of the scene and the type of the anomalous targets are application dependent. The detection process is generally performed with respect to

a predefined probabilistic model and an appropriate feature space, where the anomalous elements can be identified as different from the background clutter.

Anomaly detectors often use Bayesian classifiers, utilizing available *a priori* knowledge and *a posteriori* parametric statistics of both background clutter and anomalous targets. Single hypothesis test (SHT) [22] is employed in detection problems where the targets are substantially diverse, and not necessarily conform to a uniform model or even a characterizing subspace. Matched signal detector is employed when a typical signature of a target is available. Matched subspace detector (MSD) [3,4,23–25] and adaptive subspace detector (ASD) [26] are used when partial information about the targets allows to define a subspace for targets.

Random field models (RFM), such as simultaneous auto-regressive (SAR) and Gaussian Markov

*Corresponding author. Tel.: +972 4 8294731; fax: +972 4 8295757.

E-mail addresses: shadhan@techunix.technion.ac.il (L. Shadhan), icohen@ee.technion.ac.il (I. Cohen).

random field (GMRF), have been applied extensively for spatial analysis of textural features in texture synthesis [27,28], image segmentation [5,29–32], texture classification [33,34] and target detection [5,10,13,25,35] algorithms. Single resolution spatial analysis introduces high false alarms in semi-homogenous background textures due to deviations from the RFM. In order to overcome this, Goldman and Cohen [25] proposed a multi-resolution GMRF model and a corresponding anomaly detection algorithm using an MSD classifier, achieving better detection results when compared to single resolution spatial analysis detection algorithms. However, the SAR and the GMRF models are sensitive to the choice of their neighbor set [27,36,37] since different neighbor sets account for different textural patterns. Inappropriate choice of a neighbor set increases the prediction error and false alarm rate.

Multi-resolution decompositions, such as the wavelet transform, are often used for feature extraction. These decompositions employ a set of multi-scale bandpass oriented filters for decomposing the image, and decoupling high-order statistical features of natural images [38,39]. However, wavelet-based multi-resolution decompositions are characterized by heavy tails of the marginal probability density function of the features (known as excess kurtosis) and volatility clustering (i.e., large changes tend to follow large changes and small changes tend to follow small changes) [40]. The first phenomenon leads to high false alarm rate when Gaussian-based classifiers are used, due to model mismatch. The second phenomenon leads to deviations of the clutter image from its RFM, resulting in higher false alarm rates due to signal-to-noise ratio (SNR) degradation.

In this work, we present a multi-resolution non-casual RFM and a corresponding unsupervised anomaly subspace detection algorithm. The proposed multi-resolution RFM captures the highly correlated spatial nature of the background clutter and is less susceptible to the choice of neighbors. We utilize the redundant discrete wavelet transform (RDWT) for generating a multi-resolution feature space, and each layer is then modeled by a non-casual RFM with different sets of parameters. A multi-resolution MSD is formulated for detecting targets in the background multi-resolution RFM noise environment with possible additive subspace interference signals. The proposed algorithm is implemented and its performance is analyzed in

various scenes containing Brodatz-like background textures [41] and target anomalies. We demonstrate the improved performance of the proposed algorithm compared to using an MSD-based anomaly detector and multi-resolution GMRF model [25].

The rest of the paper is organized as follows. In Section 2, we formulate the detection problem. In Section 3, we introduce an RFM for texture images and a corresponding multi-resolution feature space. In Section 4, we present the proposed anomaly subspace detection algorithm. In Section 5, we analyze the performance of the proposed anomaly detection algorithm and compare the detection results to those obtained by using competing methods. In addition, we demonstrate the application of the proposed algorithm to automatic target detection of multiple target types using a set of subspace detectors.

2. Problem formulation

Let $\Omega = \{\mathbf{v} : 1 \leq v_1 \leq M, 1 \leq v_2 \leq M\}$ be the support of an image $\{y(\mathbf{v})\}_{\mathbf{v} \in \Omega}$ containing a background natural texture, $x(\mathbf{v})$, rare target and additive interfering signals scattered around in the image, denoted as $h(\mathbf{v})$ and $g(\mathbf{v})$, respectively. The supports of the target and interfering signals are assumed to be much smaller than the support of the background image. We define two hypotheses for each pixel $\mathbf{v} \in \Omega$:

$$\begin{aligned} H_0 : y(\mathbf{v}) &= x(\mathbf{v}) + g(\mathbf{v}), \\ H_1 : y(\mathbf{v}) &= x(\mathbf{v}) + g(\mathbf{v}) + h(\mathbf{v}), \end{aligned} \quad (1)$$

where H_0 and H_1 represent absence and, respectively, presence of an anomalous target in the image. The problem at hand is to make a decision between H_0 and H_1 for every pixel $\mathbf{v} \in \Omega$ with

$$\begin{aligned} P_D &\triangleq P(H_1|H_1), \quad P_D \geq 1 - \varepsilon_1, \\ P_{FA} &\triangleq P(H_1|H_0), \quad P_{FA} \leq \varepsilon_2, \end{aligned} \quad (2)$$

for given values of ε_1 and ε_2 .

3. Statistical model formulation

3.1. Segregation enhancement

Multi-resolution decompositions, such as the wavelet transform, are often used for feature extraction. The undecimated discrete wavelet transform is preferable to the standard decimated wavelet decomposition, mainly because it tends to

decrease the variability of the estimated texture features, hence improving texture classification performance [42]. The undecimated discrete wavelet transform also results in a texture characterization invariant under texture translation. Here we utilize the RDWT for generating a multi-resolution feature space.

Unser and Eden explored in [43] several non-linearities and their effects on texture segmentation schemes. They have concluded that squaring, followed by averaging and then by a logarithm function, improves segmentation results. Mittelman and Porat [29,44] argued that those non-linearities, when applied to wavelet coefficients of natural texture images, result in a normally distributed feature space. Motivated by the work in [29,43,44], we have formulated in [45] an RDWT-based multi-resolution feature space and a corresponding SHT anomaly detection scheme, using the same non-linearities but with additional averaging step which improved detection performance considerably. The multi-resolution feature space in [45] is a linear combination of independent Gaussian random vectors, and follows a multivariate Gaussian distribution. This suggests that the feature space which is presented in [45] is suitable for use with an MSD, since the MSD is based on a multivariate Gaussian model. However, in practice, it degrades the segregation between additive anomalies and background clutter. This is due to the averaging stages and the influence of the logarithm on the variance of the feature space. Furthermore, the use of the logarithm limits the ability to formulate appropriate subspaces for target and interference signals, and makes the MSD behave like a matched signal detector for a finite number of target signals. Yet, the use of the squaring non-linearity by itself allows for a better segregation of additive anomalous targets from the background clutter. Furthermore, the use of the squaring non-linearity reduces the RFM prediction error variance [46, p. 96] and hence results in a background clutter which can be modeled by an RFM more accurately.

3.2. Single layer 2-D RFM

The SAR and GMRF models are widely used for texture modeling and texture classification. These models were also used for detection of target signals in background clutter [5,10,13,25,35]. The models represent a pixel in a given image as a weighted sum of pixels at nearby locations and the prediction

error is referred as the innovations process. Under the SAR model the innovations are assumed uncorrelated, whereas under the GMRF model a specific correlation is defined in accordance with the Markovian assumption and the used weights. In practice, the covariance of innovations, which is derived from natural textures or from natural textures' wavelet coefficients, does not necessarily follow the SAR or the GMRF models. This is mainly a result of inappropriate choice of the neighbor set. Chellappa and Kashyap [27,36] showed that the quality of an image which is synthesized from its SAR or GMRF models varies considerably depending on the use of appropriate neighbor set, since different neighbor sets account for different textural patterns. Furthermore, correlation between image pixels is prone to change when a non-linearity is applied to the image, rendering it difficult to choose the appropriate set of neighboring pixels. The squaring non-linearity achieves an RFM with lower prediction error variance whenever a proper scaling of the modeled image is fulfilled (see Appendix A). Yet, the resulting GMRF covariance differs from the actual covariance of the innovations. We demonstrate this in Fig. 1, using the stone texture shown in Fig. 2. We observe that the GMRF covariance exhibits better resemblance to the actual innovations covariance when the squaring non-linearity is applied. In order to overcome this inconsistency, we formulate and use a more generic RFM, taking into consideration the correlation between pixels which is not accounted for in the GMRF model. This moderates the need for a proper choice of neighborhood for each background texture and yields a more robust detection algorithm when combined with the proposed squaring non-linearity.

The infinite lattice RFM for a stationary image $y(\mathbf{v})$ with a zero mean is given by

$$y(\mathbf{v}) = \sum_{\mathbf{r} \in \mathcal{R}} \theta(\mathbf{r})y(\mathbf{v} + \mathbf{r}) + \varepsilon(\mathbf{v}), \quad (3)$$

where \mathcal{R} denotes a given set of indices representing the neighborhood of a pixel, $\theta(\mathbf{r})$ denotes the weight coefficient of a neighbor $\mathbf{r} \in \mathcal{R}$ and $\varepsilon(\mathbf{v})$ is an additive spatially correlated Gaussian noise sequence with zero mean and covariance Λ . The weight coefficients $\{\theta(\mathbf{r})\}_{\mathbf{r} \in \mathcal{R}}$ and the innovations covariance Λ are unknown parameters that need to be estimated.

Let $\{y(\mathbf{v})\}_{\mathbf{v} \in \Omega}$ denote a finite stationary image, defined on an $M \times M$ toroidal lattice Ω . Eq. (3) can

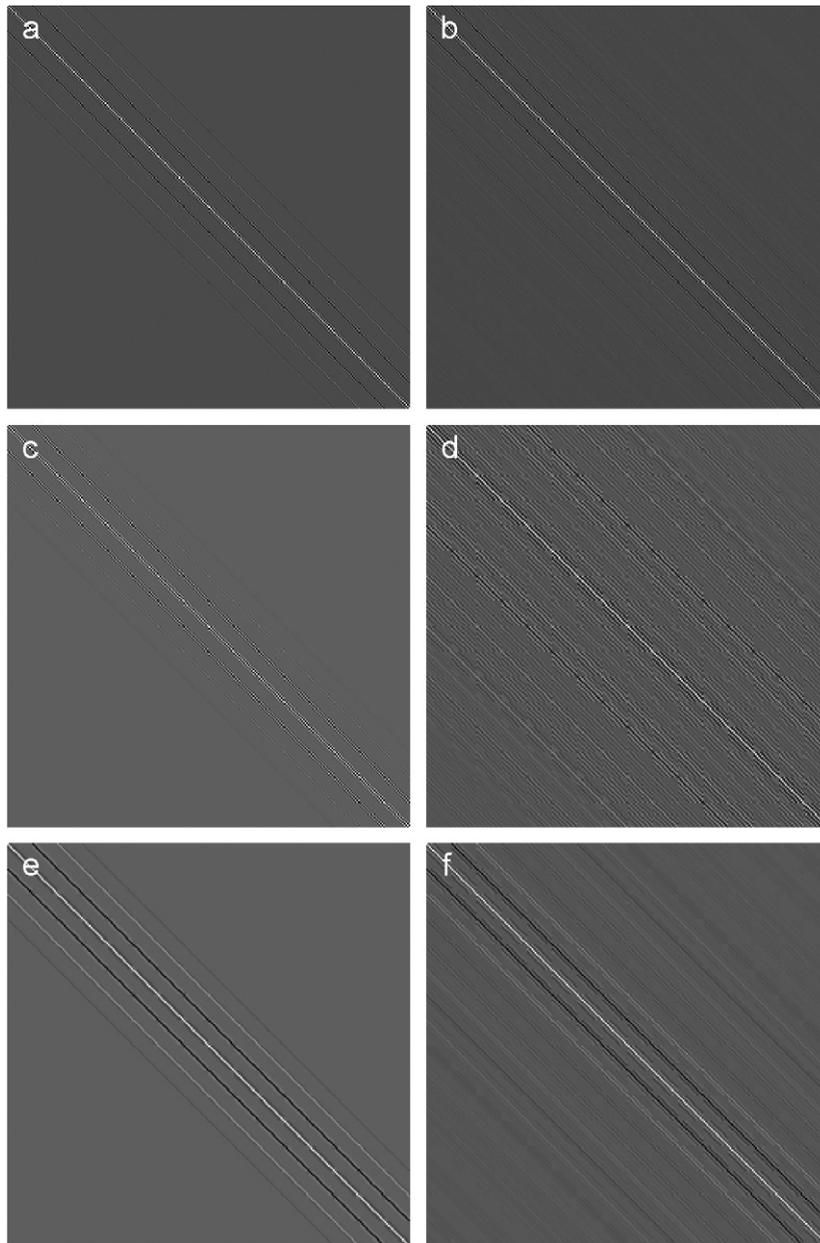


Fig. 1. Innovations covariance example. GMRF (left column) vs. actual (right column). (a and b) Derived directly from the background texture. (c and d) Derived from a layer of wavelet coefficients. (e and f) Derived from the same layer of wavelet coefficients after applying the squaring non-linearity.

then be rewritten as follows:

$$\mathbf{B}(\boldsymbol{\theta})\mathbf{y} = \boldsymbol{\varepsilon}, \quad (4)$$

where

$$\begin{aligned} \mathbf{y} &= \text{col}[y(\mathbf{v}), \mathbf{v} \in \Omega], \\ \boldsymbol{\varepsilon} &= \text{col}[\varepsilon(\mathbf{v}), \mathbf{v} \in \Omega], \\ \boldsymbol{\theta} &= \text{col}[\theta(\mathbf{r}), \mathbf{r} \in \mathcal{R}], \end{aligned} \quad (5)$$

and $\text{col}[\cdot]$ denotes an operator which organizes the entries in a column vector. The toroidal lattice determines the boundaries of the image such $y(\mathbf{v} + (M, M)) = y(\mathbf{v})$. The $M^2 \times M^2$ matrix $\mathbf{B}(\boldsymbol{\theta})$ is a block-circulant matrix which contains the weight coefficients. The neighborhood \mathcal{R} can be arbitrary as long as $(0, 0) \notin \mathcal{R}$ and the resulting block-circulant matrix $\mathbf{B}(\boldsymbol{\theta})$ is not singular. As such,

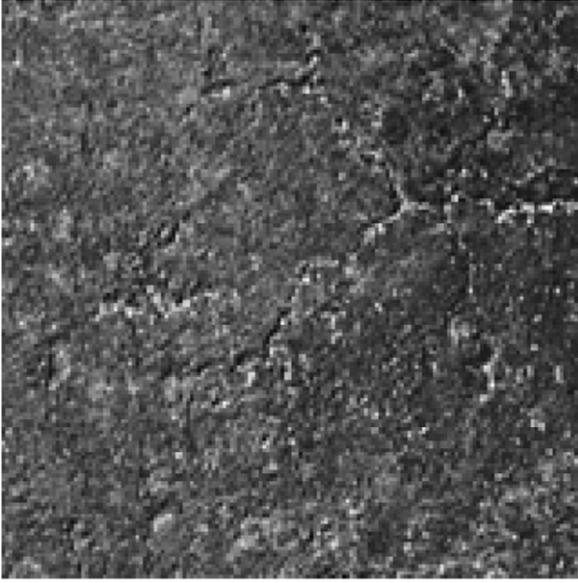


Fig. 2. Natural texture of a stone.

we require

$$\lambda(\mathbf{v}, \boldsymbol{\theta}) = 1 - \sum_{\mathbf{r} \in \mathcal{R}} \theta(\mathbf{r}) \cos\left(r_1 \frac{2\pi v_1}{M} + r_2 \frac{2\pi v_2}{M}\right) \neq 0, \quad \forall \mathbf{v} \in \Omega, \quad (6)$$

where $\lambda(\mathbf{v}, \boldsymbol{\theta})$ denotes the eigenvalues of matrix $\mathbf{B}(\boldsymbol{\theta})$ [37,47], $\mathbf{v} = (v_1, v_2)$ and $\mathbf{r} = (r_1, r_2)$. A symmetric neighborhood \mathcal{R} results in a symmetric matrix $\mathbf{B}(\boldsymbol{\theta})$ as well.

The problem of estimating the parameters of the SAR and the GMRF models was previously addressed in [5,9,25,36,39,48]. A computationally efficient method for estimating the weight coefficients is derived using the least-squares (LS) approach. The weight coefficients LS estimator (LSE) is given by [9,36]:

$$\hat{\boldsymbol{\theta}} = \left[\sum_{\mathbf{v} \in \Omega} \mathbf{w}(\mathbf{v}) \mathbf{w}(\mathbf{v})^T \right]^{-1} \left[\sum_{\mathbf{v} \in \Omega} y(\mathbf{v}) \mathbf{w}(\mathbf{v}) \right], \quad (7)$$

where

$$\mathbf{w}(\mathbf{v}) = \text{col}[y(\mathbf{v} + \mathbf{r}), \mathbf{r} \in \mathcal{R}]. \quad (8)$$

Based on Eq. (7), the estimated weight coefficients are not affected by the marginal variance of the image, but rather on the spatial interaction between neighboring pixels. The covariance $\boldsymbol{\Lambda}$ is then given by

$$\boldsymbol{\Lambda} = E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] \approx \mathbf{B}(\hat{\boldsymbol{\theta}}) \boldsymbol{\Sigma} \mathbf{B}(\hat{\boldsymbol{\theta}})^T, \quad (9)$$

where $\boldsymbol{\Sigma} = E[\mathbf{y} \mathbf{y}^T]$.

Conventional RFM, such as the SAR, GMRF and the proposed RFM, are intended to be used with micro-textures. As such, these models may not sufficiently describe the background clutter in images with periodical patterns of period lengths larger than the neighborhood \mathcal{R} . However, they should be appropriate for multi-resolution decomposition layers of such images [25].

3.3. Multi-resolution feature space

Let Ω be the $M \times M$ support lattice for a mean normalized image $\{y(\mathbf{v})\}_{\mathbf{v} \in \Omega}$. Let $\{y_j(\mathbf{v})\}_{j=1, \dots, m}$ denote the j th layer wavelet coefficients obtained from the image using an RDWT with $(m-1)/3$ levels. Let $\mathbf{y}(\mathbf{v})$ denote the resulting multi-resolution image with m layers. The multi-resolution image $\mathbf{y}(\mathbf{v})$ is defined over the same support as image $y(\mathbf{v})$ and is given by

$$\mathbf{y}(\mathbf{v}) = [y_1(\mathbf{v}), y_2(\mathbf{v}), \dots, y_m(\mathbf{v})]^T. \quad (10)$$

The wavelet coefficients at different layers are nearly de-correlated for most images and the transform can be thought of as an approximation for the Karhunen–Loève transform (KLT) [49]. We utilize the KLT as a mean for reducing the feature space, using only relevant layers in the detection scheme. Let \mathbf{K} denote a matrix whose columns are p eigenvectors taken from the covariance matrix of $\mathbf{y}(\mathbf{v})$, $\forall \mathbf{v} \in \Omega$. The KLT result, denoted as $\mathbf{t}(\mathbf{v})$, is given by

$$\mathbf{t}(\mathbf{v}) = \mathbf{K}^T \mathbf{y}(\mathbf{v}) = [t_1(\mathbf{v}), t_2(\mathbf{v}), \dots, t_p(\mathbf{v})]^T, \quad 1 \leq p \leq m, \quad (11)$$

where $\{t_k(\mathbf{v})\}_{k=1, \dots, p}$ denote the generated uncorrelated image layers. The resulting local energy measures, denoted as $\mathbf{z}(\mathbf{v})$, are given by

$$\mathbf{z}(\mathbf{v}) = [z_1(\mathbf{v}), z_2(\mathbf{v}), \dots, z_p(\mathbf{v})]^T, \quad (12)$$

where

$$z_k(\mathbf{v}) = t_k^2(\mathbf{v}), \quad k = 1, \dots, p. \quad (13)$$

We assume that each layer in $\mathbf{z}(\mathbf{v})$ is statistically homogeneous and follows the 2-D RFM. Therefore, the k th layer satisfies (Eq. (3)):

$$z_k(\mathbf{v}) - \mu_{z_k} = \sum_{\mathbf{r} \in \mathcal{R}} \theta_k(\mathbf{r}) [z_k(\mathbf{v} + \mathbf{r}) - \mu_{z_k}] + \varepsilon_k(\mathbf{v}), \quad (14)$$

where \mathcal{R} denotes the neighborhood, $\theta_k(\mathbf{r})$ denote the weight coefficients, μ_{z_k} denotes the expected value of $z_k(\mathbf{v})$ and $\varepsilon_k(\mathbf{v})$ denotes the resulting innovations

process. By applying the 2-D RFM to all layers using model parameters which are estimated for each layer separately, we produce a multi-resolution RFM. The multi-resolution RFM innovations, denoted as $\boldsymbol{\varepsilon}(\mathbf{v})$, are then given by

$$\begin{aligned}\boldsymbol{\varepsilon}(\mathbf{v}) &= [\varepsilon_1(\mathbf{v}), \varepsilon_2(\mathbf{v}), \dots, \varepsilon_p(\mathbf{v})]^T \\ &= \left(\mathbf{z}(\mathbf{v}) - \sum_{\mathbf{r} \in \mathcal{R}} \boldsymbol{\Theta}(\mathbf{r}) \mathbf{z}(\mathbf{v} + \mathbf{r}) \right) \\ &\quad - \left(\mathbf{I} - \sum_{\mathbf{r} \in \mathcal{R}} \boldsymbol{\Theta}(\mathbf{r}) \right) \boldsymbol{\mu},\end{aligned}\quad (15)$$

where

$$\boldsymbol{\mu} \triangleq [\mu_{z_1}, \mu_{z_2}, \dots, \mu_{z_p}]^T, \quad (16)$$

$$\boldsymbol{\Theta}(\mathbf{r}) \triangleq \text{diag}(\theta_1(\mathbf{r}), \theta_2(\mathbf{r}), \dots, \theta_p(\mathbf{r})). \quad (17)$$

We use the multi-resolution RFM innovations (formulated in Eq. (15)) as the feature space for the proposed anomaly detection algorithm.

4. Anomaly detection algorithm

Scharf and Friedlander introduced in [24] the MSD, formulating a class of problems for detecting subspace signals in subspace interference and broadband white Gaussian noise, laying the framework for detecting anomalies which are assumed to lie within a known subspace. Kraut et al. [26] formulated an MSD for detecting subspace signals in colored Gaussian noise with a known covariance structure. Goldman and Cohen [25] have further improved the MSD scheme and formulated a multi-resolution MSD for the detection of subspace signals in subspace interference and colored Gaussian noise with a known covariance structure, corresponding to the GMRF innovations. Here we develop a multi-resolution MSD for detecting subspace signals in subspace interference and colored Gaussian noise, corresponding to the multi-resolution RFM innovations that were presented in Section 3.3.

A block diagram of the proposed algorithm is presented in Fig. 3. Let $\{y(\mathbf{v}, \mathbf{s})\}_{\mathbf{s} \in \Omega_0}$ denote a $N \times N$ image chip, taken from image $\{y(\mathbf{v})\}_{\mathbf{v} \in \Omega}$ around the spatial location \mathbf{v} . Image chips under hypotheses H_0 and H_1 are given by

$$\begin{aligned}H_0 : y(\mathbf{v}, \mathbf{s}) &= x(\mathbf{v}, \mathbf{s}) + g(\mathbf{v}, \mathbf{s}), \\ H_1 : y(\mathbf{v}, \mathbf{s}) &= x(\mathbf{v}, \mathbf{s}) + g(\mathbf{v}, \mathbf{s}) + h(\mathbf{v}, \mathbf{s}),\end{aligned}\quad (18)$$

where $x(\mathbf{v}, \mathbf{s})$, $g(\mathbf{v}, \mathbf{s})$ and $h(\mathbf{v}, \mathbf{s})$ are the image chips of the background clutter, interference signal and target signal, respectively, around the spatial location \mathbf{v} . We assume that the chip $x(\mathbf{v}, \mathbf{s})$ conforms to a multi-resolution RFM with model parameters that were derived from the whole image. We further assume that the chip size is sufficient for containing the innovations of shapes which span the target and interference signals, taking into consideration RDWT and RFM margins. Therefore, the multi-resolution chip decomposition, denoted as $\mathbf{y}(\mathbf{v}, \mathbf{s})$, is given by

$$\begin{aligned}H_0 : \mathbf{y}(\mathbf{v}, \mathbf{s}) &= \mathbf{x}(\mathbf{v}, \mathbf{s}) + \mathbf{g}(\mathbf{v}, \mathbf{s}), \\ H_1 : \mathbf{y}(\mathbf{v}, \mathbf{s}) &= \mathbf{x}(\mathbf{v}, \mathbf{s}) + \mathbf{g}(\mathbf{v}, \mathbf{s}) + \mathbf{h}(\mathbf{v}, \mathbf{s}),\end{aligned}\quad (19)$$

where $\mathbf{x}(\mathbf{v}, \mathbf{s})$, $\mathbf{g}(\mathbf{v}, \mathbf{s})$ and $\mathbf{h}(\mathbf{v}, \mathbf{s})$ are the multi-resolution chip decompositions of the background clutter, interference signal and target signal, respectively. Let $z_k(\mathbf{v}, \mathbf{s})$ denote the chip's k th layer local energy measures, which under hypotheses H_0 and H_1 is given by

$$\begin{aligned}H_0 : z_k(\mathbf{v}, \mathbf{s}) &= t_k^2(\mathbf{x}(\mathbf{v}, \mathbf{s})) + t_k^2(\mathbf{g}(\mathbf{v}, \mathbf{s})) \\ &\quad + 2t_k(\mathbf{x}(\mathbf{v}, \mathbf{s}))t_k(\mathbf{g}(\mathbf{v}, \mathbf{s})), \\ H_1 : z_k(\mathbf{v}, \mathbf{s}) &= t_k^2(\mathbf{x}(\mathbf{v}, \mathbf{s})) + t_k^2(\mathbf{g}(\mathbf{v}, \mathbf{s})) \\ &\quad + 2t_k(\mathbf{x}(\mathbf{v}, \mathbf{s}))t_k(\mathbf{g}(\mathbf{v}, \mathbf{s})) + t_k^2(\mathbf{h}(\mathbf{v}, \mathbf{s})) \\ &\quad + 2t_k(\mathbf{x}(\mathbf{v}, \mathbf{s}))t_k(\mathbf{h}(\mathbf{v}, \mathbf{s})) \\ &\quad + 2t_k(\mathbf{g}(\mathbf{v}, \mathbf{s}))t_k(\mathbf{h}(\mathbf{v}, \mathbf{s})),\end{aligned}\quad (20)$$

where $t_k(\cdot) = [\mathbf{t}(\cdot)]_k$ and

$$\begin{aligned}\mathbf{t}(\mathbf{x}(\mathbf{v}, \mathbf{s})) &= \mathbf{K}^T \mathbf{x}(\mathbf{v}, \mathbf{s}), \\ \mathbf{t}(\mathbf{g}(\mathbf{v}, \mathbf{s})) &= \mathbf{K}^T \mathbf{g}(\mathbf{v}, \mathbf{s}), \\ \mathbf{t}(\mathbf{h}(\mathbf{v}, \mathbf{s})) &= \mathbf{K}^T \mathbf{h}(\mathbf{v}, \mathbf{s}).\end{aligned}\quad (21)$$

Let $\mathbf{n}_k(\mathbf{v})$ denote the column stack representation of the chip's k th layer RFM innovations. Let $\mathbf{z}_k(\mathbf{v})$ denote the column stack representation of $z_k(\mathbf{v}, \mathbf{s})$, the chip's k th layer local energy measures. Based on (4), (5) and (14), $\mathbf{n}_k(\mathbf{v})$ is given by

$$\mathbf{n}_k(\mathbf{v}) = \mathcal{D}_k \mathbf{z}_k(\mathbf{v}) - \mu_{z_k} \left(1 - \sum_{\mathbf{r} \in \mathcal{R}} \theta_k(\mathbf{r}) \right), \quad (22)$$

where $\mathcal{D}_k \triangleq \mathbf{B}(\theta_k(\mathbf{r}))$. From (20), $\mathbf{z}_k(\mathbf{v})$ and consequently $\mathcal{D}_k \mathbf{z}_k(\mathbf{v})$ contain derived components from the background texture, interference signals, target signals and the interaction among them. Hence, proper formulation of an MSD allows detection of target signals within $\mathbf{n}_k(\mathbf{v})$.

Let $\{h^{(l)}(\mathbf{s})\}_{\mathbf{s} \in \Omega_0, l=1, \dots, u_h}$ and $\{g^{(l)}(\mathbf{s})\}_{\mathbf{s} \in \Omega_0, l=1, \dots, u_g}$ denote sets of orthogonal image chips which span

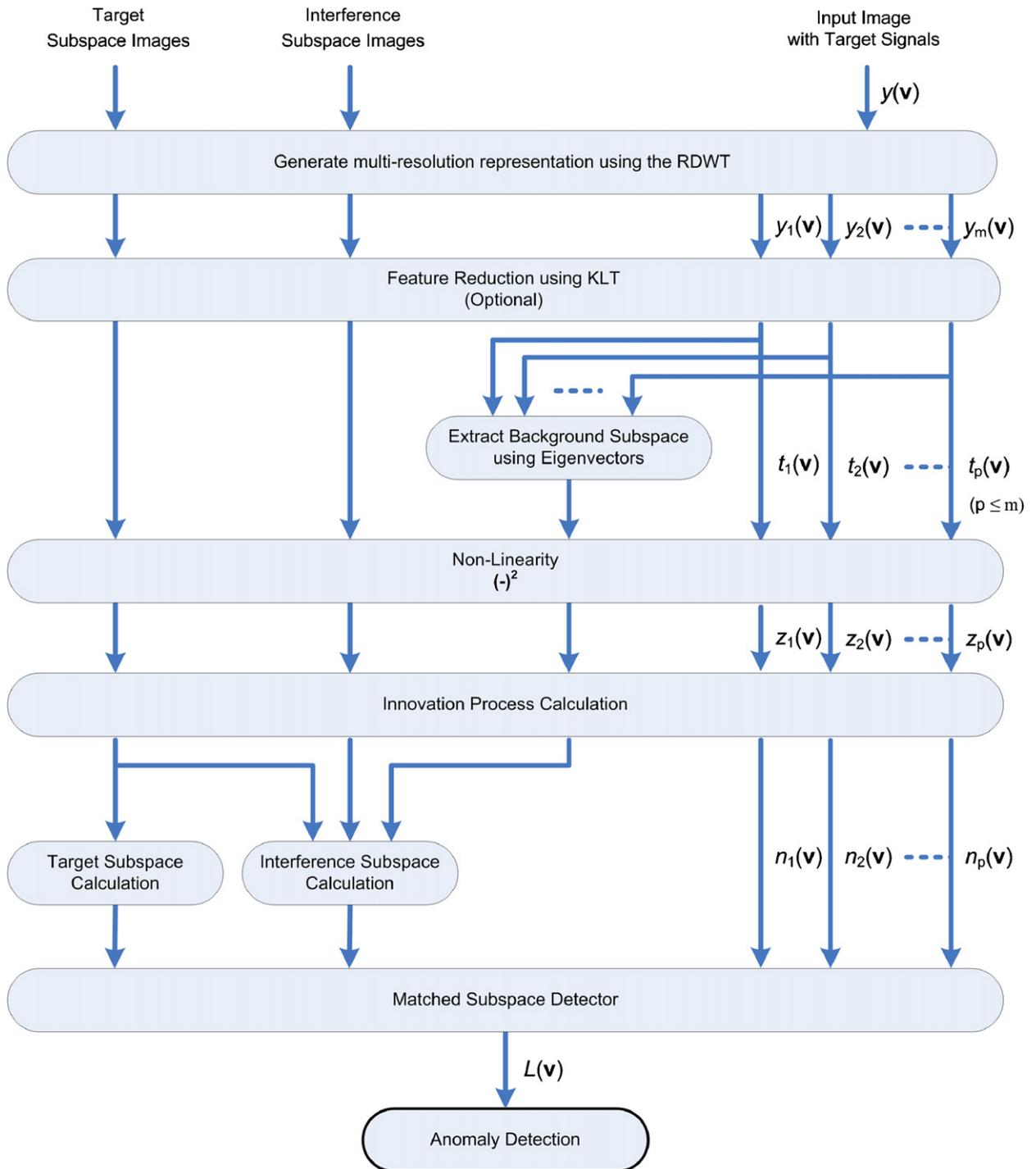


Fig. 3. Block diagram of the proposed anomaly detection algorithm.

the target and interference signals, respectively. Let $\{x_j^{(l)}(\mathbf{s})\}_{\mathbf{s} \in \Omega_0, l=1, \dots, u_x}$ denote a set of image chips which characterize the j th layer of the background natural texture multi-resolution decomposition. We use u_x significant eigenvectors, derived from the

estimated covariance matrix of a j th layer chip. Let $\mathbf{h}^{(l)}(\mathbf{s})$ and $\mathbf{g}^{(l)}(\mathbf{s})$ denote the multi-resolution chip decomposition of $h^{(l)}(\mathbf{s})$ and $g^{(l)}(\mathbf{s})$, respectively, and let $\mathbf{x}^{(l)}(\mathbf{s}) = [x_1^{(l)}(\mathbf{s}), x_2^{(l)}(\mathbf{s}), \dots, x_m^{(l)}(\mathbf{s})]^T$. We define $\langle \mathcal{A}_k \rangle$ and $\langle \mathcal{B}_k \rangle$ as the k th layer MSD subspaces,

each spanned by the columns of the full-rank matrices \mathcal{A}_k and \mathcal{B}_k , respectively [24]. We define $\gamma_k(\mathbf{v})$ as the whitened measurements, derived from $\mathbf{n}_k(\mathbf{v})$ as follows:

$$\gamma_k(\mathbf{v}) = \Lambda_k^{-1/2} \mathbf{n}_k(\mathbf{v}), \quad (23)$$

where under the assumption of sparsely placed target signals and interfering signals

$$\Lambda_k \approx E[\mathbf{n}_k(\mathbf{v})\mathbf{n}_k(\mathbf{v})^T]. \quad (24)$$

Based on (22) and (23), $\gamma_k(\mathbf{v})$ is formulated under hypotheses H_0 and H_1 as follows:

$$\begin{aligned} H_0 : \gamma_k(\mathbf{v}) &= \mathcal{A}_k \phi_k(\mathbf{v}) + \mathbf{\varepsilon}_k(\mathbf{v}), \\ H_1 : \gamma_k(\mathbf{v}) &= \mathcal{B}_k \psi_k(\mathbf{v}) + \mathcal{A}_k \phi_k(\mathbf{v}) + \mathbf{\varepsilon}_k(\mathbf{v}), \end{aligned} \quad (25)$$

where $\mathbf{\varepsilon}_k(\mathbf{v})$ denotes the whitened RFM innovations which are derived from the chip's background texture and $\phi_k(\mathbf{v})$ and $\psi_k(\mathbf{v})$ are the coordinates of the interference and target signals components with respect to subspaces \mathcal{A}_k and \mathcal{B}_k , respectively.

The interfering subspace \mathcal{A}_k should account for the interfering signals, the interaction among them and the interaction between them and the background clutter. Therefore, subspace \mathcal{A}_k is the span of

$$\begin{aligned} & [\Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\alpha}_k(\mathbf{g}^{(1)}(\mathbf{s})), & \dots & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\alpha}_k(\mathbf{g}^{(u_g)}(\mathbf{s})), \\ & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(1)}(\mathbf{s}), \mathbf{g}^{(2)}(\mathbf{s})), & \dots & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(1)}(\mathbf{s}), \mathbf{g}^{(u_g)}(\mathbf{s})), \\ & \vdots & & \vdots \\ & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(u_g-2)}(\mathbf{s}), \mathbf{g}^{(u_g-1)}(\mathbf{s})), & \dots & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(u_g-2)}(\mathbf{s}), \mathbf{g}^{(u_g)}(\mathbf{s})), \\ & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(u_g-1)}(\mathbf{s}), \mathbf{g}^{(u_g)}(\mathbf{s})), & & \\ & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(1)}(\mathbf{s}), \mathbf{x}^{(1)}(\mathbf{s})), & \dots & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(u_g)}(\mathbf{s}), \mathbf{x}^{(1)}(\mathbf{s})), \\ & \vdots & & \vdots \\ & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(1)}(\mathbf{s}), \mathbf{x}^{(u_x)}(\mathbf{s})), & \dots & \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\beta}_k(\mathbf{g}^{(u_g)}(\mathbf{s}), \mathbf{x}^{(u_x)}(\mathbf{s})), \end{aligned} \quad (26)$$

where $\boldsymbol{\alpha}_k(\cdot)$ is the column stack representation of $\alpha_k(\cdot)$, given by

$$\boldsymbol{\alpha}_k(\mathbf{g}^{(l)}(\mathbf{s})) = [\boldsymbol{\alpha}_k(\mathbf{g}^{(l)}(\mathbf{s}))]_s \triangleq t_k^2(\mathbf{g}^{(l)}(\mathbf{s})), \quad (27)$$

$\boldsymbol{\beta}_k(\cdot, \cdot)$ is the column stack representation of $\beta_k(\cdot, \cdot)$, given by

$$\begin{aligned} \boldsymbol{\beta}_k(\mathbf{g}^{(l_1)}(\mathbf{s}), \mathbf{g}^{(l_2)}(\mathbf{s})) &= [\boldsymbol{\beta}_k(\mathbf{g}^{(l_1)}(\mathbf{s}), \mathbf{g}^{(l_2)}(\mathbf{s}))]_s \\ &\triangleq t_k(\mathbf{g}^{(l_1)}(\mathbf{s})) t_k(\mathbf{g}^{(l_2)}(\mathbf{s})), \end{aligned}$$

$$\begin{aligned} \boldsymbol{\beta}_k(\mathbf{g}^{(l_1)}(\mathbf{s}), \mathbf{x}^{(l_2)}(\mathbf{s})) &= [\boldsymbol{\beta}_k(\mathbf{g}^{(l_1)}(\mathbf{s}), \mathbf{x}^{(l_2)}(\mathbf{s}))]_s \\ &\triangleq t_k(\mathbf{g}^{(l_1)}(\mathbf{s})) t_k(\mathbf{x}^{(l_2)}(\mathbf{s})) \end{aligned} \quad (28)$$

and $t_k(\cdot)$ is defined in (20) and (21).

The target subspace \mathcal{B}_k should account for the target signals, the interaction among them, the interaction between them and the background clutter and the interaction between them and the interfering signals. However, this formulation of the target subspace \mathcal{B}_k increases dramatically its dimensionality. We recall that the MSD detection performance is adversely affected by the dimensionality of its target subspace [3,4]. Therefore, we assume that the interaction among target signals, the interaction between them and the background clutter and the interaction between them and the interfering signals are negligible, searching only for a particular signature under hypothesis H_1 . Hence, subspace \mathcal{B}_k is the span of

$$[\Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\alpha}_k(\mathbf{h}^{(1)}(\mathbf{s})), \dots, \Lambda_k^{-1/2} \mathcal{D}_k \boldsymbol{\alpha}_k(\mathbf{h}^{(u_h)}(\mathbf{s}))], \quad (29)$$

where $\boldsymbol{\alpha}_k(\cdot)$ is defined in (27). We note that under this formulation of subspaces, scaled image sets can be used for the purpose of characterizing the target and interference signals in spite of the introduced

non-linearity that is used in the process of feature space creation.

Based on (25) and subspace formulations (26) and (29), we formulate the detection problem as a likelihood ratio test (LRT) between hypotheses H_0 and H_1 . The likelihood ratio (LR) is given by

$$L_k(\mathbf{v}) \triangleq 2 \log \frac{P(\boldsymbol{\varepsilon}_k(\mathbf{v})|H_1)}{P(\boldsymbol{\varepsilon}_k(\mathbf{v})|H_0)}. \quad (30)$$

We assume that the whitened innovations $\mathbf{e}_k(\mathbf{v})$ follow the 2-D RFM and therefore $\mathbf{e}_k(\mathbf{v}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. We derive the generalized LR (GLR) from (30) using the maximum likelihood (ML) estimates of $\mathbf{e}_k(\mathbf{v})$ under hypotheses H_0 and H_1 . Let $\mathbf{P}_{\mathcal{A}_k}$ and $\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k}$ denote the projection operators onto subspaces $\langle \mathcal{A}_k \rangle$ and $\langle \mathcal{A}_k, \mathcal{B}_k \rangle$, respectively. Subspace $\langle \mathcal{A}_k \rangle$ is spanned by the columns of matrix \mathcal{A}_k and subspace $\langle \mathcal{A}_k, \mathcal{B}_k \rangle$ is spanned by the columns of the concatenated matrix $[\mathcal{A}_k, \mathcal{B}_k]$. The projection operators $\mathbf{P}_{\mathcal{A}_k}$ and $\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k}$ are given by [4,24]:

$$\begin{aligned} \mathbf{P}_{\mathcal{A}_k} &= [\mathcal{A}_k][[\mathcal{A}_k]^T[\mathcal{A}_k]]^{-1}[\mathcal{A}_k]^T, \\ \mathbf{P}_{\mathcal{A}_k\mathcal{B}_k} &= [\mathcal{A}_k, \mathcal{B}_k][[\mathcal{A}_k, \mathcal{B}_k]^T[\mathcal{A}_k, \mathcal{B}_k]]^{-1} \\ &\quad \times [\mathcal{A}_k, \mathcal{B}_k]^T. \end{aligned} \quad (31)$$

Let $\widehat{\mathbf{e}}_k^{H_0}(\mathbf{v})$ and $\widehat{\mathbf{e}}_k^{H_1}(\mathbf{v})$ denote the ML estimates of $\mathbf{e}_k(\mathbf{v})$ under hypotheses H_0 and H_1 , respectively. These estimates are obtained by subtracting the components which lie in the target and interference derived subspaces from the sample vector $\gamma_k(\mathbf{v})$ as follows:

$$\begin{aligned} \widehat{\mathbf{e}}_k^{H_0}(\mathbf{v}) &= (\mathbf{I} - \mathbf{P}_{\mathcal{A}_k})\gamma_k(\mathbf{v}) \triangleq \mathbf{P}_{\mathcal{A}_k}^\perp \gamma_k(\mathbf{v}), \\ \widehat{\mathbf{e}}_k^{H_1}(\mathbf{v}) &= (\mathbf{I} - \mathbf{P}_{\mathcal{A}_k\mathcal{B}_k})\gamma_k(\mathbf{v}) \triangleq \mathbf{P}_{\mathcal{A}_k\mathcal{B}_k}^\perp \gamma_k(\mathbf{v}), \end{aligned} \quad (32)$$

where \perp denotes the projection operator onto the orthogonal complement subspace, i.e., $(\mathbf{P}_{\mathcal{A}_k}\gamma_k(\mathbf{v}))^T (\mathbf{P}_{\mathcal{A}_k}^\perp \gamma_k(\mathbf{v})) = 0$. Based on (30)–(32), the k th layer GLR is given by

$$L_k(\mathbf{v}) = \widehat{\mathbf{e}}_k^{H_0}(\mathbf{v})^T \widehat{\mathbf{e}}_k^{H_0}(\mathbf{v}) - \widehat{\mathbf{e}}_k^{H_1}(\mathbf{v})^T \widehat{\mathbf{e}}_k^{H_1}(\mathbf{v}) \quad (33)$$

and can be regarded as the difference between the Mahalanobis distances under each hypotheses. We develop Eq. (33) as follows:

$$\begin{aligned} L_k(\mathbf{v}) &= \|\widehat{\mathbf{e}}_k^{H_0}(\mathbf{v})\|^2 - \|\widehat{\mathbf{e}}_k^{H_1}(\mathbf{v})\|^2, \\ &= \gamma_k(\mathbf{v})^T (\mathbf{P}_{\mathcal{A}_k}^\perp - \mathbf{P}_{\mathcal{A}_k\mathcal{B}_k}^\perp) \gamma_k(\mathbf{v}), \\ &= \gamma_k(\mathbf{v})^T (\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k} - \mathbf{P}_{\mathcal{A}_k}) \gamma_k(\mathbf{v}). \end{aligned} \quad (34)$$

We define the MSD's k th layer SNR as the ratio between the energy of the signal which does not lie in the interference subspace and the background innovations covariance in the k th layer. Therefore,

$$\begin{aligned} \text{SNR}(k, \mathbf{v}) &\triangleq [\mathcal{B}_k \boldsymbol{\psi}_k(\mathbf{v})]^T \mathbf{P}_{\mathcal{A}_k}^\perp [\mathcal{B}_k \boldsymbol{\psi}_k(\mathbf{v})], \\ &= [\mathcal{B}_k \boldsymbol{\psi}_k(\mathbf{v})]^T (\mathbf{I} - \mathbf{P}_{\mathcal{A}_k}) [\mathcal{B}_k \boldsymbol{\psi}_k(\mathbf{v})]. \end{aligned} \quad (35)$$

The quadratic form of the k th layer GLR may be thought of as the norm-squared of $(\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k} - \mathbf{P}_{\mathcal{A}_k}) \gamma_k(\mathbf{v})$. Hence, $L_k(\mathbf{v})$ is chi-square distributed with r

degrees of freedom [4,24,50]:

$$L_k \sim \begin{cases} \chi_r^2(0) & \text{under } H_0, \\ \chi_r^2(\text{SNR}(k, \mathbf{v})) & \text{under } H_1, \end{cases} \quad (36)$$

where $r = u_h$ and $\text{SNR}(k, \mathbf{v})$ is the non-centrality parameter of the chi-square distribution under hypothesis H_1 . The number of degrees of freedom r is also the rank and trace of the idempotent operator $(\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k} - \mathbf{P}_{\mathcal{A}_k})$ [51].

We define the GLR for image chip $y(\mathbf{v}, \mathbf{s})$, which is derived using all layers, as follows:

$$\begin{aligned} L(\mathbf{v}) &= \sum_{k=1}^p L_k(\mathbf{v}), \\ &= \sum_{k=1}^p \gamma_k(\mathbf{v})^T (\mathbf{P}_{\mathcal{A}_k\mathcal{B}_k} - \mathbf{P}_{\mathcal{A}_k}) \gamma_k(\mathbf{v}). \end{aligned} \quad (37)$$

It was previously mentioned that the wavelet coefficients at different layers are nearly de-correlated for most images [49]. Therefore, the GLR is the sum of p independent chi-square distributions. As such, it is also chi-square distributed with q degrees of freedom,

$$L(\mathbf{v}) \sim \begin{cases} \chi_q^2(0) & \text{under } H_0, \\ \chi_q^2\left(\sum_{k=1}^p \text{SNR}(k, \mathbf{v})\right) & \text{under } H_1, \end{cases} \quad (38)$$

where $q = p \cdot u_h$. The non-centrality parameter of the chi-square distribution is zero under hypothesis H_0 , and it equals $\sum_{k=1}^p \text{SNR}(k, \mathbf{v})$ under hypothesis H_1 . The detection statistic will be maximized when a target is centrally located in the processed image chip.

We define the decision rule for distinguishing between the two hypotheses as a threshold criterion, converting the GLR grayscale image into a binary image. The GLRT is given by

$$L(\mathbf{v}) \underset{H_0}{\overset{H_1}{\geq}} \eta. \quad (39)$$

The threshold parameter η determines both the probability of detection P_D and the probability of false alarm P_{FA} . The detection rate in (2) is given for each image chip $y(\mathbf{v}, \mathbf{s})$ by

$$P_D(\mathbf{v}) = 1 - P\left(\chi_q^2\left(\sum_{k=1}^p \text{SNR}(k, \mathbf{v})\right) \leq \eta\right). \quad (40)$$

The appropriate false alarm is given by

$$P_{FA}(\mathbf{v}) = 1 - P(\chi_q^2(0) \leq \eta). \quad (41)$$

The false alarm rate in (41) is uniquely defined by the threshold η and the dimensionality q . The detection rate in (40) is a monotonically increasing function of the GLR non-centrality parameter for a given false alarm rate and dimensionality q . This is expected since the GLR non-centrality parameter is defined as the sum of the MSD's layers SNR and high SNR values represent a better segregation of the anomalous targets. The detection rate is also a monotonically decreasing function of the dimensionality q for a given false alarm rate and MSD's layers SNR values. This is also expected since the dimensionality represents the available *a priori* information about the targets and this information decreases as we increase the rank of the subspace [4]. We note that the detection and false alarm rates cannot be easily found for the general case due to possible statistical dependency between the different layers and inconsistency with the Gaussian assumption. However, the detector performance can be evaluated using extensive computer simulations, as we present in Section 5. Since not all layers of the feature space usually contribute the same amount of information to the detection process, it may be beneficial to use only a subset of those layers [40], aiming for the reduction of false alarms and dimensionality. Criterion for selecting the subset of layers is application dependent. This selection can be made *a priori*, thus reducing the computational complexity of the proposed method, or it can be made based on in-process data such as layers with highest average SNR, highest point SNR, etc., in which case the decision can only be made after some calculations have been made. A SNR-based criterion is given in Appendix B. Nevertheless, even when utilizing all layers, the proposed scheme outperforms other recently published algorithms, as we present in Section 5.

5. Anomaly detection experimental results

In this section, we study the performance of the proposed algorithm using a large set of background textures. We first explore the effects of the KLT on the performance of the proposed algorithm. Then we explore the GLR non-centrality parameter of the proposed algorithm, compared to its RDWT equivalent. We investigate the receiver operating characteristics (ROC) curves and the MSD performance of the proposed algorithm, compared to recently published work [25,45]. Finally, we demonstrate the performance of the proposed

algorithm using a bank of low-dimensional subspace detectors.

5.1. Data generation

We have qualitatively investigated the performance of the proposed algorithm using synthesized anomalous targets, randomly placed in a set of 40 Brodatz-like textures [41]. The synthesized targets were randomly created in a subspace spanned by the image chips in Fig. 4 and were scaled to achieve desired SNR values. Actual performance statistics were derived using 250 synthesized anomalous targets for each of the 40 Brodatz-like textures. No interference subspace is assumed. The background textures were scaled to values within $[0, 1]$ and then normalized to a zero mean, prior to anomaly insertion. We have used the following algorithm parameters:

- A neighbor set concurrent with N_{S7} in [27], given by $\mathcal{R} = \{-\mathcal{R}_h, \mathcal{R}_h\}$, where $\mathcal{R}_h = \{(0, 1), (1, 0), (1, -1), (1, 1), (0, 2), (2, 0), (-2, 1), (2, 1), (1, 2), (-1, 2), (2, 2), (-2, 2), (3, 0), (0, 3), (1, 3), (3, 1), (-1, 3), (-3, 1)\}$.
- An image lattice size of $M \times M$, where $M = 160$.
- A chip size of $N \times N$, where $N = 16$.
- An RDWT decomposition with two levels, using the Symlet wavelets with eight taps.

The image lattice size and chip size are adequate for assuming $\widehat{\Lambda}_k \simeq \Lambda_k$ when using the ML estimator for the innovations covariance [52]. RFM parameters were estimated as described in Section 3.3. Anomaly detection was then performed using information from all layers, as detailed in Section 4.

5.2. SNR and SNER

The SNR and signal-to-noise energy ratio (SNER) of a single resolution analysis provide a reference point for investigating the performance of

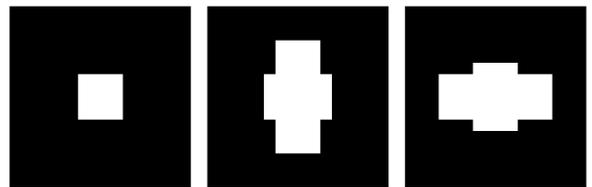


Fig. 4. Image chips used for the target subspace. Each chip has 16×16 pixels.

multi-resolution algorithms. We define the SNR as the ratio between the energy of the anomalous target signal and the energy of the background clutter with respect to the background clutter covariance. Based on (18), the SNR around the spatial location \mathbf{v} is given by

$$\text{SNR}(\mathbf{v}) = \|\Gamma^{-1/2}\mathbf{h}(\mathbf{v})\|^2, \quad (42)$$

where $\mathbf{h}(\mathbf{v})$ denotes the column stack representation of the target chip $h(\mathbf{v}, \mathbf{s})$ and Γ denotes the background clutter covariance of a background image chip. We define the SNER as the ratio between the energy of the anomalous target signal and the energy of the background clutter. Using (18), the SNER around the spatial location \mathbf{v} is given by

$$\text{SNER}(\mathbf{v}) = \|\mathbf{h}(\mathbf{v})\|^2 / \|\mathbf{x}(\mathbf{v})\|^2, \quad (43)$$

where $\mathbf{x}(\mathbf{v})$ denotes the column stack representation of the background chip $x(\mathbf{v}, \mathbf{s})$. Under the framework of synthesized anomalous targets and Brodatz-like textures, using the algorithm parameters as stated above, the resulting averaged SNER value for a given SNR value is: $\text{SNER} \approx \text{SNR} - 26.3$ (dB). This is the outcome of the correlation between pixels in the background clutter, indicating that the use of a RFM, such as the proposed RFM, is in place. The proposed detection scheme utilizes this spatial information of both anomalous targets and background clutter, further improving the SNR. Therefore, the proposed detection scheme outperforms energy-based detectors which ignore available spatial information.

5.3. Performance analysis

In [25], it was argued that the use of eigenvectors associated with the highest eigenvalues yields better detection results, whereas in [8,10] it was argued that rare anomalous targets reside in the subspace spanned by eigenvectors associated with the lowest eigenvalues. Chang and Chiang have shown in [8] that the RX detector [1] is mainly affected by data associated with smaller eigenvalues. Since both the RX detector and the MSD use a similar Bayesian

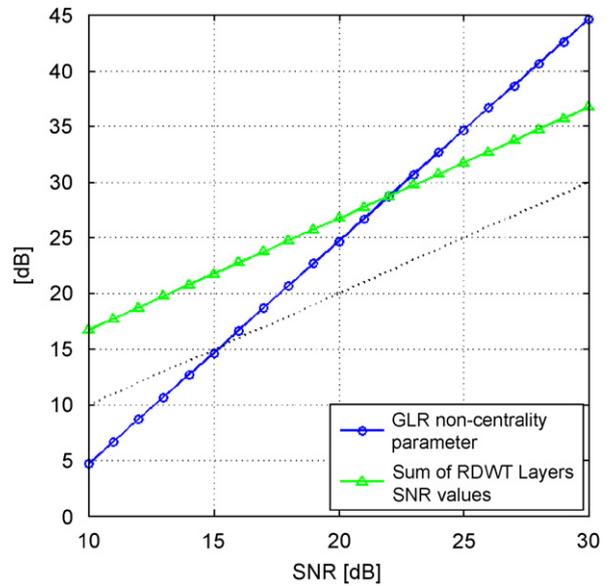


Fig. 6. GLR non-centrality parameter vs. the sum of RDWT layers SNR values, under the hypothesis H_1 . Results are averaged over a set of 40 Brodatz-like textures.

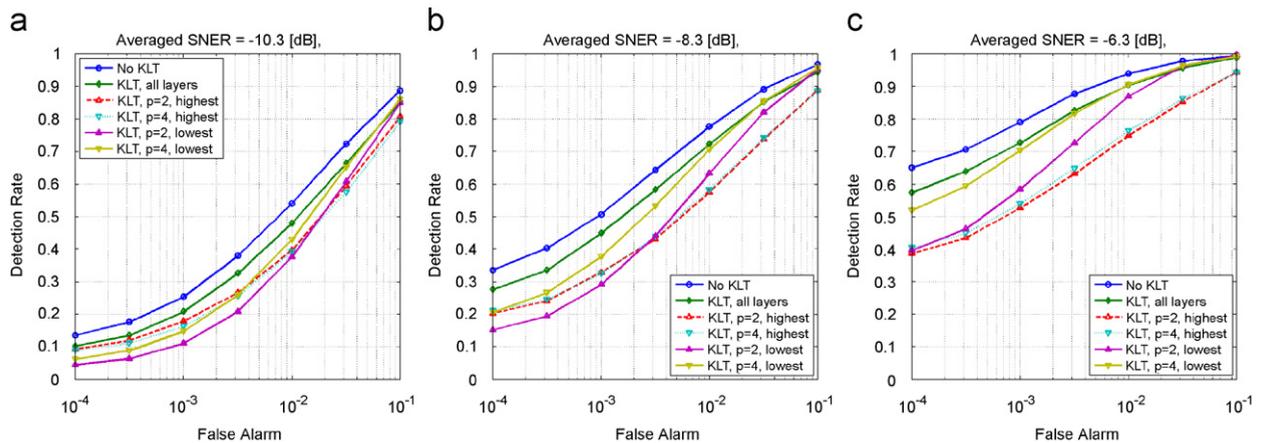


Fig. 5. ROC curves of the proposed algorithm for various values of p -number of layers after KLT, averaged over a set of 40 Brodatz-like textures. Results are given for various SNR values: (a) 16 dB, (b) 18 dB, and (c) 20 dB.

framework, we have expected similar results. Fig. 5 shows averaged ROC curves of the proposed algorithm for various KLT configurations and SNR values. We achieved better averaged detection performance using eigenvectors associated with the lowest eigenvalues rather than the highest eigenvalues when considering a minimum required value for the detection rate. Furthermore, the use of additional layers improves the detection performance. This improvement is due to additional information which is concealed in each layer. The best averaged detection performance was achieved without applying the KLT and when all layers were used. Therefore, we have not used the KLT in the rest of our analysis.

Theoretical MSD performance is highly affected by the MSD’s GLR non-centrality parameter, as can be seen in (40). Fig. 6 shows the averaged GLR non-centrality parameter for various SNR values along with the RDWT equivalent of a multi-resolution MSD framework which is employed directly on the RDWT coefficients. We observe that the GLR non-centrality parameter exceeds single resolution analysis SNR values for SNR values larger than approximately 15 dB. Furthermore, the GLR non-centrality parameter exceeds its RDWT equivalent values for SNR values larger than approximately 22 dB. Therefore, we could expect for improved detection results whenever a certain minimum SNR threshold is assured. This is demonstrated in Fig. 7 using theoretical curves for

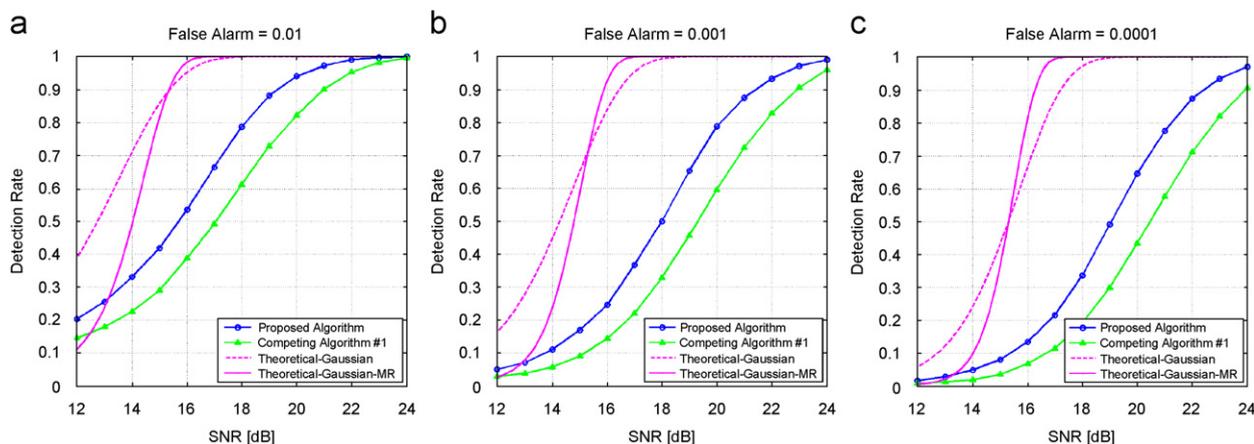


Fig. 7. MSD performance. Proposed algorithm vs. competing algorithm [25] and theoretical results with equal target subspace dimensionality, averaged over a set of 40 Brodatz-like textures. Results are given for various false alarm rates: (a) 10^{-2} , (b) 10^{-3} , and (c) 10^{-4} .

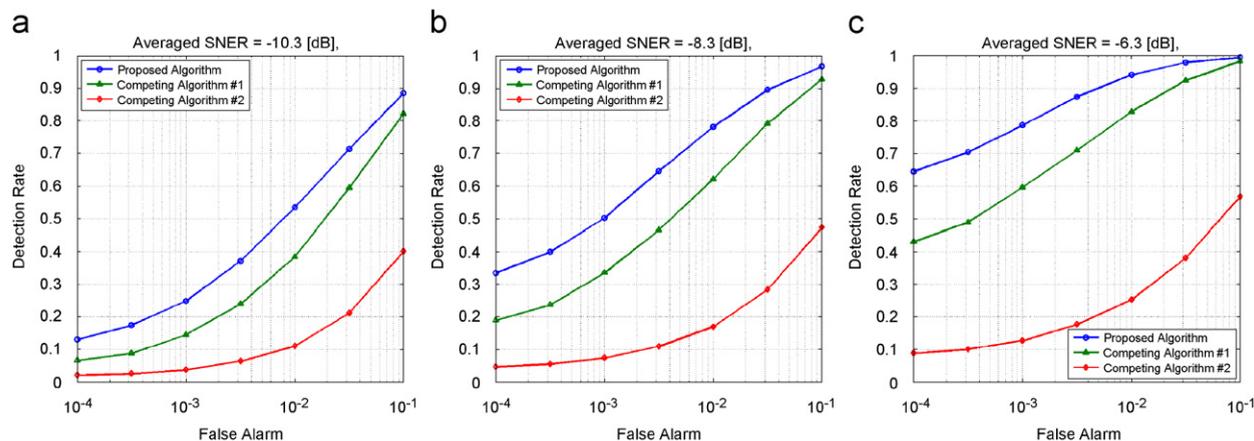


Fig. 8. ROC curves for the proposed algorithm vs. competing algorithms [25,45], averaged over a set of 40 Brodatz-like textures. Results are given for various SNR values: (a) 16 dB, (b) 18 dB, and (c) 20 dB.

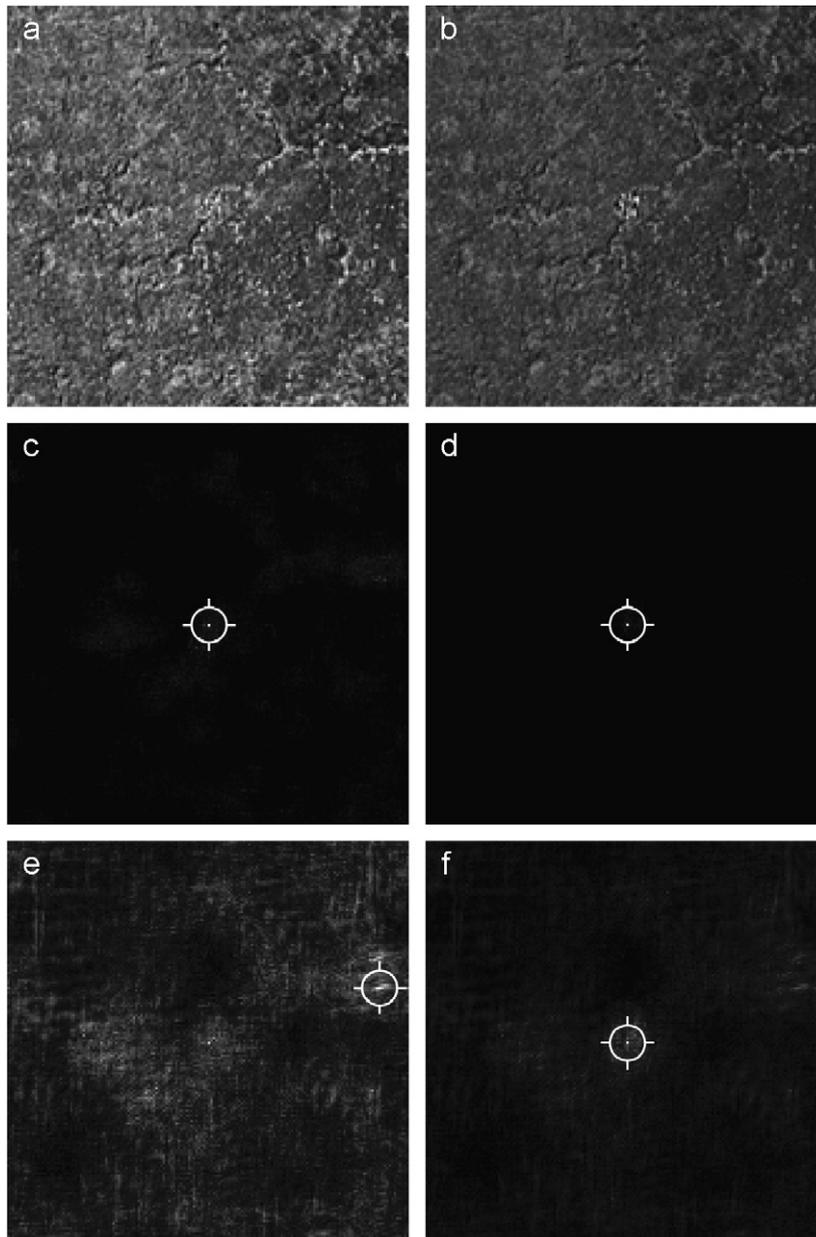


Fig. 9. An example of detection improvement in low SNER environment using a background natural texture of a stone with an additive squared shaped synthetic anomaly in its center. The synthetic anomaly was created using image chips taken from a texture of metal, not having visual resemblance with the stone texture. The anomaly was scaled to achieve: (a) $\text{SNR} = 24 \text{ dB}$, (b) $\text{SNR} = 30 \text{ dB}$. Corresponding anomaly detection results using various methods: (c and d) proposed algorithm, (e and f) Goldman and Cohen [25]. In both methods the detection was performed using a target subspace which was created from the same image chips that were used for synthesizing the anomaly. The white target mark is attached only to the highest value in each detection image.

single resolution and multi-resolution (MR) analyses. The theoretical curves were calculated using (40) and (41) and the averaged GLR non-centrality parameter values, as given in Fig. 6. The minimum SNR threshold is both anomaly and background texture dependent.

We formulate the desired minimum SNR threshold for each layer in Appendix B. A layer which contains a dim target signal contributes to the GLR less than a layer with a clear target signal and might even introduce undesired false alarms to the detector. Although this

can be resolved by applying a proper threshold to the GLR, the use of such layers affects the robustness of the algorithm. We note that a larger GLR non-centrality parameter is achieved when defining the target subspace to contain interaction between target signals. However, as a result, the target subspace dimensionality is increased from $q = p \cdot u_h$ to $q = p \cdot u_h(u_h - 1)/2$, degrading the overall performance. Hence, we have omitted interaction between target signals in the rest of our analysis.

We have tested the performance of the proposed anomaly detection algorithm against two recently published competing algorithms [25,45]. The algorithm in [25] is based on a multi-resolution GMRF and MSD. The algorithm in [45] is based on a multi-resolution feature space and SHT using the Mahalanobis distance, not utilizing any available information on the target signals other than their size. The three algorithms were tested under the same conditions, using the same parameters that are described in Section 5.1. Fig. 7 shows plots of the averaged probability of detection as a function of SNR for fixed values of false alarm rates. Fig. 8 shows ROC curves for fixed values of SNRs. As expected, the detection rate increases with the SNR. The proposed algorithm outperformed the other algorithms, demonstrating an averaged improvement of up to approximately 2 dB when compared to the algorithm presented in [25]. This improvement is crucial in low SNR environments, as we demonstrate in Fig. 9. Still, there is a margin between the theoretical detection rate and the experimental results. This margin results from inconsistency between the Gaussian assumption and the actual distribution of the innovations, which exhibits heavier tails. As such, the background innovations should follow an elliptical multivariate t -distribution and the k th layer GLR (see (34)) then follows a univariate F -distribution [53]. The resulting GLR then follows a mixture of p F -distributions and a new threshold η can be derived to ensure a required false alarm rate. We demonstrate this in Fig. 10 using the stone texture shown in Fig. 2. A possible enhancement for achieving closely Gaussian distributed feature space is the use of a non-stationary local mean [1]. Under this approach, (14) becomes

$$z_k(\mathbf{v}) - \mu_{z_k}(\mathbf{v}) = \sum_{\mathbf{r} \in \mathcal{R}} \theta_k(\mathbf{r}) [z_k(\mathbf{v} + \mathbf{r}) - \mu_{z_k}(\mathbf{v} + \mathbf{r})] + \varepsilon_k(\mathbf{v}), \quad (44)$$

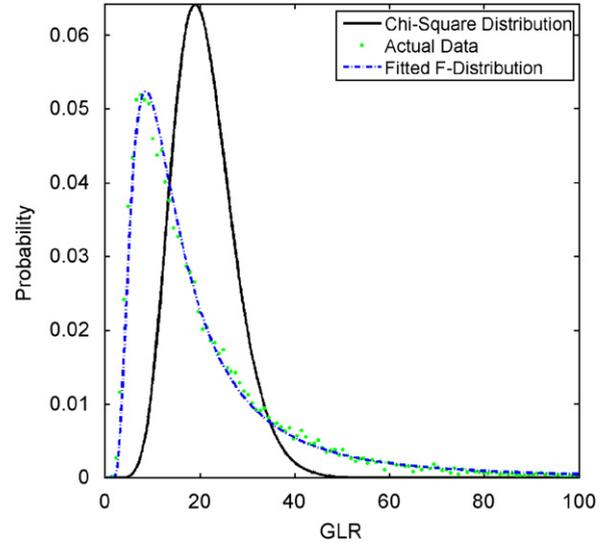


Fig. 10. GLR distribution example using the stone texture shown in Fig. 2.

where

$$\mu_{z_k}(\mathbf{v}) = \frac{\sum_{\mathbf{r} \in \mathcal{R}_z} z_k(\mathbf{v} + \mathbf{r})}{|\mathcal{R}_z|} \quad (45)$$

and the window \mathcal{R}_z is optimized to achieve

$$E[(z_k(\mathbf{v}) - \mu_{z_k}(\mathbf{v}))^3] = 0. \quad (46)$$

However, this approach is useful only when the window \mathcal{R}_z is much larger than the anomalous targets. The margin can also be explained by the use of semi-homogenous textures as a background clutter, which affects the covariance calculation. Nevertheless, the proposed algorithm yields superior detection results, implying that the proposed statistical model better suites the underlying background texture when considering natural textures.

5.4. Anomaly detection examples

We demonstrate the robustness of the proposed algorithm using a diverse set of three Brodatz-like texture images: fabric, raffia and woodgrain. Each background texture contains two additive round shaped synthetic anomalies with zero mean. The synthetic anomalies for each background texture were randomly created using image chips taken from the two other textures. The anomalies were scaled to achieve $\text{SNER} = 0$ dB. Here we have used an image lattice size of $M \times M$, where $M = 192$ and a chip size of $N \times N$, where $N = 24$. An image with

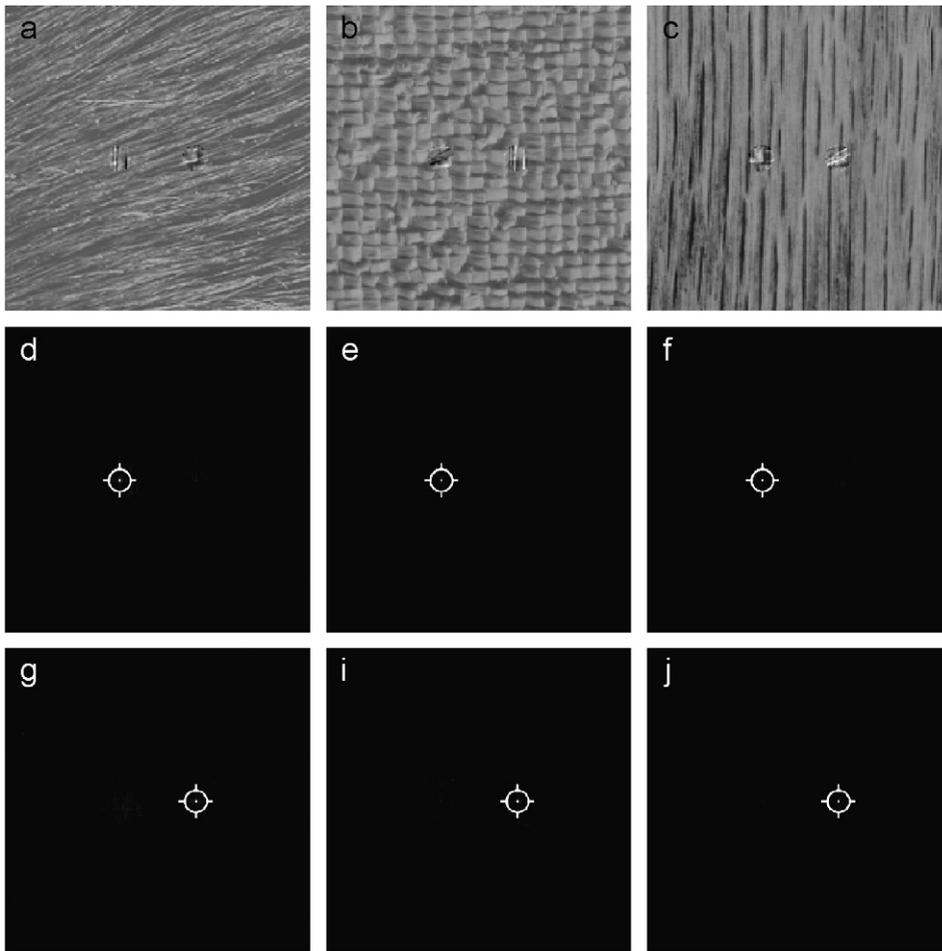


Fig. 11. An example of anomaly detection using a bank of subspace detectors, each best formulated for detecting a specific anomaly type. (a–c) Background natural textures of a fabric, a raffia and a woodgrain correspondingly. Each background texture contains two additive round shaped synthetic anomalies with zero mean. The synthetic anomalies for each background texture were created using image chips taken from the two other textures. The anomalies were scaled to achieve $\text{SNER} = 0 \text{ dB}$. (d–f) Anomaly detection results using the first subspace detector. (g–i) Anomaly detection results using the second subspace detector. The subspace detectors were formulated based on the same image chips that were used for synthesizing the anomalies. The white target mark is attached only to the highest value in each detection image.

various types of anomalies can be processed using a bank of subspace detectors, each designed for detecting a specific anomaly type [3]. Fig. 11 shows the set of test images and the detection results of the subspace detectors. All the anomalous targets were detected. These results demonstrate the robustness of the proposed model and detection algorithm, allowing for the detection of various anomalies in various background textures using the same multi-resolution decomposition and neighbor set of pixels.

6. Conclusion

We have introduced a multi-resolution feature space and a corresponding unsupervised anomaly

detection method. The multi-resolution feature space is based on the RDWT and a proposed multi-resolution RFM which better describes the background clutter in natural images than other models such as the SAR and GMRF. Our detection method is based on a multi-resolution MSD classifier, formulated for detecting subspace targets in multi-resolution RFM innovations. The MSD enables the incorporation of *a priori* information into the detection process. *A priori* information includes the target and interference characterizing subspaces, and multi-resolution layers which are most significant to the detection process. We have investigated the influence of different parameters on the detection performance, and compared the

performance of the proposed method to competing methods. We have demonstrated the robustness of the proposed algorithm using a bank of subspace detectors applied to various background textures. The results show the capability of the proposed method to detect a variety of targets in diverse background clutter patterns using the same multi-resolution decomposition and neighbor set of pixels. The proposed algorithm can be used in conjunction with texture segmentation algorithms for detection of anomalous targets in multi textural patterns.

Acknowledgments

The authors thank the anonymous reviewers for their helpful comments.

Appendix A. Image scaling for reduced prediction error

Let \mathbf{x} denote a column stack representation of an image with zero mean. Let δ denote a scaling adjustment factor. Let \mathbf{y} denote the scaled image $\mathbf{y} = \delta\mathbf{x}$. Let $\boldsymbol{\theta}_x$ denote the RFM weight coefficients of image \mathbf{x} and $\mathbf{B}(\boldsymbol{\theta}_x)$ denote the appropriate RFM matrix. Based on (4) and (7), the RFM innovations are given by

$$\boldsymbol{\varepsilon}_y = \mathbf{B}(\boldsymbol{\theta}_x)\mathbf{y} = \delta\mathbf{B}(\boldsymbol{\theta}_x)\mathbf{x} = \delta\boldsymbol{\varepsilon}_x. \quad (47)$$

We define \mathbf{z}_y and \mathbf{z}_x by

$$\begin{aligned} [\mathbf{z}_x]_i &= [\mathbf{x}]_i^2 - E[[\mathbf{x}]_i^2], \\ [\mathbf{z}_y]_i &= [\mathbf{y}]_i^2 - E[[\mathbf{y}]_i^2] = \delta^2[\mathbf{z}_x]_i, \end{aligned} \quad (48)$$

where $[\cdot]_i$ denotes a vector element in the i th row.

Let $\boldsymbol{\theta}_{z_x}$ denote the RFM weight coefficients of image \mathbf{z}_x and $\mathbf{B}(\boldsymbol{\theta}_{z_x})$ denote the appropriate RFM matrix. From (4), (7) and (48), the RFM innovations are given by

$$\boldsymbol{\varepsilon}_{z_y} = \mathbf{B}(\boldsymbol{\theta}_{z_x})\mathbf{z}_y = \delta^2\mathbf{B}(\boldsymbol{\theta}_{z_x})\mathbf{z}_x = \delta^2\boldsymbol{\varepsilon}_{z_x}. \quad (49)$$

The innovations covariances follow

$$\begin{aligned} \text{cov}[\boldsymbol{\varepsilon}_y] &= \mathbf{B}(\boldsymbol{\theta}_x)\text{cov}[\mathbf{y}]\mathbf{B}(\boldsymbol{\theta}_x)^T \\ &= \delta^2\mathbf{B}(\boldsymbol{\theta}_x)\text{cov}[\mathbf{x}]\mathbf{B}(\boldsymbol{\theta}_x)^T = \delta^2\text{cov}[\boldsymbol{\varepsilon}_x], \\ \text{cov}[\boldsymbol{\varepsilon}_{z_y}] &= \mathbf{B}(\boldsymbol{\theta}_{z_x})\text{cov}[\mathbf{z}_y]\mathbf{B}(\boldsymbol{\theta}_{z_x})^T \\ &= \delta^4\mathbf{B}(\boldsymbol{\theta}_{z_x})\text{cov}[\mathbf{z}_x]\mathbf{B}(\boldsymbol{\theta}_{z_x})^T = \delta^4\text{cov}[\boldsymbol{\varepsilon}_{z_x}]. \end{aligned} \quad (50)$$

In order to achieve smaller prediction error, we require $[\text{cov}[\boldsymbol{\varepsilon}_{z_y}]]_{i,i} \leq [\text{cov}[\boldsymbol{\varepsilon}_y]]_{i,i}$, where $[\cdot]_{i,j}$ denotes the matrix element in the i th row and the j th

column. This results in the following requirement:

$$|\delta| \leq \sqrt{\frac{[\text{cov}[\boldsymbol{\varepsilon}_x]]_{i,i}}{[\text{cov}[\boldsymbol{\varepsilon}_{z_x}]]_{i,i}}}. \quad (51)$$

Using (51), the squaring non-linearity achieves a RFM with lower prediction error variance whenever a proper scaling of the modeled image is performed and the same neighbor set is used.

Appendix B. SNR criterion for improved performance

A simple SNR criterion can be applied on each RDWT layer for determining whether or not the proposed algorithm yields improved detection results when compared to a multi-resolution analysis without the squaring non-linearity. The criterion is given for a specific set of anomalous targets and a background clutter. Decomposition layers which do not conform to this criterion might degrade the performance of the proposed algorithm. This allows for an *a priori* selection of layers that are employed for computing the GLR. Other anomaly detection algorithms may be used for the remaining layers, followed by a fusion process of detection results.

Let $\boldsymbol{\psi}$ denote the target subspace coefficients vector. Let ρ denote an SNR adjustment factor. Let \mathcal{A} denote the target subspace in an RDWT layer. Let Σ denote the background clutter covariance in an RDWT layer. The SNR of a target signal vector $\mathbf{y} = \rho(\mathcal{A}\boldsymbol{\psi})$ in an RDWT layer is then given by

$$\mathbf{y}^T \Sigma^{-1} \mathbf{y} = \rho^2 (\boldsymbol{\psi}^T \mathcal{A}^T \Sigma^{-1} \mathcal{A} \boldsymbol{\psi}). \quad (52)$$

We define subspace \mathcal{B} using the squaring non-linearity such $[\mathcal{B}]_{i,j} = [\mathcal{A}]_{i,j}^2$, where $[\cdot]_{i,j}$ denotes the matrix element in the i th row and the j th column. We define the signal vector \mathbf{z} using the squaring non-linearity such $[\mathbf{z}]_i = [\mathbf{y}]_i^2$, where $[\cdot]_i$ denotes the vector element in the i th row. We define the coefficients vector $\boldsymbol{\phi}$ using the squaring non-linearity such $[\boldsymbol{\phi}]_i = [\boldsymbol{\psi}]_i^2$, where $[\cdot]_i$ denotes the vector element in the i th row. Hence,

$$\mathbf{z} = \rho^2(\mathcal{B}\boldsymbol{\phi}) + \rho^2(\text{residual}). \quad (53)$$

Let \mathcal{D} denote the RFM operator. Let $\boldsymbol{\varepsilon}$ denote the RFM innovations vector. Let \mathcal{A} denote the background clutter RFM innovations covariance in an RDWT layer. We have

$$\boldsymbol{\varepsilon} = \mathcal{D}\mathbf{z} = \rho^2(\mathcal{D}\mathcal{B})\boldsymbol{\phi} + \rho^2\mathcal{D}(\text{residual}). \quad (54)$$

The MSD's layer SNR is then given by

$$(\mathbf{P}_{\mathcal{D}\mathcal{B}}\boldsymbol{\varepsilon})^T \Lambda^{-1}(\mathbf{P}_{\mathcal{D}\mathcal{B}}\boldsymbol{\varepsilon}) \approx \rho^4 (\boldsymbol{\phi}^T (\mathcal{D}\mathcal{B})^T \Lambda^{-1} (\mathcal{D}\mathcal{B}) \boldsymbol{\phi}), \quad (55)$$

where $\mathbf{P}_{\mathcal{D}\mathcal{B}}$ denotes the projection operator onto the subspace spanned by the columns of $[\mathcal{D}\mathcal{B}]$. We define the criterion as

$$\mathbf{y}^T \Sigma^{-1} \mathbf{y} \leq (\mathbf{P}_{\mathcal{D}\mathcal{B}}\boldsymbol{\varepsilon})^T \Lambda^{-1} (\mathbf{P}_{\mathcal{D}\mathcal{B}}\boldsymbol{\varepsilon}). \quad (56)$$

The criterion is satisfied as long as

$$|\rho| \geq \sqrt{\frac{\boldsymbol{\psi}^T \mathcal{A}^T \Sigma^{-1} \mathcal{A} \boldsymbol{\psi}}{\boldsymbol{\phi}^T (\mathcal{D}\mathcal{B})^T \Lambda^{-1} (\mathcal{D}\mathcal{B}) \boldsymbol{\phi}}}. \quad (57)$$

Using (57), the proposed algorithm achieves improved detection results whenever the SNR is above a certain threshold. This threshold depends on the type of anomalous targets, background clutter and multi-resolution layers.

References

- [1] I.S. Reed, X. Yu, Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, *IEEE Trans. Acoust. Speech Signal Process.* 38 (10) (October 1990) 1760–1770.
- [2] E.A. Ashton, Detection of subpixel anomalies in multi-spectral infrared imagery using an adaptive Bayesian classifier, *IEEE Trans. Geosci. Remote Sensing* 36 (2) (March 1998) 506–517.
- [3] D.W.J. Stein, S.G. Beaven, L.E. Hoff, E.M. Winter, A.P. Schaum, A.D. Stocker, Anomaly detection from hyperspectral imagery, *IEEE Signal Process. Mag.* (January 2002) 58–69.
- [4] D. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, *IEEE Signal Process. Mag.* 19 (1) (January 2002) 29–43.
- [5] G.G. Hazel, Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection, *IEEE Trans. Geosci. Remote Sensing* 38 (3) (May 2000) 1199–1211.
- [6] H. Ren, C.I. Chang, Automatic spectral target recognition in hyperspectral imagery, *IEEE Trans. Aerospace Electronic Systems* 39 (4) (October 2003) 1232–1249.
- [7] H. Kwon, S.Z. Der, N.M. Nasrabadi, Adaptive anomaly detection using subspace separation for hyperspectral imagery, *Opt. Eng. Soc. Photo-Opt. Eng.* 42 (11) (November 2003) 3342–3351.
- [8] C.I. Chang, S.S. Chiang, Anomaly detection and classification for hyperspectral imagery, *IEEE Trans. Geosci. Remote Sensing* 40 (6) (June 2002) 1314–1325.
- [9] S.M. Schweizer, J.M.F. Moura, Hyperspectral imagery: clutter adaptation in anomaly detection, *IEEE Trans. Inform. Theory* 46 (5) (August 2000) 1855–1871.
- [10] S.M. Schweizer, J.M.F. Moura, Efficient detection in hyperspectral imagery, *IEEE Trans. Image Process.* 10 (4) (April 2001) 584–597.
- [11] A. Banerji, J. Goutsias, A morphological approach to automatic mine detection problems, *IEEE Trans. Aerospace Electronic Systems* 34 (4) (October 1998) 1085–1096.
- [12] M. Mignotte, C. Collet, Three-class Markovian segmentation of high-resolution sonar images, *Comput. Vision Image Understanding* 76 (3) (December 1999) 191–204.
- [13] S. Reed, Y. Petillot, J. Bell, An automatic approach to the detection and extraction of mine features in sidescan sonar, *IEEE J. Oceanic Eng.* 28 (1) (January 2003) 90–105.
- [14] E. Dura, Z. Yan, L. Xuejun, G.J. Dobeck, L. Carin, Active learning for detection of mine-like objects in side-scan sonar imagery, *IEEE J. Oceanic Eng.* 30 (2) (April 2005) 360–371.
- [15] A. Goldman, I. Cohen, Anomaly detection based on an iterative local statistics approach, *Signal Process.* 84 (7) (July 2004) 1225–1229.
- [16] J. Song, Q.H. Liu, P. Torrione, L. Collins, Two-dimensional and three-dimensional NUFFT migration method for landmine detection using ground-penetrating radar, *IEEE Trans. Geosci. Remote Sensing* 44 (6) (June 2006) 1462–1469.
- [17] P. Torrione, L. Collins, Application of texture feature classification methods to landmine/clutter discrimination in off-lane GPR data, in: *Proceedings of IEEE IGARSS September 2004*, pp. 1621–1624.
- [18] P.A. Torrione, C.S. Throckmorton, L.M. Collins, Performance of an adaptive feature-based processor for a wide-band ground penetrating radar system, *IEEE Trans. Aerospace Electronic Systems* 42 (2) (April 2006) 644–658.
- [19] P. Gader, W.H. Lee, J.N. Wilson, Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening, *IEEE Trans. Geosci. Remote Sensing* 42 (11) (November 2004) 2522–2534.
- [20] F. Berizzi, M. Martorella, G. Bertini, A. Garzelli, F. Nencini, F.D. Acqua, P. Gamba, Sea SAR image analysis by fractal data fusion, in: *Proceedings of IEEE IGARSS, September 2004*, p. 96.
- [21] R. Paget, J. Homer, D. Crisp, (automatic) Target detection in synthetic aperture radar imagery via terrain recognition, in: *Proceedings of the International Conference on Image Processing, October 2001*, pp. 54–57.
- [22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second ed., Academic Press, New York, 1990.
- [23] B. Porat, B. Friedlander, Performance analysis of a class of transient detection algorithms—a unified framework, *IEEE Trans. Signal Process.* 40 (10) (October 1992) 2536–2546.
- [24] L.L. Scharf, B. Friedlander, Matched subspace detectors, *IEEE Trans. Signal Process.* 42 (8) (August 1994) 2146–2156.
- [25] A. Goldman, I. Cohen, Anomaly subspace detection based on a multi-scale Markov random field model, *Signal Process.* 85 (3) (March 2005) 463–479.
- [26] S. Kraut, L.L. Scharf, L.T. McWhorter, Adaptive subspace detectors, *IEEE Trans. Signal Process.* 49 (1) (January 2001) 1–16.
- [27] R. Chellappa, R.L. Kashyap, Texture synthesis using 2-D noncausal autoregressive models, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-33 (1) (February 1985) 194–203.
- [28] K.B. Eom, Long-correlation image models for textures with circular and elliptical correlation structures, *IEEE Trans. Image Process.* 10 (7) (July 2001) 1047–1055.
- [29] R. Mittelman, M. Porat, A new method for multi-resolution texture segmentation using Gaussian Markov random fields, in: *EUSIPCO, 2005*.

- [30] C. Kervrann, F. Heitz, A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics, *IEEE Trans. Image Process.* 4 (6) (June 1995) 856–862.
- [31] A. Sarkar, M.K. Biswas, K.M.S. Sharma, A simple unsupervised MRF model based image segmentation approach, *IEEE Trans. Image Process.* 9 (5) (May 2000) 801–812.
- [32] D.A. Clausi, B. Yue, Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery, *IEEE Trans. Geosci. Remote Sensing* 42 (1) (January 2004) 215–228.
- [33] A. Speis, G. Healey, Feature extraction for texture discrimination via random field models with random spatial interaction, *IEEE Trans. Image Process.* 5 (4) (April 1996) 635–645.
- [34] F.S. Cohen, Z. Fan, M.A. Patel, Classification of rotated and scaled textured images using Gaussian Markov random field models, *IEEE Trans. Pattern Anal. Machine Intell.* 13 (2) (February 1991) 192–202.
- [35] M.G. Bello, A random-field model-based algorithm for anomalous complex image pixel detection, *IEEE Trans. Image Process.* 1 (2) (April 1992) 186–196.
- [36] R.L. Kashyap, R. Chellappa, Estimation and choice of neighbors in spatial-interaction models of images, *IEEE Trans. Inform. Theory* IT-29 (1) (January 1983) 60–72.
- [37] R. Chellappa, R.L. Kashyap, Digital image restoration using spatial interaction models, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-30 (3) (June 1982) 461–472.
- [38] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli, Image de-noising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Process.* 12 (11) (2003) 1338–1351.
- [39] R. Mittelman, On multi-resolution statistical models for image processing, Master's Thesis, Technion - Israel Institute of Technology, August 2005.
- [40] A. Noiboar, I. Cohen, Anomaly detection based on wavelet domain GARCH random field modeling, *IEEE Trans. Geosci. Remote Sensing* 45 (5) (May 2007) 1361–1373.
- [41] Brodatz-like textures (Online). Available: (<http://www-dbv.cs.uni-bonn.de/image/texture.tar.gz>).
- [42] M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans. Image Process.* 4 (11) (November 1995) 1549–1560.
- [43] M. Unser, M. Eden, Nonlinear operators for improving texture segmentation based on features extracted by spatial filtering, *IEEE Trans. Systems Man Cybernet.* 20 (4) (July–August 1990) 804–815.
- [44] R. Mittelman, M. Porat, A new approach to feature extraction for wavelet-based texture classification, in: *ICIP*, September 2005, pp. 1128–1131.
- [45] L. Shadhan, I. Cohen, Detection of anomalies in textures based on multi-resolution features, in: *Proceedings of the 24th IEEE Convention of the Electrical and Electronic Engineers in Israel, Eilat, Israel*, 15–17 November 2006.
- [46] L. Shadhan, Detection of anomalies in texture images using multi-resolution features, Research Thesis, Technion - Israel Institute of Technology, January 2007.
- [47] R.M. Gray, Toeplitz and Circulant Matrices: A Review, Department of Electrical Engineering, Stanford University (Online). Available: (<http://www-ee.stanford.edu/gray/toeplitz.pdf>).
- [48] M. Hassner, J. Sklansky, The use of Markov random fields as models of textures, *Comput. Graphics Image Process.* 12 (April 1980) 357–370.
- [49] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7) (July 1989) 674–693.
- [50] J.K. Patel, C.B. Read, *Handbook of the Normal Distribution*, second ed., CRC Press, Boca Raton, FL, 1996.
- [51] C.R. Rao, *Linear Statistical Inference and its Applications*, second ed., Wiley, New York, 1972.
- [52] C.D. Richmond, Derived PDF of maximum likelihood signal estimator which employs an estimated noise covariance, *IEEE Trans. Signal Process.* 44 (2) (February 1996) 305–315.
- [53] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley, New York, 1982.