

# Anomaly Detection Based on Wavelet Domain GARCH Random Field Modeling

Amir Noiboar and Israel Cohen, *Senior Member, IEEE*

**Abstract**—One-dimensional Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is widely used for modeling financial time series. Extending the GARCH model to multiple dimensions yields a novel clutter model which is capable of taking into account important characteristics of a wavelet-based multiscale feature space, namely heavy-tailed distributions and innovations clustering as well as spatial and scale correlations. We show that the multidimensional GARCH model generalizes the casual Gauss Markov random field (GMRF) model, and we develop a multiscale matched subspace detector (MSD) for detecting anomalies in GARCH clutter. Experimental results demonstrate that by using a multiscale MSD under GARCH clutter modeling, rather than GMRF clutter modeling, a reduced false-alarm rate can be achieved without compromising the detection rate.

**Index Terms**—Anomaly detection, Gaussian Markov random field (GMRF), Generalized Autoregressive Conditional Heteroscedasticity (GARCH), image segmentation, image texture analysis, matched subspace detector, multiscale representation, object recognition.

## I. INTRODUCTION

**I**MAGE ANOMALY detection is the process of distilling a small number of clustered pixels, which differ from the image general characteristics. The type of image, its characteristics, and the type of anomalies depend on the application at hand. Common applications include detection of targets in images [1]–[4], detection of mine features in side-scan sonar [5]–[7], detection of tumorous areas in medical imaging [8], and detection of faults in seismic data [9], [10]. Anomaly-detection algorithms generally consist of some or all of the following steps [1]–[6], [13]–[16]: 1) selection of an appropriate feature space where the distinction between an anomaly and clutter is possible; 2) selection of a statistical model for the image clutter and particular anomalies relevant for the given application; and 3) selection of a detection algorithm.

Feature spaces employed in anomaly and object detection applications are often related to multiresolution representations for various reasons. In an image, features of interest are generally present in different sizes [11]. A multiresolution decomposition of the image is thus an efficient way to analyze such features. Using a multiresolution representation also allows processing of different scales and orientations in parallel, resulting in a more efficient implementation [11]. In [8], it is shown that a certain wavelet transform can be used to approximate the matched filter for detecting Gaussian objects in

Markov noise. It is also claimed that objects in imagery create a response over several scales in a multiresolution representation of an image, and therefore, the wavelet transform can serve as a means for computing a feature set for input to a detector. In [17], a multiscale wavelet representation is utilized to capture periodical patterns of various period lengths, which often appear in natural clutter images. In [12], the orientation and scale selectivity of the wavelet transform are related to the biological mechanisms of the human visual system and are utilized to enhance mammographic features.

Statistical models for clutter and anomalies are usually related to the Gaussian distribution due to its mathematical tractability. In [1], the RX algorithm for detecting anomalies in multiband images was established. The RX algorithm assumes two hypotheses. Under target presence hypothesis, the observations are assumed to be a linear combination of the target signature and clutter, while under the clutter hypothesis, the observations are assumed to consist of only clutter. The statistical distributions of the two hypotheses are Gaussian with common covariance matrices and different means. The Gaussian mixture model [3] is another Gaussian-based feature space model, which has been used for modeling nonhomogeneous data consisting of multiple scenes. In a Gaussian mixture model, each observation is modeled as a linear combination of Gaussian distributions. The Gauss Markov random field (GMRF) is also a well-known Gaussian model, which has been extensively used in the context of texture analysis and anomaly and object detection. The 2-D GMRF has been introduced by Woods [13]. It assumes a stationary image background where every image pixel is represented as a weighted sum of neighboring pixels and additive colored noise. Schweizer and Moura [14], [15] utilize a 3-D GMRF for detecting anomalies in multi- and hyperspectral images, where a moving window over the image is employed. Inside the window, a stationary GMRF is assumed, and the center of the window is tested as an anomaly. Hazel [16] used a multivariate GMRF model of vector observations for the application of image segmentation and anomaly detection. A good review of multiresolution Markov Models for signal and image processing can be found in [18].

Over the years, several anomaly-detection methods have been developed. The single hypothesis test (SHT) [19] assumes a background model and detects areas within the image which differ from the background clutter. The SHT does not allow any use of *a priori* information about the anomalies. The counter approach is the matched signal detector, where an exact pattern of the anomaly is assumed available and utilized in the detection process. Unfortunately, the matched signal detector is not useful for detecting anomalies which significantly

Manuscript received April 9, 2006; revised December 20, 2006.

The authors are with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: noiboar@tx.technion.ac.il; icohen@ee.technion.ac.il).

Digital Object Identifier 10.1109/TGRS.2007.893741

differ from the assumed pattern. Scharf and Friedlander [20] presented the matched subspace detector (MSD). The MSD is appropriate when the anomalies are assumed to lie within a known subspace. The MSD allows for *a priori* information about the anomalies to be used without the single pattern limitation of the matched signal detector. Kraut *et al.* [21] extended the MSD for the case of unknown noise covariance matrix and developed the adaptive subspace detector.

The aforementioned statistical models may account for spatial as well as depth correlation in a multidimensional feature space. However, these models are not appropriate for modeling two common phenomena of often-used feature spaces: heavy tails of the probability density function of the features (known as excess kurtosis) and volatility clustering (a property of many heteroscedastic stochastic processes, which means that large changes tend to follow large changes and small changes tend to follow small changes). Detection algorithms based on these models may result in high false-alarm rates due to mismatch between the model and the data. In particular, it was observed that the wavelet transform, which is often used as a feature space in applications dealing with natural images, yields wavelet coefficients that show excess kurtosis [18], [22]–[24]. It is also argued that spatial- and scale-to-scale statistical dependences of wavelet coefficients exist. That is, coefficients of large magnitudes tend to appear at close spatial locations and at adjacent scales and orientations. These characteristics of a wavelet-based feature space cannot be appropriately modeled by a Gaussian distribution and therefore call for an alternative multidimensional statistical model.

In this paper, we introduce an  $N$ -dimensional ( $N$ -D) Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model and an appropriate detection approach. The 1-D GARCH model [25]–[27] is widely used for modeling financial time series. We have extended this model to 2-D in [28]. Extending the GARCH model to  $N$  dimensions yields a novel clutter model which is capable of taking into account important characteristics of a 3-D feature space, namely heavy-tailed distributions and innovations clustering as well as spatial and depth correlations. We utilize an undecimated wavelet transform and present a 3-D wavelet-based feature space. The undecimated wavelet transform has the property of translation invariance, which is important in the context of anomaly detection.

Since we assume that the multiscale feature space follows a GARCH distribution, we are faced with the challenge of developing an appropriate detection approach. In [20], an MSD is developed for the detection of signals in subspace interference and additive white Gaussian noise (WGN). Here, we introduce a set of multiscale MSDs operating in subspace interference and additive GARCH noise. We show that the GARCH model is more appropriate for the background clutter than the Gaussian model. Since the statistical model is not limited to 2-D, our multiscale MSDs can utilize the correlation within and between layers (for example, every scale and orientation in the wavelet domain can be regarded as a feature space layer). This means that detection at each location in the feature space may be based on feature space data from adjacent layers and is not limited to a single layer. Our multiscale MSD approach takes into

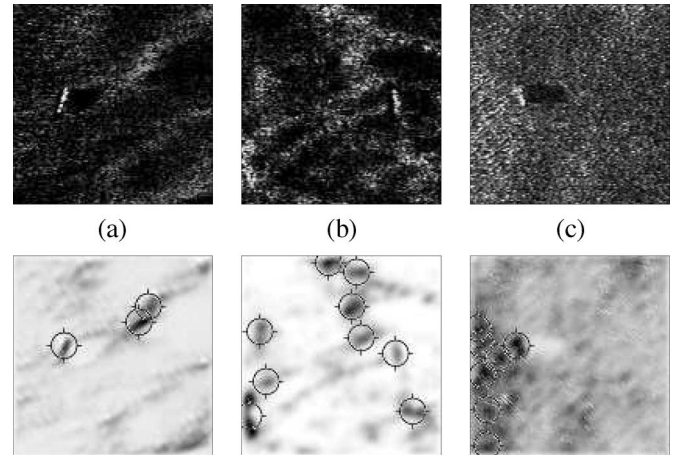


Fig. 1. Detection results on side-scan sea-mine sonar images. (Top row) Original side-scan sea-mine sonar images. (Bottom row) Detection results using a GMRF-based method.

consideration the fact that not all feature space layers contribute uniformly to the detection process. It allows for selection of the most relevant layers, where the relevance criteria are application dependent and independent of the detection algorithm. A separate anomaly subspace is assumed for each multiscale MSD (operating on a set of feature space layers), thus allowing incorporation of *a priori* information into the detection process. These anomaly subspaces need not be of the same size so that greater adaptivity of the anomaly subspace to the characteristics of the feature space, namely, scale and orientation, can be implemented.

This paper is organized as follows: Section II describes the motivation for using the GARCH model. Section III introduces the  $N$ -D GARCH model and a maximum likelihood model estimation. Our multiscale MSD anomaly-detection approach is developed in Section IV. Finally, in Section V, we demonstrate the performance of our method using synthetic data and real sea-mine side-scan sonar images.

## II. MOTIVATION FOR A GARCH MODEL

Anomaly detection is often applied using a Gaussian distribution due to its mathematical tractability. As an example, consider the GMRF-based algorithm presented in [17]. This algorithm is based on 2-D GMRF modeling of uncorrelated layers in a multiscale representation of the image. Correlation between layers is reduced by means of the Karhunen–Loève transform (KLT). Anomaly detection is performed by means of an MSD followed by a threshold operation. We next apply this algorithm to the sea-mine sonar images in the top row of Fig. 1. The side-scan sonar images presented in this paper are from the Sonar-5 database collected by the Naval Surface Warfare Center Coastal System Station (Panama City, FL). The images are 8-bit grayscale. An elongated sea mine (such as those presented in the top row of Fig. 1) is characterized by a bright line (the highlight or echo), corresponding to the scattering response of the mine to the acoustic insonification, and a shadow behind it, corresponding to the blocking of sonar waves by the mine. The image background corresponds

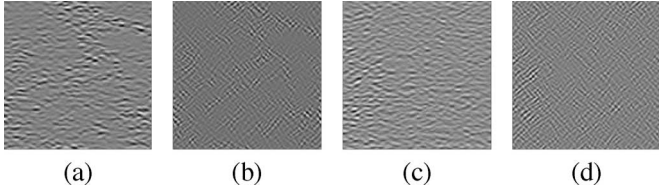


Fig. 2. Example of layers from an undecimated wavelet-transform representation of sea-mine sonar images. (a)–(b) Two layers from the multiresolution representation of Fig. 1(b). (c)–(d) Two layers from the multiresolution representation of Fig. 1(c).

to the reverberation from the seabed [5], [6]. A description of the acquisition process of side-scan sonar imagery and a discussion on the various shapes of minelike objects in such imagery can be found in [6] and [29]. Further technical and navigational information about the specific database used is not available. It is worth noting that the anomaly, being the mine and its shadow, is skewed. We shall not pursue this further in this paper since our goal is to propose a novel clutter model and a corresponding detection approach without the specifics of a certain application. However, for specific applications, when information about the statistical characteristics of the anomaly is available *a priori*, it can be accounted for in order to improve detection results. Applying the anomaly-detection algorithm described above to the sea-mine sonar images in the top row of Fig. 1 results in a high false-alarm rate. This high false-alarm rate is demonstrated at the bottom row of Fig. 1, where the dark target like symbol marks locations where an anomaly has been detected. To explain this high false-alarm rate, let us first look into the statistical characteristics of the feature space of Fig. 1(a) and (b). The kurtosis (fourth moment divided by the square second moment) is a measure of fat tail behavior [22]. The sample kurtosis of the multiresolution feature space of these two figures (obtained by means of an undecimated wavelet transform) is 9.8 and 10.9, respectively. The expected kurtosis value for the Gaussian distribution is three. The high kurtosis values of the feature space for these two images imply a distribution with much heavier tails than the Gaussian distribution. Second, let us examine Fig. 1(c). The sample kurtosis value for the multiresolution feature space of these images is about 4.2, meaning that the distribution is not highly leptokurtic. However, in the areas where false alarms are detected, clustering of innovations occurs. Clustering of innovations is clearly seen in the image itself and is also apparent in the layers of the multiresolution representation, as demonstrated in Fig. 2(c) and (d). This phenomenon is also present in Fig. 1(b) and in its corresponding multiresolution layers presented in Fig. 2(a) and (b). We note that the clustering of innovations phenomena demonstrated in Fig. 2 appears at the same spatial locations in the different multiscale representation layers. This demonstrates scale-to-scale dependence of the wavelet coefficients. The two characteristics of the feature space, namely, heavy tailed distribution and clustering of innovations, cannot be accounted for by the GMRF model underlying the detection algorithm of [17] and therefore call for an alternative statistical model. For that purpose, we introduce in this paper the multidimensional GARCH model.

### III. $N$ -DIMENSIONAL GARCH MODEL

The 1-D GARCH, which is often used as a statistical model for time series, allows for the conditional variance to change as a function of past squared field values and past conditional variance values. It has been shown to be useful in modeling different economic phenomena. In [28], we have extended this model to 2-D. In this paper, we propose a 3-D multiresolution feature space for a given 2-D image. We assume that our feature space follows a 3-D GARCH model, that is, the conditional variance at every location within the feature space depends on the squared field values and the conditional variance values of neighboring locations, where the neighborhood is 3-D. In order to model the 3-D feature space, we next present an extension of the GARCH model to the general case of  $N$ -D and use a 3-D GARCH for modeling our feature space in subsequent sections. Following model definition, we present maximum likelihood model estimation.

#### A. Model Definition

Let  $\mathbf{q} = (q_1, q_2, \dots, q_N)$ ,  $q_i \geq 0$ ,  $i = 1, \dots, N$ ;  $\mathbf{p} = (p_1, p_2, \dots, p_N)$ ,  $p_i \geq 0$ ,  $i = 1, \dots, N$  denote the order of an  $N$ -D GARCH model, and let  $\Gamma_1$  and  $\Gamma_2$  denote two neighborhood sets, such that

$$\begin{aligned} \Gamma_1 &= \{\mathbf{k} | 0 \leq k_i \leq q_i, \quad i = 1, \dots, N \text{ and } \mathbf{k} \neq \mathbf{0}\} \\ \Gamma_2 &= \{\mathbf{k} | 0 \leq k_i \leq p_i, \quad i = 1, \dots, N \text{ and } \mathbf{k} \neq \mathbf{0}\}. \end{aligned}$$

Define an  $N$ -D index vector  $\mathbf{i} = (i_1, i_2, \dots, i_N)$ . Let  $\epsilon_{\mathbf{i}}$  represent a random variable on an  $N$ -D lattice, and let  $h_{\mathbf{i}}$  denote its variance conditioned upon the information set  $\psi_{\mathbf{i}} = \{\{\epsilon_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k} \in \Gamma_1}, \{h_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k} \in \Gamma_2}\}$ . Define the  $N$ -D causal neighborhood of location  $\mathbf{i}$  as:  $\Gamma = \Gamma(\mathbf{i}) = \{\mathbf{k} | k_j \leq i_j, j = 1, \dots, N\}$  and let  $\eta_{\mathbf{i}} \stackrel{iid}{\sim} N(0, 1)$  be another random variable on an  $N$ -D lattice independent of  $\{h_{\mathbf{k}}\}_{\mathbf{k} \in \Gamma}$ . An  $N$ -D GARCH( $\mathbf{p}; \mathbf{q}$ ) process is defined as

$$\epsilon_{\mathbf{i}} = \sqrt{h_{\mathbf{i}}} \eta_{\mathbf{i}} \quad (1)$$

$$h_{\mathbf{i}} = \alpha_0 + \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \epsilon_{\mathbf{i}-\mathbf{k}}^2 + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{k}} \quad (2)$$

and is therefore conditionally distributed as

$$\epsilon_{\mathbf{i}} | \psi_{\mathbf{i}} \sim N(0, h_{\mathbf{i}}). \quad (3)$$

In order to guarantee a nonnegative conditional variance, the model parameters must satisfy

$$\begin{aligned} \alpha_0 &> 0 \\ \alpha_{\mathbf{k}} &\geq 0, \quad \mathbf{k} \in \Gamma_1 \\ \beta_{\mathbf{k}} &\geq 0, \quad \mathbf{k} \in \Gamma_2. \end{aligned} \quad (4)$$

From (2), we see that at every location ( $\mathbf{i}$ ), both the  $N$ -D neighboring squared field values and the  $N$ -D neighboring conditional variances play a role in the current conditional variance. This yields clustering of variations, which is an important characteristic of the GARCH process.

A special case of the GARCH model is when  $\mathbf{q} = \mathbf{p} = \mathbf{0}$ . In this case,  $\epsilon_i$  is simply WGN. Another worth-noting case is when  $N = 1$ , that is,  $\mathbf{q} = q_1$  and  $\mathbf{p} = p_1$  in which case the multidimensional GARCH model resorts to the 1-D GARCH model of Bollerslev [25].

Note that although causality may seem an unnatural model limitation, it is a means of guarantying nonnegativity of the conditional variance in the above model (see Appendix I for more details). The causality of the model may lead to different results, depending on the data orientation. This is demonstrated in Section V-B. Depending on the application and on the data at hand, it may be appropriate to consider more than one data orientation when performing anomaly detection based on the casual GARCH model.

### B. Estimation of an $N$ -D GARCH Model

In this section, we find a maximum likelihood estimate for the GARCH model. We let  $\epsilon_i$  be innovations of a linear regression on an  $N$ -D lattice, where  $y_i$  is the dependent variable,  $\mathbf{x}_i$  is a vector of explanatory variables, and  $\mathbf{b}$  is a vector of unknown parameters, such that  $\epsilon_i$  can be viewed as innovations of an autoregressive process:

$$\epsilon_i = y_i - \mathbf{x}_i^T \mathbf{b}. \quad (5)$$

Note that if  $\epsilon_i$  in (5) is WGN (as described in Section III-A), the regression model is a casual GMRF. This is a special case of the GARCH process.

The conditional distribution of  $y_i$  is Gaussian with mean  $\mathbf{x}_i^T \mathbf{b}$  and variance  $h_i$

$$f(y_i | \mathbf{x}_i^T \mathbf{b}, \psi_i) = \frac{1}{\sqrt{2\pi h_i}} \exp\left(-\frac{(y_i - \mathbf{x}_i^T \mathbf{b})^2}{2h_i}\right). \quad (6)$$

Define the sample space  $\Omega_s$  as an  $N$ -D lattice of size  $K_1 \times K_2 \times \dots \times K_N$  such that:  $\Omega_s = \{\mathbf{i} | 1 \leq i_j \leq K_j, j = 1, \dots, N\}$ , and let  $\boldsymbol{\theta} = [\mathbf{b}^T, \alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}]^T$  be a vector of the unknown parameters, where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are column vectors of the parameter sets  $\{\alpha_{\mathbf{k}}\}_{\mathbf{k} \in \Gamma_1}$  and  $\{\beta_{\mathbf{k}}\}_{\mathbf{k} \in \Gamma_2}$ , respectively. The conditional sample log likelihood is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{\mathbf{i} \in \Omega_s} \log f(y_i | \mathbf{x}_i, \psi_i) \\ &= -\frac{1}{2} \left[ (K_1 + \dots + K_N) \log(2\pi) - \sum_{\mathbf{i} \in \Omega_s} \log(h_i) \right. \\ &\quad \left. - \sum_{\mathbf{i} \in \Omega_s} (y_i - \mathbf{x}_i^T \mathbf{b})^2 / h_i \right]. \end{aligned} \quad (7)$$

Substituting (2) and (5) into (7) together with the constraints in (4) may seem enough to estimate the model parameters. However, due to the structure of the conditional variance (2), wide sense stationarity (WSS) is a necessary condition for guarantying bounded variance for an infinite lattice, and therefore, conditions for WSS should be included in the model estimation process. It is shown in [25] and [28] that a sufficient condition for WSS of the 1-D and 2-D GARCH process, respectively, is

that the sum of all model parameters is smaller than one. A similar result is obtained here for the  $N$ -D case as we prove in the following theorem.

*Theorem 1:* The GARCH( $\mathbf{p}; \mathbf{q}$ ) process, as defined in (1) and (2), is wide-sense stationary with

$$\begin{aligned} E(\epsilon_i) &= 0 \\ \text{var}(\epsilon_i) &= \alpha_0 \left[ 1 - \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} - \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right]^{-1} \\ \text{cov}(\epsilon_i, \epsilon_{\mathbf{k}}) &= 0 \quad \forall \mathbf{i} \neq \mathbf{k} \end{aligned}$$

if and only if

$$\mathbf{1}^T (\boldsymbol{\alpha} + \boldsymbol{\beta}) < 1.$$

*Proof:* See Appendix II. ■

The parameter vector  $\boldsymbol{\theta}$  is found by numerically solving a constrained maximization problem on the log likelihood function with respect to the unknown parameters (see, for example, [30]). The constraints used are those presented in (4) and in Theorem 1. To solve the maximization problem, knowledge of  $\epsilon_i$  and  $h_i$ , where  $i_1, \dots, i_N \leq 0$ , is required. We set these boundaries in a similar way to the studies in [25] and [28] such that

$$\epsilon_i = h_i = \frac{1}{\prod_{\ell=1}^N K_{\ell}} \sum_{\mathbf{k} \in \Omega_s} (y_i - \mathbf{x}_i^T \mathbf{b})^2 \quad \forall i_1, \dots, i_N \leq 0. \quad (8)$$

Since GARCH model estimation requires an iterative procedure to solve the constrained maximization problem presented above, it may be desirable to test if it is appropriate and to estimate the model order before going into the effort of estimating it. Several tests and model-order selection methods have been proposed for the 1-D GARCH model [25]–[27]. The problem of testing for GARCH and model-order estimation is beyond the scope of this paper and may be the subject of future research.

### C. Example

We next demonstrate the heavy tails and volatility clustering properties of the multidimensional GARCH model. Fig. 3 shows seven layers of a 3-D synthetic GARCH data. These synthetic data were generated using the regression and GARCH parameters shown in Tables I and II, respectively. The regression parameters are of low values so that the GARCH behavior can be easily detected in the examples. The sum of the GARCH parameters is  $\mathbf{1}^T (\boldsymbol{\alpha} + \boldsymbol{\beta}) = 0.98$  such that the condition stated in Theorem 1 is satisfied. The parameter values in  $\boldsymbol{\alpha}$  compared to those in  $\boldsymbol{\beta}$  allow the neighboring square field values to have a larger influence on the conditional variance than the neighboring conditional variances. A  $7 \times 7$  random Gaussian-shaped anomaly is planted to the lower left of the image center in all layers of the synthetic image, as shown in, for example, Fig. 3(f). The sample kurtosis of the complete data set is 26.87. The kurtosis of each of the 2-D layers in Fig. 3 is shown in Table III. The kurtosis values are much larger

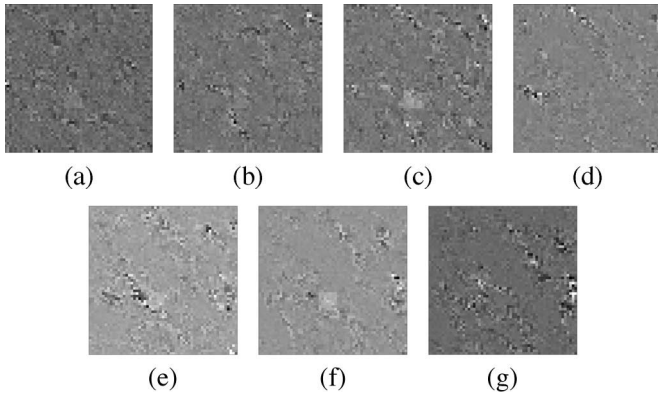


Fig. 3. Seven layers of a GARCH synthetic image with a Gaussian-shaped anomaly.

TABLE I  
REGRESSION PARAMETERS USED FOR GENERATING SYNTHETIC IMAGE

Parameter	$b_0$	$b_{110}$	$b_{100}$	$b_{010}$
value	0	0.05	0.03	0.04

TABLE II  
GARCH PARAMETERS USED FOR GENERATING SYNTHETIC IMAGE

Parameter	$\alpha_0$	$\alpha_{110}$	$\alpha_{100}$	$\alpha_{010}$	$\alpha_{101}$	$\alpha_{111}$	$\alpha_{011}$	$\alpha_{001}$	$\beta_{110}$	$\beta_{100}$	$\beta_{010}$	$\beta_{101}$	$\beta_{111}$	$\beta_{011}$	$\beta_{001}$
value	0.002	0.1	0.2	0.2	0.2	0.02	0.02	0.02	0.1	0.02	0.02	0.02	0.02	0.02	0.02

TABLE III  
KURTOSIS VALUES OF EACH OF THE 2-D LAYERS IN FIG. 3

Layer	Kurtosis
1	7.81
2	9.57
3	17.44
4	15.91
5	11.94
6	14.59
7	16.28

than the value of three characterizing the Gaussian distribution, demonstrating the heavy tails property of the GARCH model. The heavy tails of these sample data can also be viewed from the data's histogram shown in Fig. 4. The volatility clustering property of the GARCH model, which is due to the special structure of the conditional variance, is apparent from Fig. 3, where clustered areas of high variations in gray-scale levels are easily noticed. To further demonstrate it, Fig. 5 shows the seven layers of the conditional variance field based on the estimated model parameters. Darker areas in Fig. 5 represent areas of high conditional variance. These darker areas appear in clusters and not as scattered pixels. The match between darker areas in Fig. 5 and the areas of clustered variations in Fig. 3 is obvious.

#### IV. MULTISCALE MATCHED SUBSPACE ANOMALY DETECTION

In this section, we develop our anomaly-detection approach, which is based on modeling the image feature space (presented in the next section) as a 3-D causal autoregressive model with GARCH innovations and an interference subspace. We assume that the anomalies are sparse within the image, and therefore,

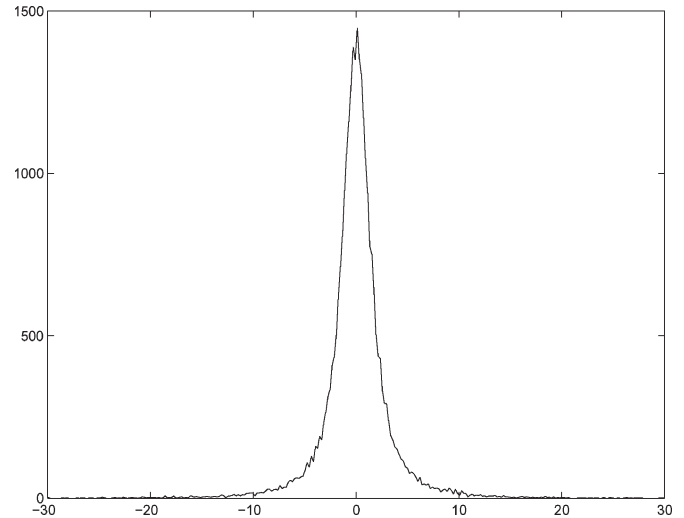


Fig. 4. Histogram of the 3-D GARCH data shown in Fig. 3. The sample kurtosis of these data is 26.87.

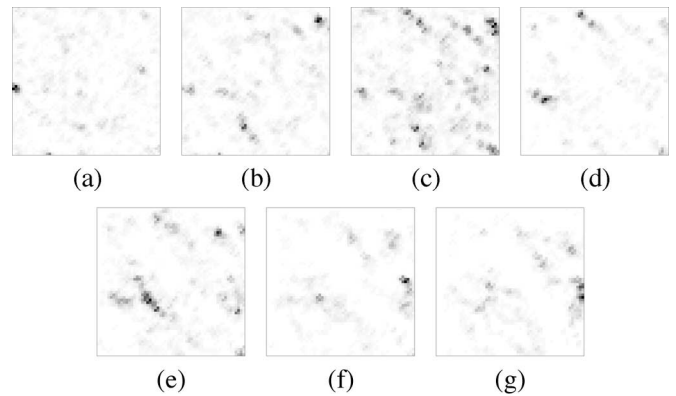


Fig. 5. Seven layers of the conditional variance field of the synthetic GARCH data presented in Fig. 3. Darker areas represent higher conditional variance values.

their influence on the model estimation and on the estimated conditional variance field is negligible. Model estimation is performed as described in Section III-B. The conditional variance field  $h_{i_1, i_2, i_3}$  is calculated based on the estimated model parameters using (2) and (8) and is later used in our detection process. The following sections present our multiscale MSD anomaly-detection approach.

##### A. Statistical Model in the Wavelet Domain

In this section, we present the wavelet-based multiresolution feature space. The suggested multiresolution feature space is an example of a multidimensional feature space that can be statistically modeled using the proposed GARCH model. Let  $Y$  be a 2-D image of size  $K_1 \times K_2$ . We use an undecimated wavelet transform into  $z$  levels to create a multiresolution representation of  $Y$  [8]. The undecimated wavelet transform yields four subband images at every analysis level. These four subband images are labeled  $d_{LH}^i$ ,  $d_{HL}^i$ ,  $d_{HH}^i$ , and  $s_{LL}^i$ , where the subscripts L and H stand for low- and high-pass filtering, respectively,  $d$  labels a detail subband,  $s$  represents the

“smooth” subband, and the superscript  $i$  specifies the analysis level. The undecimated wavelet transform yields a redundant representation. However, the same analysis and synthesis filters are used as in the decimated wavelet transform, and since the transform preserves the spatial dimensions, it is easy to work with. Furthermore, the undecimated wavelet transform has an additional property, namely, translation invariance, which is important in the context of anomaly detection. In [8], the subband images of an undecimated wavelet transform were used to create a 3-D lattice  $\mathbf{Y}$  of size  $K_1 \times K_2 \times (2 * z + 1)$  by creating a feature vector at every spatial location  $(i_1, i_2)$ :

$$\epsilon_{i_1, i_2} = [d_{LH}^1 + d_{HL}^1, d_{HH}^1, d_{LH}^2 + d_{HL}^2, d_{HH}^2, \dots, d_{LH}^z + d_{HL}^z, d_{HH}^z, s_{LL}^z]_{(i_1, i_2)}^T \quad (9)$$

Depending on the application and on the anomalies and clutter characteristics, it may be more appropriate to use each subband image as a separate feature, yielding a 3-D lattice  $\mathbf{Y}$  of size  $K_1 \times K_2 \times (3 * z + 1)$ . If such an approach is selected then the following represents the vector at every spatial location  $(i_1, i_2)$ :

$$\epsilon_{i_1, i_2} = [d_{LH}^1, d_{HL}^1, d_{HH}^1, d_{LH}^2, d_{HL}^2, d_{HH}^2, \dots, d_{LH}^z, d_{HL}^z, d_{HH}^z, s_{LL}^z]_{(i_1, i_2)}^T \quad (10)$$

The transformation from  $Y$  to  $\mathbf{Y}$  generates a multiresolution representation with  $K_3$  layers, where  $K_3 = 2 * z + 1$  or  $K_3 = 3 * z + 1$ , depending on the selected feature vector (9) or (10), respectively.

We assume that there is a set of wavelet filters such that  $\mathbf{Y}$  can be modeled as a 3-D GARCH process (examples are provided in Section V-B). It has been noted by researchers that the distribution of wavelet coefficients of natural images is characterized by heavier tails than the often applied Gaussian distribution (e.g., [18], [22]–[24]). It is also argued that spatial and scale-to-scale statistical dependences of wavelet coefficients exist. That is, coefficients of large magnitudes tend to appear at close spatial locations and at adjacent scales and orientations.

### B. Anomaly and Interference Subspaces

Our anomaly-detection approach introduces a designated best fit multiscale anomaly subspace for each feature space layer. The anomaly subspaces for different layers can be based on different anomaly dimensions. This results in greater adaptivity of the anomaly subspace to the wavelet feature space and improved incorporation of *a priori* information, thus potentially reducing the false-alarm rate of the detection algorithm. The anomaly subspace for layer  $\ell$  is spanned by a training set of  $G_\ell$  anomaly chips. The anomaly chips are denoted  $w_g^\ell$ ,  $g = 1, 2, \dots, G_\ell$  and are each of size  $L_1^\ell \times L_2^\ell \times L_3^\ell$ , where  $L_1^\ell \ll K_1$ ,  $L_2^\ell \ll K_2$ , and  $L_3^\ell \leq K_3$ . To create these anomaly chips, we may start with a training set of images containing anomalies at known image locations. These images are passed through the process of undecimated wavelet transform, and an anomaly chip of size  $L_1^\ell \times L_2^\ell \times L_3^\ell$  is cut around the spatial center of

the anomaly in layer  $\ell$ . Alternatively, we may try to create these anomaly chips synthetically by using prior knowledge. Each anomaly chip is reshaped in a consistent order into a column vector of size  $L_1^\ell L_2^\ell L_3^\ell \times 1$ . The  $G_\ell$  vectors associated with layer  $\ell$  are arranged as columns in a matrix  $H_\ell$ , such that the columns or  $H_\ell$  spans the anomaly subspace for layer  $\ell$ . This procedure is performed for every layer  $\ell = 1, \dots, K_3$ .

When the number of available image chips for a certain layer is high, such that  $\text{rank}(H_\ell) \approx L_1^\ell L_2^\ell L_3^\ell$ , the subspace practically spans the entire space, and anomalies may be falsely detected everywhere within layer  $\ell$ . In this case, dimensionality reduction methods (for example, principal component analysis [31], [32]) are utilized.

An interference subspace is modeled in a similar manner using  $T_\ell$  subspace chips  $s_t^\ell$ ,  $t = 1, 2, \dots, T_\ell$ , each of size  $L_1^\ell \times L_2^\ell \times L_3^\ell$ . A matrix spanning the interference subspace  $S_\ell$  is created accordingly.

### C. Multiscale Matched Subspace Detection

In this section, we introduce an anomaly-detection approach based on an MSD and the multidimensional GARCH statistical model previously presented. In [20], an MSD is developed for the detection of signals in subspace interference and additive WGN. Here, the underlying statistics is more appropriate for the background clutter. We derive a modified MSD operating in subspace interference and additive GARCH noise.

Let  $y_{\ell, s}$  represent a pixel at layer  $\ell$  and spatial location  $s$  in the 3-D lattice  $\mathbf{Y}$ . For each pixel  $y_{\ell, s}$ , we create a column vector  $\mathbf{y}_{\ell, s}$  by row-stacking an image chip of size  $L_1^\ell \times L_2^\ell \times L_3^\ell$  centered around  $(\ell, s)$ . Let  $\epsilon_{\ell, s}$  be a result of row-stacking a chip of a GARCH field of size  $L_1^\ell \times L_2^\ell \times L_3^\ell$  centered around  $(\ell, s)$ . Similarly, let  $\mathbf{u}_{\ell, s}$  be a vector representing the explanatory variable field  $(\mathbf{x}_{i_1, i_2, i_3}^T \mathbf{b})$  in the  $L_1 \times L_2 \times L_3$  neighborhood of  $(\ell, s)$ . Let  $\phi_{\ell, s}$ ,  $\psi_{\ell, s}$  be vectors locating the interference and anomaly within their subspaces  $\langle S_\ell \rangle = \text{span}\{S_\ell\}$ ,  $\langle H_\ell \rangle = \text{span}\{H_\ell\}$ , respectively. We define two hypotheses  $H_0$  and  $H_1$ , which represent the absence and the presence of an anomaly, respectively.

$$\begin{aligned} H_0 : \mathbf{y}_{\ell, s} &= S_\ell \phi_{\ell, s} + \mathbf{u}_{\ell, s} + \epsilon_{\ell, s} \\ H_1 : \mathbf{y}_{\ell, s} &= H_\ell \psi_{\ell, s} + S_\ell \phi_{\ell, s} + \mathbf{u}_{\ell, s} + \epsilon_{\ell, s}. \end{aligned} \quad (11)$$

Let  $\mathbf{h}_{\ell, s}$  represent a row stack of the conditional variance field  $h_{i_1, i_2, i_3}$  around  $(\ell, s)$ , and let  $\Sigma_{\ell, s}$  be a diagonal matrix whose main diagonal equals the elements of  $\mathbf{h}_{\ell, s}$ . Under the two hypotheses, the sample conditional distribution of  $\mathbf{y}_{\ell, s}$  is Gaussian with identical covariance matrices and with different means:

$$\begin{aligned} H_0 : \mathbf{y}_{\ell, s} &\sim N(S_\ell \phi_{\ell, s} + \mathbf{u}_{\ell, s}, \Sigma_{\ell, s}) \\ H_1 : \mathbf{y}_{\ell, s} &\sim N(H_\ell \psi_{\ell, s} + S_\ell \phi_{\ell, s} + \mathbf{u}_{\ell, s}, \Sigma_{\ell, s}). \end{aligned}$$

Note that although  $\Sigma_{\ell, s}$  is a diagonal matrix, the vector elements in  $\mathbf{y}_{\ell, s}$  are only conditionally uncorrelated. Unconditionally, these vector elements may be correlated, such that correlation within and between layers plays a role in our

detection algorithm. This is due to the regression process represented by  $\mathbf{u}_{\ell,s}$  in (11) and to the fact that, under the GARCH assumption, the true distribution of the data is unknown and only the conditional distribution is known. Define  $P_{S_\ell}$  as the projection into the subspace spanned by the columns of  $S_\ell$ , and define  $P_{H_\ell S_\ell}$  as the projection into the subspace spanned by the columns of the concatenated matrix  $[H_\ell S_\ell]$ , that is,

$$P_{S_\ell} = S_\ell (S_\ell^T S_\ell)^{-1} S_\ell^T$$

$$P_{H_\ell S_\ell} = [H_\ell S_\ell] ([H_\ell S_\ell]^T [H_\ell S_\ell])^{-1} [H_\ell S_\ell]^T. \quad (12)$$

From (11) and (12), we find the GARCH innovation field under any one of the hypotheses:

$$H_0 : \boldsymbol{\epsilon}_{\ell,s}^0 = \mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s} - S_\ell \boldsymbol{\phi}_{\ell,s}$$

$$= (I - P_{S_\ell})[\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s}] \quad (13)$$

$$H_1 : \boldsymbol{\epsilon}_{\ell,s}^1 = \mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s} - S_\ell \boldsymbol{\phi}_{\ell,s} - H_\ell \boldsymbol{\psi}_{\ell,s}$$

$$= (I - P_{H_\ell S_\ell})[\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s}]. \quad (14)$$

The conditional likelihood function of  $\boldsymbol{\epsilon}$  under any one of the hypotheses is

$$H_0 : \ell_{\ell,s}^0 = (2\pi)^{-L_1^\ell L_2^\ell L_3^\ell / 2} |\Sigma_{\ell,s}|^{-1/2}$$

$$\times \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{\ell,s}^{0T} \Sigma_{\ell,s}^{-1} \boldsymbol{\epsilon}_{\ell,s}^0 \right]$$

$$H_1 : \ell_{\ell,s}^1 = (2\pi)^{-L_1^\ell L_2^\ell L_3^\ell / 2} |\Sigma_{\ell,s}|^{-1/2}$$

$$\times \exp \left[ -\frac{1}{2} \boldsymbol{\epsilon}_{\ell,s}^{1T} \Sigma_{\ell,s}^{-1} \boldsymbol{\epsilon}_{\ell,s}^1 \right] \quad (15)$$

where  $|\Sigma_{\ell,s}|$  denotes the determinant of  $\Sigma_{\ell,s}$ .

The generalized likelihood ratio (GLR) is defined as

$$L_{\ell,s} = 2 \log \left( \frac{\ell_{\ell,s}^1}{\ell_{\ell,s}^0} \right). \quad (16)$$

Substituting (15) into (16) yields

$$L_{\ell,s} = \boldsymbol{\epsilon}_{\ell,s}^{0T} \Sigma_{\ell,s}^{-1} \boldsymbol{\epsilon}_{\ell,s}^0 - \boldsymbol{\epsilon}_{\ell,s}^{1T} \Sigma_{\ell,s}^{-1} \boldsymbol{\epsilon}_{\ell,s}^1$$

$$= [(P_{H_\ell S_\ell} - P_{S_\ell})(\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s})]^T \Sigma_{\ell,s}^{-1}$$

$$\times [(P_{H_\ell S_\ell} - P_{S_\ell})(\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s})]$$

$$= [\Sigma_{\ell,s}^{-1/2}(\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s})]^T (P_{H_\ell S_\ell} - P_{S_\ell})$$

$$\times [\Sigma_{\ell,s}^{-1/2}(\mathbf{y}_{\ell,s} - \mathbf{u}_{\ell,s})]. \quad (17)$$

The signal-to-noise ratio (SNR) is the ratio between the signal and the noise in terms of intensity. We define the point SNR as the ratio between the energy of the signal which does not lie in the interference subspace  $[(H_\ell \boldsymbol{\psi}_{\ell,s})(I - P_{S_\ell})]^T \times$

$[(H_\ell \boldsymbol{\psi}_{\ell,s})(I - P_{S_\ell})]$  and the innovations' conditional variance  $\Sigma_{\ell,s}$ , such that

$$\text{SNR}_{\ell,s} = [(H_\ell \boldsymbol{\psi}_{\ell,s})(I - P_{S_\ell})]^T \Sigma_{\ell,s}^{-1} [(H_\ell \boldsymbol{\psi}_{\ell,s})(I - P_{S_\ell})]. \quad (18)$$

The GLR is a sum of squared conditionally independent normally distributed variables and therefore is conditionally chi-square distributed with  $\mu_\ell = \text{rank}(H_\ell)$  degrees of freedom, as follows:

$$H_0 : L_{\ell,s} \sim \chi_{\mu_\ell}^2(0)$$

$$H_1 : L_{\ell,s} \sim \chi_{\mu_\ell}^2(\text{SNR}_{\ell,s}). \quad (19)$$

Under hypothesis  $H_1$ , the noncentrality parameters of the chi-square distributions of  $L_{\ell,s}$  is equal to the SNR [20].

The GLR is a 3-D lattice. Our goal is to unify the detection results for multiple layers into a single 2-D detection image corresponding to the original image in size. 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 the layers. Criteria 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. Define the selected subset of layers as  $\Omega \subset \{1, 2, \dots, K_3\}$  such that the final detection image is

$$D_{i_1, i_2} = \sum_{k \in \Omega} L_{i_1, i_2, k} \quad \forall i_1, i_2. \quad (20)$$

The elements summed in (20) are, in general, statistically dependent. This is due to the 3-D neighborhood used to create  $\mathbf{y}_{\ell,s}$ . However, under certain conditions, these elements are conditionally statistically independent. Consider the case where  $\Omega$  consists of  $p$  layers, which are mutually farther apart than the corresponding depth dimension of the 3-D neighborhoods  $L_3^\ell$ ,  $\ell \in \Omega$ . For example, let  $K_3 = 7$  and  $\{L_3^\ell = 3, \forall \ell\}$ . Choosing  $\Omega = \{2, 6\}$  would yield conditionally independent layers in (20). Another example is when  $p = 1$  and  $\Omega = \{\ell\}$ . That is, layer  $\ell$  of the GLR is selected and used as the detection image:

$$D_{i_1, i_2} = L_{i_1, i_2, \ell} \quad \forall i_1, i_2. \quad (21)$$

Depending on the application and available *a priori* information, different selections of  $\ell$  may be appropriate. For example, if  $L_3 = K_3$  such that the anomaly subspace and the feature space have the same depth dimension, it may be appropriate to select  $\ell = \lfloor K_3/2 \rfloor + 1$ , where  $\lfloor \cdot \rfloor$  stands for integer division.

Detection is performed by applying a threshold  $\eta$  to  $D_{i_1, i_2}$ , yielding

$$D_{i_1, i_2} \stackrel{H_1}{\underset{H_0}{\gtrless}} \eta. \quad (22)$$

The threshold is determined by the tradeoff between the desired conditional detection and false-alarm rates. For the



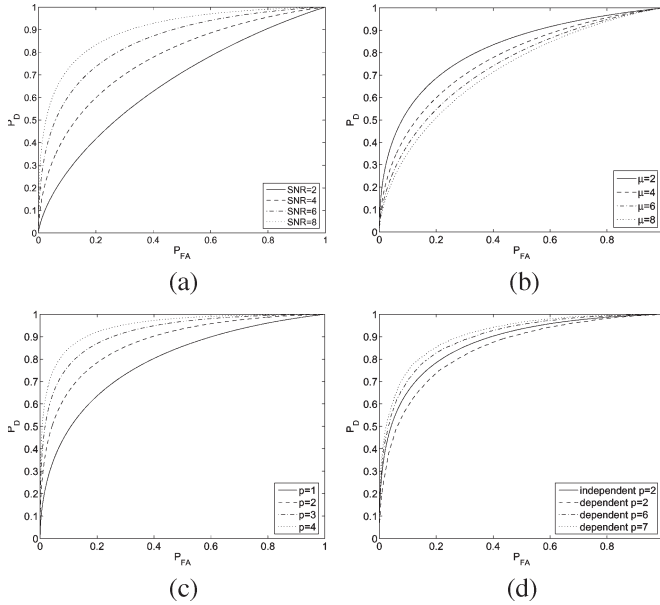


Fig. 6. ROC curves of (a) single-layer detector for different values of SNR, (b) single-layer detector for different anomaly subspace ranks, (c) single-layer detector and an independent layer detector, and (d) dependent layer detectors versus an independent layer detector.

case where independent layers are used, these rates can be calculated by

$$P_{FA} = 1 - P[\chi_{\mu_{tot}}^2(0) \leq \eta] \quad (23)$$

$$P_D = 1 - P[\chi_{\mu_{tot}}^2(\text{SNR}_{tot}) \leq \eta] \quad (24)$$

where  $\mu_{tot} = \sum_{k \in \Omega} \mu_k$  and  $\text{SNR}_{tot} = \sum_{k \in \Omega} \text{SNR}_k$ .

These rates cannot be easily found for the general case due to the conditional statistical dependence of the elements summed in (20). However, computer simulations can present receiver-operating-characteristic (ROC) curves for the general case as we present next, when discussing the performance of the proposed detection approach.

Performance analysis: We shall first look into the performance of a single-layer detection. Fig. 6(a) presents ROC curves for different values of the SNR. These curves were generated using  $\mu_l = 4$ , and the SNR was varied from 2 to 8 in steps of 2. The values of  $P_{FA}$  and  $P_D$  were calculated using (23) and (24), respectively. As expected, the detection rate increases with the SNR. Fig. 6(b) presents ROC curves for different anomaly subspace ranks (different number of degrees of freedom) while the SNR is preserved at a constant value. It is clearly seen that for a constant SNR, the detector's performance increases with decreasing rank of the signal subspace. This is expected since as the anomaly subspace rank increases (under the constant SNR constraint), it is more likely for a false alarm to occur since the anomaly subspace covers a larger portion of the feature space. It is important to note that usually, in real applications, increasing the signal subspace rank results in an increase of the SNR. We next discuss the case of conditionally statistical independent layers. Fig. 6(c) compares the ROC curve of a single-layer detection ( $p = 1$ ) with those of conditionally statistical independent layers ( $p = 2, 3, 4$ ). The

values of  $P_{FA}$  and  $P_D$  were calculated using (23) and (24). We assumed a constant SNR for all layers and an identical anomaly subspace rank for all layers. It is clear that additional independent layers improve the detector's performance. This improvement is due to the additional information concealed in every additional independent layer. ROC curves for the general case, where the sum in (20) contains dependent layers, are presented in Fig. 6(d). These curves are generated by computer simulations under similar assumptions to those used for generating Fig. 6(c) (constant SNR and identical anomaly subspace ranks for all layer). The ROC curve for the former case of two conditionally independent layers is also presented for comparison. To generate Fig. 6(d), we used  $\mu_k = 3; \forall k \in \Omega$  such that the ROC curve representing two independent layers is based on information from six different layers. For the ROC curve representing two dependent layers, we chose to use information from only five different layers ( $\Omega = 2, 4$ ). This explains the apparent advantage of the two independent layers over the two dependent layers. However, under certain conditions, for example,  $K_3 = 8$  and  $L_3^\ell = 3; \forall \ell \in \Omega$ , if we wish to choose independent layers, the maximum value of  $p$  is two (six layers are used in the detection process). Under such conditions, it appears that using a larger number of dependent layers (for example,  $p = 6$  such that  $\Omega = \{2, 3, 4, 5, 6, 7\}$  and 8 layers are used in the detection process or  $p = 7$  such that  $\Omega = \{2, 3, 4, 5, 6, 7, 8\}$  and 9 layers are used in the detection process) may be beneficial as seen in the ROC curves of Fig. 6(d). The advantage of using dependent layers under these conditions is due to the fact that more information is made available in the detection process.

## V. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed anomaly-detection approach on synthetic and real data.

### A. Synthetic Data

We demonstrate the performance of the multiscale MSD on the synthetic data presented in Fig. 3, which exactly matches our model assumptions, and qualitatively investigate the detection performance for different selections of  $\Omega$ . As described in Section III-C, the clutter in Fig. 3 clearly contains areas of clustered variations. These areas may generate high rate of false alarms when conventional GMRF-based anomaly-detection algorithms are deployed [28]. We recall the  $7 \times 7$  random Gaussian-shaped anomaly planted to the lower left of the image center in all layers of the synthetic image as shown in, for example, Fig. 3(f). Note that the anomaly does not stand out in all layers; specifically in Fig. 3(d) and (g), it can hardly be noticed. For anomaly detection, we set the anomaly size to  $L_1 = L_2 = 7$ ,  $L_3 = 3$  and create an anomaly subspace using four image chips. No interference subspace is assumed. We perform parameter estimation as described in Section III-B and anomaly detection as detailed in Section IV-C. Fig. 7 shows layers 2–6 of the GLR. Layers 1 and 7 are not considered here since they suffer from boundary effects due to the 3-D nature of the anomaly subspace. The target mark on each detection image shows the detection result when  $\Omega$  contains this layer



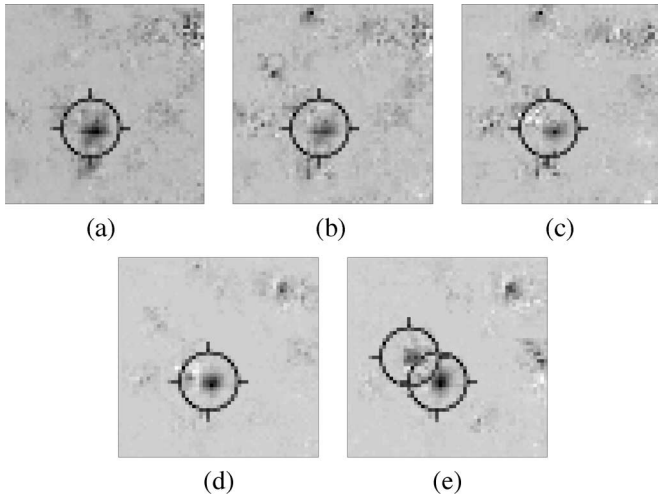


Fig. 7. Layers 2–6 of the GLR with detected anomalies marked by a dark target sign.

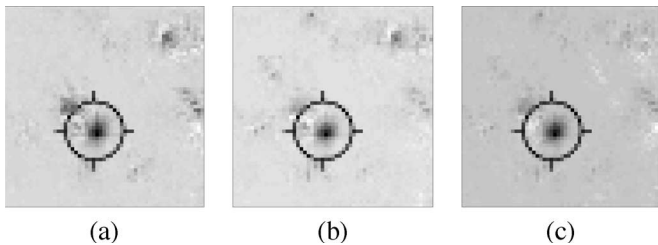


Fig. 8. Detection using a sum of GLR layers. (a) Independent layers 2 and 6. (b) Dependent layers 2–6. (c) Layer 4 of the GLR using a seven-layer anomaly subspace.

only. Note that positive detection is achieved in all layers, while layer 6 includes a false alarm. Fig. 8(a) shows detection results when performing detection using layers 2 and 6, that is,  $\Omega = \{2, 6\}$  and  $L_3 = 3$ . These two detection layers are conditionally statistically independent, and the detection image is clearly detected. Fig. 8(b) presents the results where  $\Omega$  includes dependent layers 2–6 and  $L_3 = 3$ . The 2-D detection result is achieved by means of (20). The detection of Fig. 8(b) seems clearer than that of Fig. 8(a), which qualitatively demonstrates the potential of using dependent layers. Fig. 8(c) presents the detection result for the same synthetic data as above, only here,  $\Omega = \{4\}$  and  $L_3 = K_3 = 7$ , that is, we have used seven layers anomaly subspace, and the detection image is layer 4 of the GLR. Due to the choice of  $\Omega = \{4\}$  and the fact that the anomaly and feature subspaces have the same depth dimension, information from all layers is used in the detection process. The anomaly is clearly detected, and it seems that this detection image is clearer than those presented earlier. For detection in real images, presented next, we use a single layer of the GLR with  $L_3 = K_3$  and  $\Omega = \{\lfloor K_3/2 \rfloor + 1\}$ .

### B. Real Data

The following examples demonstrate the potential of the proposed anomaly-detection approach on real sea-mine sonar images. Automatic detection of sea mines in side-scan sonar

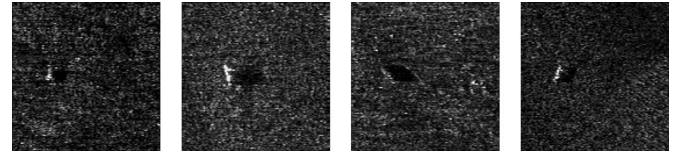


Fig. 9. Original sea-mine sonar images from which an image chip is cut to create the anomaly subspace.

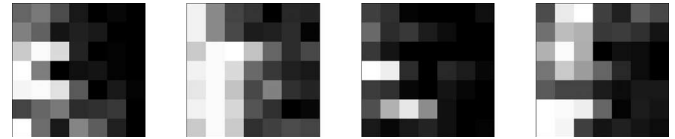


Fig. 10. Image chips cut from the sea-mine sonar images presented in Fig. 9.

imagery is a challenging task due to the high variability of the target and seabottom reverberation (background). An example of this variability is shown in the top row of Fig. 1, which shows three sea-mine sonar images. In [5], a two-phase three-class Markovian segmentation algorithm for the detection of sea mines in side-scan sonar is presented. In the first phase, the data are segmented into two classes: shadow and reverberation, where the latter consists of both echo and seabottom-reverberation regions. In the second phase, the reverberation class is segmented into two classes: seabottom reverberation and echo. In [6], a three-phase procedure for detecting sea mines in side-scan sonar data is presented. In the first stage, suspected mine objects are detected. The shadow cast by the mine is extracted in the second stage. In the third stage, shadow information is used to provide classification information on the shape and dimensions of the detected object. In [17], a competing method based on 2-D GMRF modeling of independent layers in a multiscale representation of the image is presented. Independence of layers is achieved by means of the KLT. Anomaly detection is performed by using an appropriate subspace detector for each layer.

The proposed method has been applied to the images shown in the top row of Fig. 1. A five-layer feature space is created ( $K_3 = 5$ ) for each image by using the biorthogonal spline wavelet transform domain and (9). We note that in our experiments, using different wavelet filters produces similar detection results. The anomaly subspace is created from arbitrarily selected four real examples of sea mines. The images used for creating the subspace are taken from a training set which is mutually exclusive with the images presented in the detection examples. These four images are presented in Fig. 9. The spatial size of the image chip is  $7 \times 7$ . The four chips in the image domain are presented in Fig. 10. They all consist of a portion of a sea-mine highlight and a portion of a sea-mine shadow and thus represent the sea mine properly.  $\Omega = \{3\}$  is used for a single-layer detection since no special information is used for the different layers. We choose  $K_3 = 5$  such that all layers contribute to the detection process. To create the anomaly subspace, a wavelet-based feature space is created for each of the four images in Fig. 9, in a similar manner to that used for the images in Fig. 1. Anomaly chips of size  $7 \times 7 \times 5$  are cut from the four feature spaces. The center of the chip is located in layer 3 at the spatial location corresponding to the center of the image



Fig. 11. Detection results using the proposed method on the sea-mine sonar images of Fig. 1.

chip presented in Fig. 10. These four chips are consistently reordered into column vectors of size  $245 \times 1$  and are set as column vectors in a matrix  $H_{\ell=3}$ , which spans the anomaly subspace. A GARCH(1,1,1;1,1,1) was chosen for modeling the image clutter. For the 1-D GARCH, the GARCH(1;1) is often enough to capture characteristics of financial time series [33]. In a similar manner, we utilize a GARCH(1,1,1;1,1,1) since it allows demonstrating the 3-D model and its advantage over the GMRF model while keeping the calculations simple. Choosing a higher order GARCH model may be more appropriate for the data; however, the results obtained by using this simple model are very promising. We also note that the sea-mine sonar images are noisy. Using a complex high-order model may result in unreliable parameter estimation. Detection results of the proposed approach are presented in Fig. 11. A black target like symbol marks the location of the detected anomaly. Note that the positive detection is achieved in all three images (emphasis is given on the highlight region corresponding to the selected subspace). To further improve the detection results, the proposed method can be aided by inference on the object's shadow made available by published algorithms such as those presented in [5], [6], and [34]. We chose to compare our results with those of the GMRF-based multiscale detection method in [17]. Detection results of the GMRF-based method on the sea-mine sonar images presented in the top row of Fig. 1 are shown in the bottom row of Fig. 1. We have used the same multiscale image representation, subspace image chips, and anomaly spatial size for both the proposed approach and the GMRF-based method. It is clearly demonstrated by these figures that the GMRF-based method may result in high false-alarm rate, while the proposed method potentially reduced the false-alarm rate. The high false-alarm rate of the GMRF-based method may be due to the inability of the GMRF statistics to properly model the leptokurtic feature space of Fig. 1(a) and (b). High kurtosis values correspond to non-Gaussian distributions; therefore, the underlying GMRF model of the GMRF-based method does not allow for accurate detection. Another reason for the high false alarm of the GMRF-based method is the clustering of innovation phenomena apparent in the feature spaces of all three images. The GMRF cannot properly model clustering of innovations. Information on statistical values and examples of clustering of innovations in the feature spaces of these images are presented in Section II. The examples presented here demonstrate the potential of the proposed statistical model and detection method in a variable background.

To further demonstrate the robustness of the proposed method, Fig. 12(a) presents a sea-mine sonar image, in which

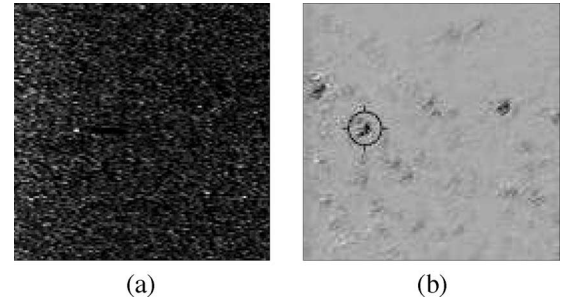


Fig. 12. Original side-scan sonar image of a spherical object and a corresponding detection image. (a) Original sea-mine sonar image. (b) Detection results using the proposed method on the sea-mine sonar images.

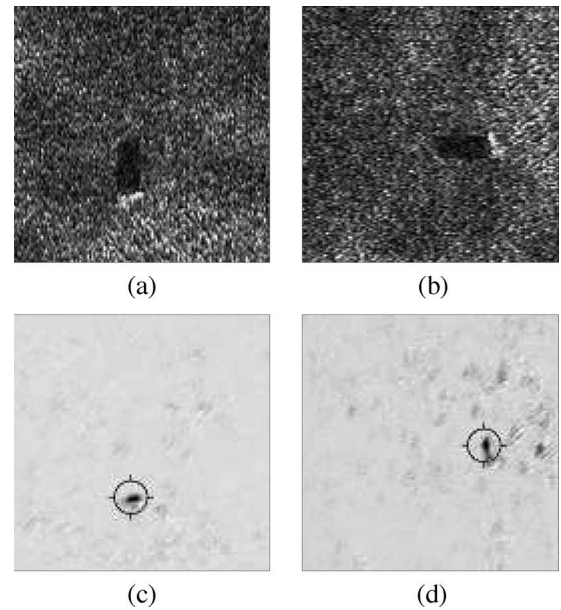


Fig. 13. Detection results on rotated versions of the side-scan sea-mine sonar image presented in Fig. 1(c). (a) Original image rotated by  $90^\circ$ . (b) Original image rotated by  $180^\circ$ . (c) Detection results for (a) using the proposed method. (d) Detection results for (b) using the proposed method.

the mine object (probably a spherical object) differs from the mine objects used to create the anomaly subspace (elongated mines). Fig. 12(b) shows the detection results using the exact same anomaly subspace used in the detection process leading to Fig. 11. This demonstrates the potential of detecting minelike objects in sonar imagery using a subspace, which does not contain exact examples of such objects.

As discussed in Section III-A, causality seems an unnatural model limitation. We therefore demonstrate detection results for different image orientations. The side-scan sonar image of Fig. 1(c) is rotated by  $90^\circ$  and  $180^\circ$ , and the resulting images are presented in the top row of Fig. 13. Detection is performed using the exact same procedures as above, only that the subspace images are rotated according to the image orientation in order to isolate the effect of model causality on the detection results. Detection results are shown in the bottom row of Fig. 13. Although positive detection without false alarms is achieved in all orientations ( $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ ), the detection images differ. In particular, Fig. 13(c) produces the best detection

results, while Fig. 13(d) produces the worst detection of the three images. Due to the model causality, it may be appropriate in some applications to consider all four possible orientations of a given image.

## VI. CONCLUSION

We have introduced a multidimensional GARCH model and a corresponding anomaly subspace detection method. The GARCH statistical model is characterized by heavy tails and clustering of innovations. These characteristics are of interest since they are common in image multiresolution representations and cannot be well modeled by Gaussian-based statistical models such as the GMRF. The multiresolution representation and clutter modeling based on multiscale GARCH model allow for correlation in and between layers of the multiresolution representation in addition to the GARCH characteristics. Our detection method is based on the MSD in GARCH background noise. The MSD enables incorporation of *a priori* information into the detection process. A separate anomaly subspace is assumed for each layer in the multiresolution representation. Since not all layers contribute uniformly to the detection process, we allow for a selection of only those layer which are most significant to the detection. Layers are selected *a priori* based on intermediate results obtained for each layer. We have demonstrated the performance of the proposed statistical model and detection approach on synthetic images and real sea-mine side-scan sonar imagery. Automatic detection of sea mines in side-scan sonar imagery is a challenging task due to the high variability of the target and seabottom reverberation. Compared with a GMRF-based method, we presented improved performance, i.e., a reduced false-alarm rate while retaining a high detection rate.

## APPENDIX I

For ease of notation, let us explore the GARCH model in 1-D and show that there is no way to guaranty a nonnegative conditional variance without the causality constraint.

Let  $q, p \geq 0$  denote the order of a noncausal symmetric GARCH model, and let  $\Gamma_1$  and  $\Gamma_2$  denote two neighborhood sets which are defined by

$$\begin{aligned}\Gamma_1 &= \{k \mid -q \leq k \leq q, \quad k \neq 0\} \\ \Gamma_2 &= \{k \mid -p \leq k \leq p, \quad k \neq 0\}.\end{aligned}$$

Let  $\epsilon_t$  represent a stochastic process, and let  $h_t$  denote its variance conditioned upon the information set  $\psi_t = \{\{\epsilon_{t-k}\}_{k \in \Gamma_1}, \{h_{t-k}\}_{k \in \Gamma_2}\}$ . Let  $\eta_t \stackrel{iid}{\sim} N(0, 1)$  be a stochastic process independent of  $h_k, \forall k \neq t$ . The noncausal GARCH( $p, q$ ) process is defined as

$$\epsilon_t = \sqrt{h_t} \eta_t \quad (25)$$

$$h_t = \alpha_0 + \sum_{k \in \Gamma_1} \alpha_k \epsilon_{t-k}^2 + \sum_{k \in \Gamma_2} \beta_k h_{t-k} \quad (26)$$

and is therefore conditionally distributed as

$$\epsilon_t | \psi_t \sim N(0, h_t). \quad (27)$$

In order to guarantee a nonnegative conditional variance, we require that

$$h(t) \geq 0 \quad \forall t. \quad (28)$$

We need to find conditions on the parameters space  $\{\alpha_0, \{\alpha_k\}_{k \in \Gamma_1}, \{\beta_k\}_{k \in \Gamma_2}\}$  such that (28) holds. Substituting (25) into (26) yields

$$h_t = \alpha_0 + \sum_{k \in \Gamma_1} \alpha_k \eta_{t-k}^2 h_{t-k} + \sum_{k \in \Gamma_2} \beta_k h_{t-k}. \quad (29)$$

This is a set of linear equations in  $h_t$ . To the best of our knowledge, there are no known conditions on the parameters of a set of linear equations in order to guarantee a nonnegative solution. In addition, the equation parameters include a stochastic process  $\eta_{t-k}^2$ , which cannot be limited in any way. This means that causality is a necessary constraint to guarantee a nonnegative conditional variance.

## APPENDIX II

In this appendix, we prove Theorem 1 presented in Section III.

Repeating substitutions of (3) into (2) yields

$$\begin{aligned}h_{\mathbf{i}} &= \alpha_0 + \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{k}}^2 h_{\mathbf{i}-\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{k}} \\ &= \alpha_0 + \sum_{\mathbf{r} \in \Gamma_1} \alpha_{\mathbf{r}} \eta_{\mathbf{i}-\mathbf{r}}^2 \\ &\quad \times \left[ \alpha_0 + \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{r}-\mathbf{k}}^2 h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} \right] \\ &\quad + \sum_{\mathbf{r} \in \Gamma_2} \beta_{\mathbf{r}} \left[ \alpha_0 + \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{r}-\mathbf{k}}^2 h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} \right] \\ &= \alpha_0 \sum_{g=0}^{\infty} M(\mathbf{i}, g)\end{aligned} \quad (30)$$

where  $M(\mathbf{i}, g)$  involves all terms of the form

$$\prod_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}}^{a_{\mathbf{k}}} \prod_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}}^{b_{\mathbf{k}}} \prod_{r=1}^n \eta_{\mathbf{i}-\mathbf{s}_r}^2$$

for

$$\begin{aligned}\sum_{\mathbf{k} \in \Gamma_1} a_{\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} b_{\mathbf{k}} &= g \\ \sum_{\mathbf{k} \in \Gamma_1} a_{\mathbf{k}} &= n\end{aligned}$$

and

$$\begin{aligned} \mathbf{0} &< |\mathbf{s}_1| \leq |\mathbf{s}_2| \leq \dots \leq |\mathbf{s}_n| \\ \mathbf{s}_r &\equiv (s_{r_1}, \dots, s_{r_N}) \\ s_{r_1} &\leq \max \{gq_1, (g-1)q_1 + p_1\} \\ &\vdots \\ s_{r_N} &\leq \max \{gq_N, (g-1)q_N + p_N\}. \end{aligned}$$

Thus

$$\begin{aligned} M(\mathbf{i}, 0) &= 1 \\ M(\mathbf{i}, 1) &= \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{k}}^2 + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \\ M(\mathbf{i}, 2) &= \sum_{\mathbf{r} \in \Gamma_1} \alpha_{\mathbf{r}} \eta_{\mathbf{i}-\mathbf{r}}^2 \left[ \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{r}-\mathbf{k}}^2 + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right] \\ &\quad + \sum_{\mathbf{r} \in \Gamma_2} \beta_{\mathbf{r}} \left[ \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{r}-\mathbf{k}}^2 + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right] \end{aligned}$$

and in general

$$M(\mathbf{i}, g+1) = \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} \eta_{\mathbf{i}-\mathbf{k}}^2 M(\mathbf{i}-\mathbf{k}, g) + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} M(\mathbf{i}-\mathbf{k}, g). \quad (31)$$

Since  $\eta_{\mathbf{i}}$  is independent identically distributed (i.i.d.), the moments of  $M(\mathbf{i}, g)$  are not dependent on  $(\mathbf{i})$ , and in particular

$$E \{M(\mathbf{i}, g)\} = E \{M(\mathbf{k}, g)\} \quad \forall \mathbf{i}, \mathbf{k}, g. \quad (32)$$

From (31) and (32), we obtain

$$\begin{aligned} E \{M(\mathbf{i}, g+1)\} &= \left[ \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right] E \{M(\mathbf{i}, g)\} \\ &= \left[ \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right]^{g+1} E \{M(\mathbf{i}, 0)\} \\ &= \left[ \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right]^{g+1}. \end{aligned} \quad (33)$$

Finally, by (1), (30), and (33)

$$\begin{aligned} E \{\epsilon_{\mathbf{i}}^2\} &= \alpha_0 E \left\{ \sum_{g=0}^{\infty} M(\mathbf{i}, g) \right\} \\ &= \alpha_0 \sum_{g=0}^{\infty} E \{M(\mathbf{i}, g)\} \\ &= \alpha_0 \left[ 1 - \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} - \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right]^{-1} \end{aligned} \quad (34)$$

if and only if

$$\sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} < 1$$

and

$$\begin{aligned} E(\epsilon_{\mathbf{i}}) &= 0 \\ \text{var}(\epsilon_{\mathbf{i}}) &= \alpha_0 \left[ 1 - \sum_{\mathbf{k} \in \Gamma_1} \alpha_{\mathbf{k}} - \sum_{\mathbf{k} \in \Gamma_2} \beta_{\mathbf{k}} \right]^{-1} \\ \text{cov}(\epsilon_{\mathbf{i}}, \epsilon_{\mathbf{k}}) &= 0 \quad \text{for } (\mathbf{i}) \neq (\mathbf{k}) \end{aligned}$$

follows immediately.

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and helpful suggestions.

#### REFERENCES

- [1] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [2] E. A. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 506–517, Mar. 1998.
- [3] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.
- [4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [5] M. Mignotte, C. Collet, P. Perez, and P. Boutheymy, "Three-class Markovian segmentation of high-resolution sonar images," *Comput. Vis. Image Process.*, vol. 76, no. 3, pp. 191–204, Nov. 1999.
- [6] 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.
- [7] 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, Apr. 2005.
- [8] R. N. Strickland and H. I. Hahn, "Wavelet transform methods for object detection and recovery," *IEEE Trans. Image Process.*, vol. 6, no. 5, pp. 724–735, May 1997.
- [9] I. Cohen and R. R. Coifman, "Local discontinuity measures for 3-D seismic data," *Geophysics*, vol. 67, no. 6, pp. 1933–1945, Nov./Dec. 2002.
- [10] A. Noiboar and I. Cohen, "Anomaly detection in three dimensional data based on Gauss Markov random field modeling," in *Proc. 23rd IEEE Conv. Elect. and Electron. Eng.*, Tel-Aviv, Israel, Sep. 2004, pp. 448–451.
- [11] X. Yu, I. S. Reed, W. Karske, and A. D. Stocker, "A robust adaptive multi-spectral object detection by using wavelet transform," in *Proc. IEEE ICASSP*, 1992, pp. V-141–V-144.
- [12] A. F. Laine, S. Schuler, J. Fan, and W. Huda, "Mammographic feature enhancement by multiscale analysis," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 725–740, Dec. 1994.
- [13] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 2, pp. 232–240, Mar. 1972.
- [14] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1855–1871, Aug. 2000.
- [15] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [16] G. G. Hazel, "Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1199–1211, May 2000.
- [17] A. Goldman and I. Cohen, "Anomaly subspace detection based on a multi-scale Markov random field model," *Signal Process.*, vol. 85, no. 3, pp. 463–479, Mar. 2005.
- [18] A. S. Willisky, "Multiresolution Markov models for signal and image processing," *Proc. IEEE*, vol. 90, no. 8, pp. 1396–1458, Aug. 2002.
- [19] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
- [20] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

- [21] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.
- [22] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characteristics in the wavelet domain," *IEEE Trans. Image Process.*, vol. 8, no. 12, pp. 1688–1701, Dec. 1999.
- [23] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
- [24] A. Srivastava, "Stochastic models for capturing image variability," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 63–76, Sep. 2002.
- [25] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *J. Econom.*, vol. 31, no. 3, pp. 307–327, Apr. 1986.
- [26] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, Jul. 1982.
- [27] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.
- [28] A. Noiboar and I. Cohen, "Two-dimensional GARCH model with application to anomaly detection," presented at the 13th European Signal Processing Conf., Istanbul, Turkey, Sep. 2005, Paper 1594.
- [29] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *Proc. Inst. Electr. Eng.—Radar, Sonar Navigation*, vol. 151, no. 1, pp. 48–56, Feb. 2004.
- [30] E. K. Berndt, B. H. Hall, R. E. Hall, and J. A. Hausman, "Estimation and inference in nonlinear structural models," *Ann. Econ. Soc. Meas.*, vol. 3/4, pp. 653–665, 1974.
- [31] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K. R. Castleman, "Subspace-based prototyping and classification of chromosome images," *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1277–1287, Sep. 2005.
- [32] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [33] P. R. Hansen and A. Lunde, *A Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?* Aarhus, Denmark: Center Analytical Finance, Univ. Aarhus, Mar. 2001. working paper.
- [34] E. Dura, J. M. Bell, and D. M. Lane, "Superellipse fitting for the classification of mine-like shapes in side-scan sonar images," in *Proc. Oceans MTS/IEEE*, 2002, vol. 1, pp. 23–28.



Electronic Engineers in Israel, in September 2004.

**Amir Noiboar** received the B.Sc. degree from Tel Aviv University, Tel Aviv, Israel, and the M.Sc. degree from Technion–Israel Institute of Technology, Haifa, Israel, both in electrical engineering, in 1996 and 2007, respectively.

Since 1996, he has been an R&D Officer with the Israeli Defense Forces working in the field of information systems. His academic interest is in the field of statistical signal processing.

Mr. Noiboar received the best student paper award at the 23rd IEEE Convention of the Electrical and



**Israel Cohen** (M'01–SM'03) received the B.Sc. (*Summa Cum Laude*), M.Sc., and Ph.D. degrees from Technion–Israel Institute of Technology, Haifa, Israel, in 1990, 1993, and 1998, respectively, all in electrical engineering.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001, he joined the Department of Electrical

Engineering, Technion–Israel Institute of Technology, Haifa, where he is currently an Associate Professor. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification, and adaptive filtering.

Dr. Cohen received in 2005 and 2006 the Technion Excellent Lecturer awards. He serves as Associate Editor of the *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING* and the *IEEE SIGNAL PROCESSING LETTERS*, and as Guest Editor of a Special Issue on the *EURASIP Journal on Applied Signal Processing* on Advances in Multimicrophone Speech Processing and a Special Issue on the *EURASIP Speech Communication Journal* on Speech Enhancement. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing*.