ANOMALY DETECTION BASED ON WAVELET DOMAIN GARCH RANDOM FIELD MODELING

AMIR NOIBOAR

ANOMALY DETECTION BASED ON WAVELET DOMAIN GARCH RANDOM FIELD MODELING

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

IN ELECTRICAL ENGINEERING

AMIR NOIBOAR

SUBMITTED TO THE SENATE OF THE TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY

SHEBAT, 5767 HAIFA JANUARY, 2007

THIS RESEARCH THESIS WAS SUPERVISED BY ASSOCIATE PROFESSOR ISRAEL COHEN UNDER THE AUSPICES OF THE ELECTRICAL ENGINEERING DEPARTMENT

ACKNOWLEDGMENT

I am grateful to Prof. Israel Cohen for his guidance throughout all stages of this research. I would also like to thank the examiners of this thesis, Prof. David Malah and Prof. Amir Averbuch, for their constructive comments and helpful suggestions.
Special thanks to my wife Anat for her love and support.

THE GENEROUS FINANCIAL HELP OF THE TECHNION IS GRATEFULLY ACKNOWLEDGED

dedicated to my wife Anat

Contents

List of Figures

List of Tables

Abstract

Image anomaly detection is the process of distilling a small number of clustered pixels, which differ from the image's 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, detection of defects in silicon wafers, detection of mine features in side-scan sonar and detection of tumorous areas in medical imaging. Anomaly detection algorithms generally consist of three stages: selection of an appropriate feature space in which the distinction between the anomaly and the general clutter is possible; selection of a statistical model for the feature space representing the image clutter and selection of a detection algorithm. This last stage implies a selection of an anomaly model, which defines the type of anomaly or anomalies relevant for the application. This work focuses on the latter two stages.

The Gaussian distribution is a common basis for feature space statistical models due to its mathematical tractability. A major drawback of using the Gaussian distribution lays in its inability to appropriately model 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 Gaussian models may result in high false alarm rates when applied to such feature spaces, due to the inadequacy 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. It is also observed that spatial and scale-to-scale statistical dependencies 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 commonly used feature spaces cannot be appropriately modeled by a Gaussian distribution and therefore call for an alternative multi-dimensional statistical model.

We thus introduce an N dimensional generalized autoregressive conditional heteroscedasticity (GARCH) model. The 1D GARCH model is widely used for modeling financial time series. Extending the GARCH model into N dimensions yields a novel clutter model which is capable of taking into account important characteristics of commonly used feature spaces, namely heavy-tailed distributions and innovations clustering as well as spatial and depth correlations. In this work we utilize an undecimated wavelet transform and present a 3D wavelet-based feature space. The undecimated wavelet transform has the property of translation invariance, which is important in the context of anomaly detection. A 3D GARCH model is used as the underlying statistics of this feature space.

Once statistical modeling is accomplished, we are faced with the challenge of developing an appropriate detection approach. In practice it often happens that the class of anomalies to be detected is not well defined. Algorithms, which assume a specific anomaly pattern, such as the matched signal detector are therefore inapplicable. Furthermore, algorithms such as the single hypothesis test, which assume that *a priory*

information about the anomalies is not available, are also not sufficient, since some examples of typical anomalies are often available. In this work we develop two detectors, each is comprised of a set of multiscale Matched Subspace Detectors (MSDs). The MSD was originally developed for the detection of signal in subspace interference and additive white Gaussian noise. Our MSDs operate in additive colored Gaussian noise (corresponding to a 2-D Gauss Markov Random Field (GMRF) model) and in additive GARCH noise. For every feature space layer (for example, every scale and orientation in the wavelet domain may be considered a feature space layer) a separate anomaly subspace is used, thus allowing for better incorporation of *a priory* information into the process. When a large training set of anomalies is available *a priory*, we utilize a procedure known as eigen-pictures to create an anomaly subspace of a desired rank. Our approach takes into consideration the fact that not all feature space layers contribute uniformly to the detection process and allows for selection of the most relevant layers, where the relevance criterion is application dependent.

We demonstrate the performance of the set of MSDs operating in GMRF noise by applying it to the detection of defects in wafer images and to detection of faults in 3D seismic data. These images are 3-D in nature and the image data itself is used as a feature space. Although the potential of the proposed method is demonstrated on these examples, applying GMRF based detection methods when the underlying statistics is characterized by excess kurtosis and innovation's clustering produces poor results. This is due to the fact that the image clutter cannot be well modeled using a Gaussian distribution. For such environments we propose using a GARCH based detection approach. We demonstrate it using synthetic and real sea-mine side-scan sonar images. For the sonar images, a wavelet based feature space is created, which

shows high kurtosis values and clustering of innovations. Our results show the potential of the set of MSDs, the importance of an appropriate statistical model for the background and the advantages of the GARCH statistical model.

List of Papers

The following papers have been published to date on the results of some of the work described in this thesis:

1. A. Noiboar and I. Cohen. Anomaly Detection in Three Dimensional Data Based on Gauss Markov Random Field Modeling. Proc. 23rd IEEE Convention of the Electrical and Electronics Engineers, pages 448-451, Israel, Sep. 2004.

2. A. Noiboar and I. Cohen. Two-Dimensional GARCH Model with Application to Anomaly Detection. Proc. 13th European Signal Processing Conference, Turkey, Sep. 2005.

3. A. Noiboar and I. Cohen. Anomaly Detection Based on Wavelet Domain GARCH Random Field Modeling. Submitted to IEEE Transactions on Geoscience and Remote Sensing.

List of Symbols

Abbreviations

AMF	adaptive matched filter
AOI	area of interest
$ARCH$	autoregressive conditional heteroscedasticity
ASD	adaptive subspace detector
$CFAR$	constant false alarm rate
DWT	discrete wavelet transform
EM	expectation maximization
$GARCH$	generalized autoregressive conditional heteroscedasticity
GLR	generalized likelihood ratio
$GLRT$	generalized likelihood ratio test
$GMRF$	Gauss Markov random field
GMM	Gaussian mixture model
$IDWT$	inverse discrete wavelet transform
KLT	Karhunen-Loeve transform
LMM	linear mixing model
MF	matched filter
$MMSE$	minimum mean square error
ML	maximum likelihood
MSD	matched subspace detector
ND	N-Dimensional
PCA	principal component analysis
PDF	probability distribution function
ROI	region of interest
ROC	receiver operating characteristics
SHT	single hypothesis test
SNR	signal to noise ratio
WGN	white Gaussian noise
WSS	wide sense stationarity
iid	independent identically distributed
1D	1-Dimensional
2D	2-Dimensional
3D	3-Dimensional

Alphabetic Symbols

$\mathcal{L}(\mathbf{\Theta})$	conditional sample log likelihood
$\mathcal{L}$	likelihood ratio
$\mathcal{L}_l(s)$	GLR for spatial location s in layer l
$\mathcal{L}(s)$	GLR for spatial location s
A	matrix used for parametrical representation of the inverse covariance matrix
B_0, B_1	projection matrices
B, C	matrices used for structuring A
C	number of possible Gaussian distributions (classes)
$E\{\cdot\}$	expectation
G	number of image chips used for creating an anomaly subspace
$G(w)$	highpass filter
$H(w)$	lowpass filter
H	matrix spanning $\langle H \rangle$
H_l	matrix spanning the anomaly subspace for layer l
H_0	null hypothesis (target absent)
H_1	target presents hypothesis
$\langle H \rangle$	anomaly subspace
H_{K_1}, H_{K_2}	rectangular matrices of size K_1, K_2 respectively
I	horizontal rectangle size
I_j	identity matrix of size j
I_{K_1}, I_{K_2}	identity matrices of size K_1, K_2 respectively
J	vertical rectangle size
$J(\mathbf{c})$	cost function
J_{max}	maximum value of cost function
K	matrix whose columns are the eigenvectors of the covariance matrix of $\mathbf{Y}_{i,j}$
K_1	image feature space size (vertical dimension)
K_2	image feature space size (horizontal dimension)
K_3	image feature space size (depth dimension)
L_1	neighborhood size (vertical dimension)
L_2	neighborhood size (horizontal dimension)
L_3	neighborhood size (depth dimension)
M	number of endmembers
M^ℓ	dimension of transform feature space
$N(\cdot, \cdot)$	normal distribution
N_i	horizontal rectangle size
N_j	vertical rectangle size
N_{ref}	number of reference vectors
N_t	target rectangle dimension (inner rectangle
P_{FA}	probability of false alarm
P_D	probability of detection
P_H	projection matrix onto subspace $\langle H \rangle$
P_{HS}	projection matrix onto subspace $\langle HS \rangle$
P_S	projection matrix onto subspace $\langle S \rangle$
$P_y(s)$	statistical distribution of $y(s)$
$P_{\boldsymbol{y}}(s)$	statistical distribution of $\mathbf{Y}(s)$

p_0	distribution under H_0
p_1	distribution under H_1
Q	whitening transformation
R_1	Radius of W_1
R_2	Radius of W_2
S	matrix spanning $\langle S \rangle$
$\langle S \rangle$	interference subspace
S_l	matrix spanning the interference subspace for layer l
$S_{\boldsymbol{o}}$	total scatter matrix of $\boldsymbol{o}$
T	number of image chips used for creating an interference subspace
U	matrix of orthonormal columns
$VAR\cdot$	variance
W_1	inner circular window
W_2	outer circular window
W	Result of KLT on Y
$\mathbf{W}_{ij}$	column vector representing all data layers at spatial location (i, j) in W
Y	feature space
$\mathbf{Y}(s)$	transformed feature space
$\mathbf{Y}(ij)$	column vector representing all data layers at spatial location (i, j) in Y
X	matrix of lexicographically ordered $\boldsymbol{\eta}_{ij}$
a_k	abundance of the kth endmember
$\boldsymbol{b}$	vector of parameters
$\mathbf{c}$	vector of coefficients
d	distance measure
$d^j_{LH}, d^j_{HL}, d^j_{HH}$	details images
exp	exponent
g_k	highpass filter
h_k	lowpass filter
$h_{\mathbf{i}}$	conditional variance of $\epsilon_{\mathbf{i}}$
$\boldsymbol{h}(\ell, s)$	vector created by row stacking the conditional variance field
$\mathbf{i}$	ND index vector
ℓ_0	conditional likelihood function of the innovation field under hypothesis H_0
ℓ_1	conditional likelihood function of the innovation field under hypothesis H_1
$\boldsymbol{n}_l(s)$	column vector representing $y_l(s)$
o_g	anomaly image chip
p	number of selected top eigenvectors
q	number of degrees of freedom
$\mathbf{q}, \mathbf{p}$	vectors representing the order of an ND GARCH model
q_k, p_k	vector elements of $\mathbf{q}, \mathbf{p}$ respectively
r	an index in $\Omega_{neighbor}$
s	pixel location within the image feature space
$s^j_{LL})$	lowpass subband image
$\boldsymbol{s}_k$	spectra of the kth endmember
s_t	interference image chip
$\boldsymbol{t}$	target signature
$\boldsymbol{v}_j$	reference vectors
$y(s)$	image pixel at location s
y_l	layer l in Y

$y_l(s)$	pixel at spatial location s in y_l
$\boldsymbol{y}(s)$	column vector representing pixel neighborhood centered around s
$y_{\mathbf{i}}$	dependent variable
$u_{\mathbf{i}}$	inner product of $\boldsymbol{x_i}$ and $\boldsymbol{b}$
$\boldsymbol{v}_l(s)$	GMRF vector of size $L_1 \times L_2$
$\boldsymbol{w}$	error vector
$\boldsymbol{w}$	column vector created by consistent ordering of a neighborhood of w_{ij}
w_{ij}	image pixel at location (ij) in a 2D layer
$x(n)$	1D signal
$\boldsymbol{x_i}$	vector of explanatory variables
z_k	elements of vector $\boldsymbol{z}$
$\boldsymbol{z}$	whitened vector
Γ	causal neighborhood
Γ_1, Γ_2	neighborhood sets
Ω_s	sample space
Ω_{image}	support of the image
$\Omega_{neighbor}$	set of indices representing a pixel's neighborhood
Ω	selected subset of layers
Σ_c	covariance matrix of class c
$\mid \Sigma \mid$	determinant of Σ
α_{FA}	desired probability of false alarm
$\alpha(r)$	weight coefficient
$\alpha_{\mathbf{k}}, \beta_{\mathbf{k}}$	ND GARCH parameters
β_h, β_v	MMSE estimation parameters of w_{ij} from its four nearest neighbors
$\boldsymbol{\alpha}, \boldsymbol{\beta}$	vectors of the parameter sets $\{\alpha_{\mathbf{k}}\}_{\mathbf{k}\in\Gamma_1}$ and $\{\beta_{\mathbf{k}}\}_{\mathbf{k}\in\Gamma_2}$ respectively
$\boldsymbol{\delta}$	column vector of the GARCH model parameters
$\epsilon(s)$	innovation process at location s
$\epsilon_{\mathbf{i}}$	ND stochastic process
$\hat{\boldsymbol{\epsilon}_{H_0}}$	maximum likelihood estimate of $\boldsymbol{\epsilon}$ under hypothesis H_0
$\boldsymbol{\varepsilon}$	column vector created by consistent ordering of a neighborhood of ε_{ij}
$\hat{\boldsymbol{\epsilon}_{H_1}}$	maximum likelihood estimate of $\boldsymbol{\epsilon}$ under hypothesis H_1
ε_{ij}	GMRF estimation error for
$\boldsymbol{\epsilon_i}$, $\boldsymbol{h_i}$, $\boldsymbol{y_i}$, $\mathbf{u_i}$	neighborhood column vectors associated with location $\mathbf{i}$
$\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1$	innovation field under hypothesis H_0, H_1 respectively
η_d	threshold
$\eta_{\mathbf{i}}$	stochastic ND process
η_L	threshold
γ	threshold value
$\boldsymbol{\eta}_{ij}$	column vector of first order GMRF neighborhood
$\boldsymbol{\epsilon}$	innovation process
$\boldsymbol{\psi}$	vector of coefficients
$\boldsymbol{\psi}_l(s)$	weight vector for the anomaly subspace
$\psi_{\mathbf{i}}$	information set
$\boldsymbol{\phi}$	vector of coefficients
$\boldsymbol{\phi}_l(s)$	weight vector for the interference subspace
λ	non-centrality parameter
π	pi
π_c	probability of class c

ρ	standard deviation of the innovation process
σ_y	standard deviation of $y(s)$
$\Sigma_{\boldsymbol{y}}$	covariance matrix of $\boldsymbol{y}$
θ_0	parameters of distribution p_0
θ_1	parameters of distribution p_1
$\boldsymbol{\theta}$	vector of parameters
$\hat{\boldsymbol{\theta}}$	MMSE estimation of $\boldsymbol{\theta}$
μ_y	mean of $y(s)$
$\boldsymbol{\mu}_c$	mean of class c
χ^2	chi-square distribution
∞	infinity
$.^T$	transpose
$\sim$	distributes as
$\left\lfloor \frac{\cdot}{\cdot} \right\rfloor$	integer division
$\sim$	statistically distributed as
$()^{\cdot 2}$	vector squaring - element by element

Chapter 1

Introduction

1.1 Background

Anomaly detection algorithms generally consist of some or all of the following stages: selection of an appropriate feature space in which the distinction between the anomaly and the general clutter is possible; selection of a statistical model for the feature space representing the image clutter and selection of a detection algorithm. This last stage implies a selection of an anomaly model, which defines the type of anomaly or anomalies relevant for the application. In this section we present a short review of published work on each of these stages.

1.1.1 Feature Spaces

A proper selection of a feature space, which allows distinction of anomalies from the general clutter is an important part of an anomaly detection algorithm. We next present several image pixel feature spaces, which are constructed directly from the image pixels. We then describe commonly used transform operation. Using

transforms the feature space may be created in two different manners: It may be created by first passing the image data through the transform operation, and the feature space is created based on the data in the transform domain (using similar techniques to those used in the first approach). Alternatively, the feature space may be created by passing an image pixel feature space through the transform operation.

Image pixel feature space : As the name implies this feature space is created based on the image pixels themselves. Kazantsev et al. [20] introduced a feature space based on two circular concentric windows W_1 and W_2 with Radius R_1 and R_2, respectively, $R_1 < R_2$. These two concentric windows act as a moving window, where at each step the center is located on a different image pixel. The center at each step is suspected as part of an anomaly. For each image pixel two feature sets are created. One describing the central pixel (suspected as an anomaly) and the other describing its surrounding neighborhood (assumed to be image clutter). The first set is created by an arbitrary selection of pixel values from W_1 and the second set is created by an arbitrary selection of pixels from W_2. This method is illustrated in Figure 1.1. A similar approach is taken by Schweizer and Moura [40]. We describe their approach originally created for hyperspectral 3D data for the 2D case. Two concentric rectangles serve as the moving window. The inner rectangle represents the unknown or suspected area and is of size $N_t \times N_t$. The outer or perimeter rectangle represents the clutter region and is of size $I \times J$. Each of these regions is further divided into non-overlapping rectangles of size $N_i \times N_j$. At every pixel location two sets of feature vectors are created, one for each region. These vectors are created by a consistent ordering of the image pixels within the $N_i \times N_j$ rectangles. There are $\frac{IJ-N_t^2}{N_i N_j}$ vectors representing the clutter region, and $\frac{N_t^2}{N_i N_j}$ vectors representing the suspected region. This method is illustrated in Figure 1.2. Cohen and Coifman [8]

perform fault detection in 3-Dimensional (3D) seismic data by using real life seismic data (from the Gulf of Mexico). The data is a 3D lattice. They pre-process the data by taking a small analysis cube of size $4 \times 4 \times 15$ that moves through every spatial location in every data layer. At each location, the analysis cube is partitioned into 4 sub-cubes each of size $2 \times 2 \times 15$. The sub-cubes are arranged in a consistent fashion into 4 column vectors. A correlation matrix of the 4 vectors is calculated. The lower triangle coefficients of the correlation matrix are formed into a column vector of size 6×1. In this manner each data layer is turned into a 3D lattice with a depth dimension of 6. Hazel [16] performed scene segmentation and anomaly detection in multi and hyper spectral imagery. He viewed the 3D spectral image data as vector observations on a 2-Dimensional (2D) lattice.

The above mentioned methods create feature vectors representing the pixel to be tested and its surrounding area, assumed to consist of clutter only. We [32] have used a similar technique to create a single feature vector for every pixel location by consistent ordering of neighboring pixels. Let Ω_{image} be the spatial support of an image and Let Y be a 2D image of size $K_1 \times K_2$. Let $y(s)$ represent an image pixel at spatial location $s \in \Omega_{image}$. The feature vector $\boldsymbol{y}(s)$ at location s is created by a consistent ordering of neighboring image pixels, that is: $\boldsymbol{y}(s)$ is a column vector created by consistent ordering of the pixels in an image chip of size $L_1 \times L_2$ centered around s. The methods described above can be applied in the image pixel domain or in a transform domain (examples of which are described next) to create the feature space used for anomaly detection.

Transform feature space : We next describe two common transforms used for anomaly detection, namely the Karhunen-Loeve transform (KLT) and the discrete wavelet transform (DWT).

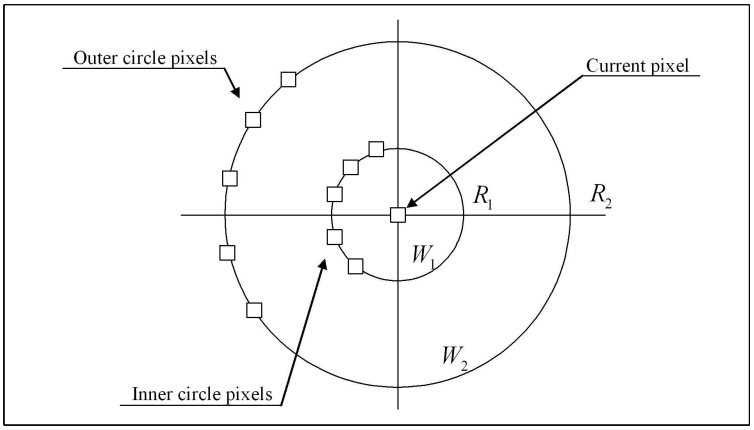


Figure 1.1: Moving window made of two concentric circles. Pixels on the inner circle represent the suspected central pixel. Pixels on the outer circle represent clutter.

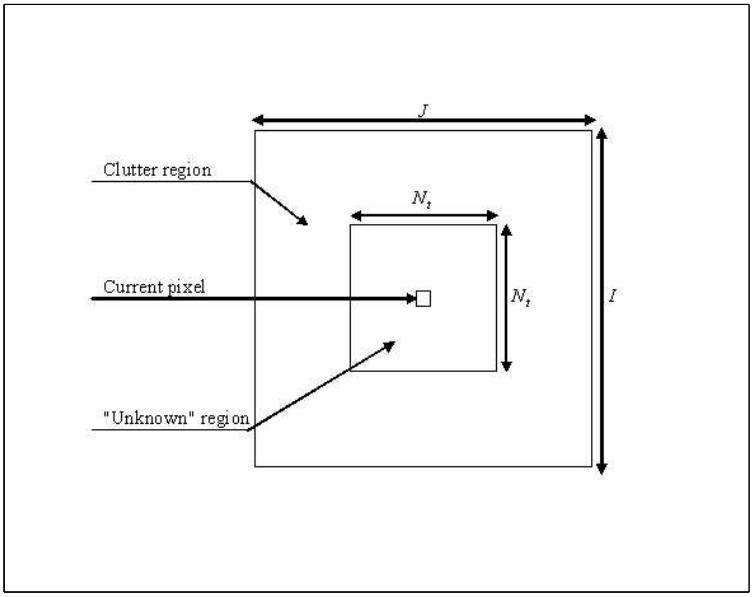


Figure 1.2: Moving window made of two rectangle. The window is divided into two regions: A clutter region and an unknown region

KLT: In the context of anomaly detection the KLT serves mainly for two purposes. First, the KLT is used to transform an n dimensional vector space into an m dimensional vector space, where $m < n$, such that the mean-square magnitude of the error resulting from representing the n dimensional vector $\boldsymbol{y}$ using only m dimensions is minimum. Second, the KLT is used to remove correlation between features. Let $\{\boldsymbol{y}(k)\}_{k=1}^{K_1K_2}$ be a set of independent identically distributed (iid) random vectors with zero mean, and let $\Sigma_{\boldsymbol{y}}$ be their covariance matrix. These vectors are the feature vectors of the K_1K_2 image pixels. Let Y be a matrix containing these feature vectors such that:

$$Y = [\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(K_1K_2)] \; . \tag{1.1}$$

Let K denote a matrix whose columns are the top p eigen vectors of the covariance matrix $\Sigma_{\boldsymbol{y}}$. The top p eigen vectors are selected by the magnitudes of their corresponding eigen values. Let $\boldsymbol{y}^{KLT}(k)$ be the KLT of $\boldsymbol{y}(k)$, then $\boldsymbol{y}^{KLT}(k)$ is given by:

$$\boldsymbol{y}^{KLT}(k) = K^T \boldsymbol{y}(k) \, . \tag{1.2}$$

If $p = n$ then the transformation preserves the feature space dimensions and eliminates correlation between image features, however, if $p < n$ then the resulting feature vectors representing image pixels, that is $\left\{\boldsymbol{y}^{KLT}(k)\right\}_{k=1}^{K_1K_2}$ has a lower dimension as well as uncorrelated features. Fukunaga [10] presents a through theoretic description of the KLT. We have used the KLT to reduce dimensionality and correlation between layers in 3D data as a pre-processing stage for anomaly detection [31]. Goldman and Cohen used the KLT to reduce dimensionality by selecting layers corresponding to the largest eigenvalues of the covariance matrix [12], and also to reduce correlation between layers [13].

DWT: In a 1D wavelet transform a signal is split into two parts: high frequencies

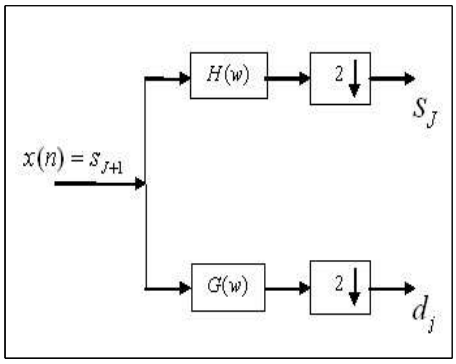


Figure 1.3: One Dimensional wavelet decomposition

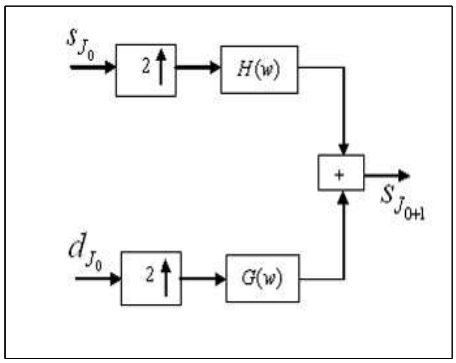


Figure 1.4: One Dimensional wavelet reconstruction

and low frequencies. The low frequencies part is split again into two parts of high and low frequencies. This process is continued an arbitrary number of times, which is usually determined by the application at hand. Furthermore, from these DWT coefficients, the original signal can be reconstructed. This reconstruction process is called the inverse discrete wavelet transform (IDWT). The DWT and IDWT can be mathematically stated as follows. Let $H(w) = \sum_k h_k e^{-jkw}$ and $G(w) = \sum_k g_k e^{-jkw}$ be a lowpass and a highpass filter, respectively, which satisfy a certain condition for reconstruction to be stated later. A signal, $x(n)$ can be decomposed recursively as:

$$s_{j-1,k} = \sum_n h_{n-2k} s_{j,n} \tag{1.3}$$

$$d_{j-1,k} = \sum_n g_{n-2k} s_{j,n}\,, \tag{1.4}$$

for $j = J{+}1, J, \ldots, J_0$, where $s_{J+1,k} = x(k)$, $k \in Z$, $J{+}1$ is the high resolution level index and J_0 is the low resolution level index. The coefficients $s_{J_0,k}, d_{J_0,k}, d_{J_0+1,k}, \ldots, d_{J,k}$ are called the DWT of signal $x(n)$, where $s_{J_0,k}$ is the lowest resolution part of $x(n)$ and $d_{j,k}$ are the details of $x(n)$ at various bands of frequencies. Furthermore, the signal $x(n)$ can be reconstructed from its DWT coefficients recursively:

$$s_{j,n} = \sum_k h_{n-2k} s_{j-1,k} + \sum_k g_{n-2k} d_{j-1,k}\,. \tag{1.5}$$

The above reconstruction is called the IDWT of $x(n)$. To ensure the above IDWT and DWT relationship, the following orthogonality condition on the filters $H(w)$ and $G(w)$ is needed:

$$|\, H(w)\,|^2 + |\, G(w)\,|^2 = 1\,. \tag{1.6}$$

An example of such $H(w)$ and $G(w)$ is the Haar wavelet filters given by:

$$H(w) = \frac{1}{2} + \frac{1}{2} e^{-jw} \tag{1.7}$$

$$G(w) = \frac{1}{2} - \frac{1}{2} e^{-jw}\,. \tag{1.8}$$

The above DWT and IDWT for a 1D signal $x(n)$ can be described in the form of two channel tree-structured filterbank as shown in Figures 1.3 and 1.4 respectively. The DWT and IDWT for 2D images $y(m,n)$ can be similarly defined by implementing the 1D DWT and IDWT for each dimension m and n separately: $DWT_n(DWT_m(y(m,n)))$, which is shown in Figure 1.5. The 2D wavelet transform is a form of multiresolution representation introduced by Mallat [23]. In the analysis phase of the 2D wavelet transform, each row of the input image is separately filtered by H and G. The resulting pair of row-transformed images are likewise filtered in the column direction yielding four subband images is the first octave level. These four images are the lowpass subband image s^j_{LL} and a set of three bandpass subband images $\left\{d^j_{LH}, d^j_{HL}, d^j_{HH}\right\}$, referred to as "details" images. The details images are termed LH (low-high), HL (high-low) and HH (high-high) and correspond to specific, non-overlapping bands in the frequency domain. The "smooth" or LL (low-low) component S^j_{LL} is a lowpass filtered version of the original image and is passed through to the next octave for further subband decomposition. The wavelet transform has been used for anomaly detection and enhancement. Lain et. al [22] used a dyadic wavelet transform in mammography to emphasize mammographic features while reducing the enhancement of noise. Strickland and Hahn [43] used an undecimated wavelet transform for detection of Gaussian objects in Markov noise. We would like to note that the presentation of the wavelet transform is based on the work by Xia et. al. [47] who used the wavelet transform to insert an undetectable watermark into digital imagery.

Once the transform domain has been created, it is regarded as the new data space, and all the above image pixel feature space methods apply for the transform domain. For example, Strickland and Hahn [43] used the wavelet decomposition as

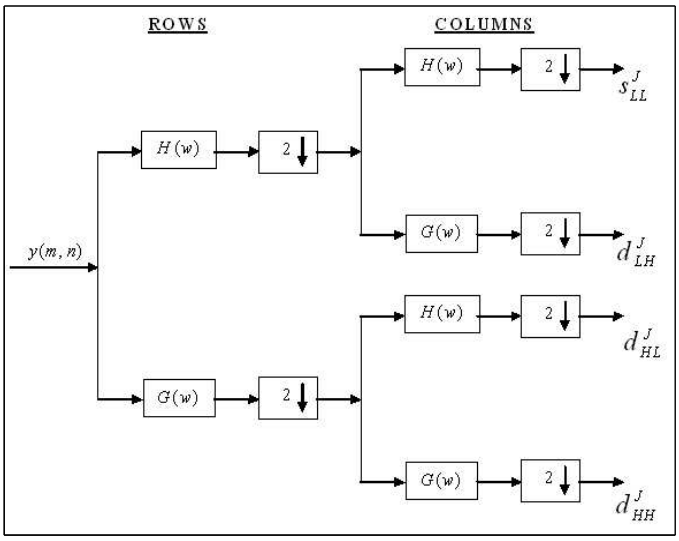


Figure 1.5: Two Dimensional wavelet decomposition

a means for computing a feature set for input to a detector. They created a feature vector for every spatial location consisting of the deferent detail subbands of the wavelet representation. The reverse order is also possible, that is, creating image pixel feature vectors and transforming the resulting feature space into a new feature space using one of the above mentioned methods. For example, Goldman and Cohen [12] created neighborhood feature vector and used the KLT to select only the most dominant principle components of the observation vectors. We now turn to explore statistical models representing the feature space of the image clutter.

1.1.2 Statistical Models of Natural Clutter

Statistical models describing the natural clutter in the selected feature space domain are usually the next step after defining the feature space in an anomaly detection scheme. In the following we describe some of the well studied and commonly applied statistical models in the field of anomaly detection.

The Gaussian distribution: is a common basis for feature space statistical models due to its mathematical tractability. In a Gaussian distribution the statistical distribution of every pixel $y(s)$ in a 1D feature space is:

$$P_y(s) = \frac{1}{\sqrt{(2\pi)}\sigma_y} \exp\left\{-\frac{(y(s)-\mu_y)^2}{2\sigma_y^2}\right\}, \tag{1.9}$$

where σ_y and μ_y are the standard deviation and mean of the random variable y.

In a 2D Gaussian feature space, every pixel location s is represented by a vector of features and the multivariate Gaussian distribution applies, such that the features vector $\boldsymbol{y}(s)$ at every location s within the image is distributed as:

$$P_{\boldsymbol{y}}(s) = (2\pi)^{-L_1 L_2/2} \mid \Sigma_{\boldsymbol{y}} \mid^{-1/2} \times \exp\left\{-\frac{1}{2}(\boldsymbol{y}(s)-\boldsymbol{\mu_y})^T \Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{y}(s)-\boldsymbol{\mu_y})\right\}, \tag{1.10}$$

where $\Sigma_{\boldsymbol{y}}$ and $\boldsymbol{\mu_y}$ are the covariance matrix and mean of the random feature vector $\boldsymbol{y}$, respectively. Ashton [1] performed subpixel anomaly detection in multispectral infrared imagery. A probabilistic background model is formed by using an adaptive Bayesian classification algorithm. Ashton assumes a multivariate Gaussian distribution of each pixel class in spectral space.

The Gaussian mixture model (GMM): The assumption of single component embedded in the multivariate Gaussian model leads to strict requirements for the phenomenon characteristics: a single basic class, which smoothly varies around the class mean. The smooth behavior is not typically the most significant problem, but the assumption of unimodality is. For multimodally distributed features the unimodality assumption may cause an intolerable error to the estimated probability distribution function (PDF). For a multimodal random variable, whose values are generated by one of several randomly occurring independent sources instead of a single source, a finite mixture model can be used to approximate the true PDF. If the Gaussian form

is sufficient for single sources, then a GMM can be used in the approximation. Stein *et al* [42] used a GMM for modeling hyperspectral imagery. They state that the GMM is a method of characterizing image clutter obtained from nonhomogeneous, multi-component scenes. This approach models each feature space vector having one of C possible multivariate Gaussian distributions (classes). The PDF of the scene is a Gaussian mixture distribution:

$$P_{\boldsymbol{y}}(s) = \sum_{c=1}^{C} \pi_c N(\boldsymbol{y} \mid \boldsymbol{\mu}_c, \Sigma_c);\ \pi_c \geq 0;\ \sum_{c=1}^{C} \pi_c = 1\,, \tag{1.11}$$

where π_c is the probability of class c. The GMM PDF can be interpreted as a weighted sum of Gaussians, where π_c is the weight of the cth component. The GMM PDF is completely defined by the parameter list: $\boldsymbol{\Theta} = \{\pi_1, \boldsymbol{\mu}_1, \Sigma_1, \ldots, \pi_C, \boldsymbol{\mu}_C, \Sigma_C\}$. A review on the GMM model and its estimation can be found in the work by Paallanen et. al. [33].

The linear mixing model (LMM): The basic premises of linear mixture modeling are that within a given scene three assumptions are met: first, the surface is dominated by a small number of materials with relatively constant spectra (endmembers). Second, most of the spectral variability within the scene results from varying proportions of the endmembers, and third, the mixing relationship is linear if the endmembers are arranged in spatially distinct patterns. In the LMM model the spectrum of a mixed pixel is represented as a linear combination of components spectra (endmembers). The weight of each endmember spectrum (abundance) is proportional to the fraction of the pixel area covered by the endmember. If there are K_3 spectral bands, the spectrum of the pixel and the spectra of the endmembers can be represented by K_3 dimensional vectors. Therefore, the general equation for mixing by area

is given by:

$$\boldsymbol{y} = \sum_{k=1}^{M} a_k \boldsymbol{s}_k + \boldsymbol{w} = \mathbf{S}\mathbf{a} + \boldsymbol{w} \tag{1.12}$$

$$\mathbf{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_M] \tag{1.13}$$

$$\mathbf{a} = [a_1, a_2, \ldots, a_M], \tag{1.14}$$

where, $\boldsymbol{y}$ is the spectrum of the mixed pixel, $\boldsymbol{s}_k$ is the spectra of the endmembers, a_k is the abundances of the endmembers, M the number of endmembers and $\boldsymbol{w}$ an K_3 dimensional error vector accounting for lack-of-fit and noise effects. Physical considerations dictate the following constraints:

$$a_k \geq 0 \qquad \text{(nonnegativity constraint)} \tag{1.15}$$

$$\sum_{k=1}^{M} a_k = 1 \qquad \text{(additivity constraint)}, \tag{1.16}$$

which can be enforced, if necessary to guarantee meaningful parameter values. These two constraints are often ignored in order to simplify calculations as in Manolakis *et al.* [25] where the LMM was applied for the problem of hyperspectral subpixel target detection. Another example of using the LMM in the context of anomaly detection can be found in the work by Stein *at el* [42] who used an LMM for detecting anomalies in hyperspectral imagery.

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 2D GMRF has been introduced by Woods [45]. It assumes a stationary image background where every image pixel is represented as a weighted sum of neighboring pixels and additive colored noise (the innovation process). Let $\Omega_{neighbor}$ be a given set of indices representing the neighborhood of a pixel. We denote

the weight coefficient of a neighbor $r \in \Omega_{neighbor}$ by $\alpha(r)$ and the innovation process by $\epsilon(s)$. Assuming an image Y can be modeled as an GMRF, a pixel $y(s)$ in the image is related to its neighboring pixels as follows:

$$y(s) = \sum_{k \in \Omega_{neighbor}} \alpha(k) y(s+k) + \epsilon(s) \tag{1.17}$$

The innovation process is spatially correlated with covariance given by:

$$E\left\{\epsilon(s)\epsilon(s+k)\right\} = \begin{cases} \rho^2, & \text{if } k = (0,0) \\ -\alpha(k)\rho^2, & \text{if } k \in \Omega_{neighbor} \\ 0, & \text{otherwise.} \end{cases} \tag{1.18}$$

Kashyap and Chellappa [17] showed that the correlation structure imposes symmetry on the neighborhood set. That is, $k \in \Omega_{neighbor}$ implies $-k \in \Omega_{neighbor}$ and $\alpha(k) = \alpha(-k)$.

A good review of multiresolution Markov Models for signal and image processing can be found in [44].

The 1D GARCH: Bollerslev [5] introduced the 1-Dimensional (1D) Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, which is often used as a statistical model for time series. It is an extension to the Autoregressive Conditional Heteroscedasticity (ARCH) model introduce by Engle [9]. The 1D GARCH has been shown to be useful in modeling different economic phenomena.

Let ϵ_t denote a real-valued discrete-time stochastic process, and ψ_t the information set of all information through time t. The 1D $GARCH(p,q)$ process is given by:

$$\epsilon_t \mid \psi_t \sim N(0, h_t) \tag{1.19}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}\,, \tag{1.20}$$

where

$$\begin{aligned} p &\geq 0\,, q > 0 \\ \alpha_0 &> 0\,, \alpha_i \geq 0\,, i = 1, \ldots, q \\ \beta_i &\geq 0\,, i = 1, \ldots, p\,. \end{aligned}$$

For $p = 0$ the process reduced to the $ARCH(q)$ process, and for $p = q = 0$ ϵ_t is simply white noise.

The GARCH model allows for the conditional variance to change as a function of past squared field values (ϵ_{t-i}^2) and past conditional variance values (h_{t-i}). This behavior creates a heavy tailed distribution characterized by clustering of innovations.

So far we have discussed feature spaces and statistical models. The final steps in the anomaly detection scheme are to define an anomaly model and an appropriate detection algorithm.

1.1.3 Anomaly Detection

The anomaly detection process can be regarded as a problem of classifying a sample into one of two classes: image clutter or an anomaly, with the assumption of low-probability anomalies. In practice, it happens that one class (the clutter) is well defined while the other (the anomaly) is not. For example, if the statistical model for the image clutter is known, its parameters can be estimated from the image data. However, without *a priory* information about the anomalies, such estimation is impossible for the anomaly class. We next present a set of detection algorithms. These algorithms differ, among other things, in their representation of the two classes.

Single Hypothesis Test: The single hypothesis test (SHT) have been proposed to solve the problem of an undefined anomaly. Typically an SHT involves measuring the distance of the sample at hand from the clutter mean (normalized by the clutter covariance matrix), and applying a threshold to determine if it is or is not an anomaly. That is:

$$d^2 = (\boldsymbol{y}(s) - \boldsymbol{\mu_y})^T \Sigma_{\boldsymbol{y}}^{-1} (\boldsymbol{y}(s) - \boldsymbol{\mu_y}) \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta_d \,, \tag{1.21}$$

where H_0 and H_1 represent target absent and target present hypotheses respectively and η_d is the selected threshold.

Define $\boldsymbol{z} = Q^T(\boldsymbol{y} - \boldsymbol{\mu_y})$, where Q is a whitening transformation, then the distance d^2 can be written as:

$$d^2 = \boldsymbol{z}^T \boldsymbol{z} = \sum_{k=1}^{L_1 L_2} z_k^2 \,, \tag{1.22}$$

where z_k are the elements of the vector $\boldsymbol{z}$. Since the expected vector and covariance matrix of $\boldsymbol{z}$ are $\mathbf{0}$ and I respectively, the z_k's are uncorrelated, $E\{z_k\} = 0$ and $VAR\{z_k\} = 1$. When $\boldsymbol{y}$ is Gaussian, d^2 is a sum of squared independent, normally distributed random variables and therefor is chi-square distributed with $L_1 L_2$ degrees of freedom, as follows:

$$d^2 \sim \chi^2_{L_1 L_2}(0) \,. \tag{1.23}$$

It is important to note, that the SHT works well when the dimension of the data ($L_1 L_2$) is very low (such as 1 or 2). As the dimension of the data increases, the error of the SHT increases significantly. The SHT assumes no knowledge about the anomalies. If information about the anomalies is made available *a priory* it cannot be incorporated into the detection scheme. A detailed discussion on the SHT and can

be found in Fukunaga [10].

Matched Filter Detector: As an opposite approach to the SHT let us now consider the detection problem specified by the following hypotheses:

$$\begin{aligned} H_0 : \boldsymbol{y} \sim &\quad N(\boldsymbol{\mu_y}, \Sigma_{\boldsymbol{y}}) \quad \text{Anomaly absent} \\ H_1 : \boldsymbol{y} \sim &\quad N(\boldsymbol{\mu_t}, \Sigma_{\boldsymbol{t}}) \quad \text{Anomaly present}\,, \end{aligned}$$

where the target and background classes follow multivariate normal distributions with different mean vectors and covariance matrices. Unlike the SHT the probability densities are completely specified under each hypothesis. The Likelihood ratio detector is given by the ratio of the conditional probability density functions of the two hypothesis:

$$\mathcal{L} = \frac{P_{\boldsymbol{y}}(\boldsymbol{y}(s) \mid \text{anomaly present})}{P_{\boldsymbol{y}}(\boldsymbol{y}(s) \mid \text{anomaly absent})} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta_d\,. \tag{1.24}$$

If $\mathcal{L}$ is larger than the threshold η_d, the "anomaly present" hypothesis is accepted. Computing the natural logarithm for the above presented PDFs leads to the quadratic detector:

$$\mathcal{L} = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu_y})^T \Sigma_{\boldsymbol{y}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu_y}) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu_t})^T \Sigma_{\boldsymbol{t}}^{-1} (\boldsymbol{y} - \boldsymbol{\mu_t})\,. \tag{1.25}$$

which compares the Mahalanobis distances of the observed feature vector $\boldsymbol{y}$ from the centers of the two classes. The required threshold η_d is determined from:

$$P_{FA} = \int_{\eta_d}^{\infty} P(\mathcal{L} \mid H_0) d\mathcal{L} = \alpha_{FA}\,, \tag{1.26}$$

where α_{FA} is the desired probability of false alarm. As a result of the quadratic

mapping, the distribution of the random variable $\mathcal{L}$ (detector output) is not normal, which makes the performance evaluation of the detector difficult.

If the target and background classes have the same covariance matrix, that is, $\Sigma_{\boldsymbol{y}} = \Sigma_{\boldsymbol{t}}$, the quadratic terms in (1.25) disappear, and the likelihood ration detector (1.24) becomes:

$$\mathcal{L} = (\boldsymbol{\mu_t} - \boldsymbol{\mu_y})^T \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{y} \,. \tag{1.27}$$

This is a linear detector:

$$\mathcal{L} = \mathbf{c}^T \boldsymbol{y} = \sum_{k \in \Omega_{image}} c_k y_k \,, \tag{1.28}$$

which is specified by the coefficient vector:

$$\mathbf{c} = \Sigma_{\boldsymbol{y}}^{-1} (\boldsymbol{\mu_t} - \boldsymbol{\mu_y}) \,. \tag{1.29}$$

The detector output is a linear combination of normal random variables and is therefor normally distributed. This result simplifies the evaluation of the detector and the computation of detection thresholds using (1.26). This detector is known as Fisher's linear discriminant and is widely used in pattern recognition application. The same result is well known in the the communication and signal processing literature, where it is termed the matched filter (MF). There the MF is usually derived by maximizing the cost function:

$$J(\mathbf{c}) = \frac{\left[E\left\{\mathcal{L} \mid H_1\right\} - E\left\{\mathcal{L} \mid H_0\right\}\right]^2}{\mathrm{var}\left\{\mathcal{L} \mid H_0\right\}} = \frac{\left[\mathbf{c}^T (\boldsymbol{\mu_t} - \boldsymbol{\mu_y})\right]^2}{\mathbf{c}^t \Sigma_{\boldsymbol{y}} \mathbf{c}} \,, \tag{1.30}$$

which measures the distance between the means of two normal distributions in units of the common variance. The maximum is obtained by substituting (1.29) into (1.30):

$$J_{max} = (\boldsymbol{\mu_t} - \boldsymbol{\mu_y})^T \Sigma_{\boldsymbol{y}}^{-1} (\boldsymbol{\mu_t} - \boldsymbol{\mu_y}) \,, \tag{1.31}$$

which is the mahalanobis squared distance between the means of the anomaly and clutter distributions.

Adaptive Matched Filter Detector: The detector in (1.27) requires the mean vector and the covariance matrix of the anomaly and clutter. In practical applications, these quantities are often unavailable. To overcome this difficulty we next present the adaptive matched filter (AMF) [24]. The above mentioned quantities are estimated from the available data. Once again, under the assumption of low-probability anomalies, we can use the available data $\boldsymbol{y}(s);\ s \in \Omega_{image}$ to determine the maximum likelihood estimates of the mean vector and covariance matrix of the clutter:

$$\begin{aligned} \boldsymbol{\mu_y} &= \frac{1}{L_1 L_2} \sum_{s \in \Omega_{image}} \boldsymbol{y}(s) \\ \Sigma_{\boldsymbol{y}} &= \frac{1}{L_1 L_2} \sum_{s \in \Omega_{image}} (\boldsymbol{y}(s) - \boldsymbol{\mu_y})(\boldsymbol{y}(s) - \boldsymbol{\mu_y})^T . \end{aligned}$$

Unfortunately, there is usually not sufficient training data to determine the mean and covariance of the target. Typically, a target signature $\boldsymbol{t}$, from a library of the mean of a small number of known target pixels observed under the same conditions is used. The resulting AMF is given by:

$$d(s) = \frac{\boldsymbol{t}^T \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{y}(s)}{\boldsymbol{t}^T \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{t}} , \tag{1.32}$$

where usually the data mean is removed from the anomaly and clutter data.

If we know the "true" covariance matrix $\Sigma_{\boldsymbol{y}}$, the output L under the "target absent" hypothesis, is distributed as $L \sim N\left(\mu_L, (\boldsymbol{t}^T \Sigma_{\boldsymbol{y}}^{-1} \boldsymbol{t})^{-1}\right)$, where $\mu_L = E\{L\}$. When the required means and covariances are estimated from the data, the resulting estimates are random quantities. If we treat them as constants, we can determine the class-conditional distribution of the detector output as in the known statistics case. However, the correct approach is to treat the estimated means and covariances as random and determine the unconditional distribution of L. Unfortunately, the

derivation of unconditional distributions is a very difficult problem even under the most simplified assumptions. For more details on the matched filter detector and adaptive matched filter detector we refer the reader to the work by Manolakis and Shaw [24].

RX Detector: Reed and Xiaoli developed the RX detector [34]. It is a generalized likelihood ratio test (GLRT) based detector, where the clutter model parameters are estimated based on test and reference data. Let $\{\boldsymbol{v}_j \in C^{L_1 L_2} \mid 1 \leq j \leq N_{ref}\}$ be a set of N_{ref} iid sample vectors having pdf $p_0(\cdot, \theta_0)$. At every pixel location s a feature vector $\boldsymbol{y}(s)$ is to be classified as arising from either PDF $p_1(\cdot, \theta_1)$ or $p_0(\cdot, \theta_0)$ (hypothesis H_1 or H_0 respectively). The GLRT is:

$$\mathcal{L} = \frac{\max\limits_{\theta_1} \left(p_1(\boldsymbol{y}(s), \theta_1) p_0\left(\{\boldsymbol{v}_j \mid 1 \leq j \leq N_{ref}\}, \theta_1\right)\right)}{\max\limits_{\theta_0} \left(p_0(\boldsymbol{y}(s), \theta_0) p_0\left(\{\boldsymbol{v}_j \mid 1 \leq j \leq N_{ref}\}, \theta_0\right)\right)} \begin{array}{c} H_1 \\ > \\ < \\ H_0 \end{array} \eta_{\mathcal{L}}\,, \tag{1.33}$$

where $\eta_{\mathcal{L}}$ is a threshold. Reed and Yu [34] developed an GLRT for multidimensional image data assuming that the target signal and the covariance of the clutter are unknown. The data under the two hypotheses is modeled as:

$$\begin{aligned} H_0 : \boldsymbol{y} \sim \quad & N(\boldsymbol{\mu_y}, \Sigma_{\boldsymbol{y}}) \quad \text{Anomaly absent} \\ H_1 : \boldsymbol{y} \sim \quad & N(\boldsymbol{\mu_t}, \Sigma_{\boldsymbol{y}}) \quad \text{Anomaly present}\,, \end{aligned}$$

such that $\Sigma_{\boldsymbol{y}}$ and $\boldsymbol{\mu_t}$ are unknown. The GLRT is:

$$\mathcal{L} = (\boldsymbol{y}-\boldsymbol{\mu_y})^T \left[\frac{N_{ref}}{N_{ref}+1}\hat{\Sigma_{\boldsymbol{y}}} + \frac{1}{N_{ref}+1}(\boldsymbol{y}-\boldsymbol{\mu_y})(\boldsymbol{y}-\boldsymbol{\mu_y})^T\right]^{-1}(\boldsymbol{y}-\boldsymbol{\mu_y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathcal{L}}\,, \tag{1.34}$$

where $\hat{\Sigma_{\boldsymbol{y}}} = \frac{1}{N_{ref}}\sum_{j=1}^{N_{ref}}(\boldsymbol{v}_j-\boldsymbol{\mu_y})(\boldsymbol{v}_j-\boldsymbol{\mu_y})^T$ is the sample covariance matrix of the reference data. As $N_{ref}\to\infty$, $\mathcal{L}$ converges to:

$$\mathcal{L} = (\boldsymbol{y}-\boldsymbol{\mu_y})^T \left[\hat{\Sigma_{\boldsymbol{y}}}\right]^{-1}(\boldsymbol{y}-\boldsymbol{\mu_y}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathcal{L}}\,. \tag{1.35}$$

Asymptotic forms, as $N_{ref}\to\infty$, of the probability distributions of $\mathcal{L}$,under the H_0 and H_1 hypotheses are given in terms of the χ^2 and noncentral χ^2 densities, respectively. Let $\chi^2_n(\cdot)$ denote the χ^2 density on n degrees of freedom, and let $\chi^2_n(\cdot,\lambda)$ denote the noncentral χ^2 density on n degrees of freedom having noncentrality parameter λ. The probability distribution of $\mathcal{L}$ under H_0 and H_1 are:

$$\mathcal{L} \sim \begin{cases} \chi^2_{\tau L_1 L_2}(0), & \text{under } H_0 \\ \chi^2_{\tau L_1 L_2}((\boldsymbol{\mu_t}-\boldsymbol{\mu_y})^T\Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{\mu_t}-\boldsymbol{\mu_y})), & \text{under } H_1 \end{cases}, \tag{1.36}$$

where $\tau = 1, 2$ if the data are real or complex values, respectively. The probabilities of false alarm and detection are:

$$P_{FA} = 1 - P\left[\chi^2_{\tau L_1 L_2}(0) \le \eta_{\mathcal{L}}\right] \tag{1.37}$$

$$P_D = 1 - P\left[\chi^2_{\tau L_1 L_2}((\boldsymbol{\mu_t}-\boldsymbol{\mu_y})^T\Sigma_{\boldsymbol{y}}^{-1}(\boldsymbol{\mu_t}-\boldsymbol{\mu_y})) \le \eta_{\mathcal{L}}\right]. \tag{1.38}$$

Note that the distribution of the test statistics under H_0 is independent of the unknown parameters, and thus the test statistics has the constant false alarm rate (CFAR) property. Chang and Chiang [7] used the RX detector for anomaly detection in hyperspectral imagery.

Matched Subspace Detector : Scharf and Friedlander [38] extended the GLRT to situations where the anomaly signature is unknown and assumed to be in a subspace. They formulated the MSD for the general problem of detecting subspace signals in subspace interference and additive white Gaussian noise (WGN). Let $\langle H \rangle$ denote the anomaly subspace, spanned by the columns of a matrix H and let $\langle S \rangle$ denote the interference subspace, spanned by the columns of a matrix S. We denote the additive iid Gaussian noise by $\boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \rho^{\mathbf{2}}\mathbf{I})$. The problem is to determine whether the sample vector $\boldsymbol{y}$ contains an anomaly signal. The anomaly signal can be described as a linear combination of the columns of H, that is, the anomaly signal is: $H\boldsymbol{\psi}$, where $\boldsymbol{\psi}$ is a vector of coefficients. The interference signal is described similarly, using the matrix S and the coefficients vector $\boldsymbol{\phi}$. We define two hypotheses, H_0 and H_1 which indicate, respectively, absence and presence of the anomaly signal in the vector $\boldsymbol{y}$:

$$H_0 : y = S\boldsymbol{\phi} + \boldsymbol{\epsilon} \tag{1.39}$$

$$H_1 : y = H\boldsymbol{\psi} + S\boldsymbol{\phi} + \boldsymbol{\epsilon}\,. \tag{1.40}$$

Let P_S denote the projection of a vector onto the subspace $\langle S \rangle$:

$$P_S\boldsymbol{y}(s) = S(S^TS)^{-1}S^T\boldsymbol{y}(s)\,, \tag{1.41}$$

and let P_{HS} denote the projection of a vector onto the subspace $\langle HS \rangle$, spanned by the columns of the concatenated matrix $[HS]$. The maximum likelihood estimates

of the additive noise vector $\boldsymbol{\epsilon}$ under H_0 and under H_1 are denoted by $\boldsymbol{\epsilon}_{\hat{H}_0}$ and $\boldsymbol{\epsilon}_{\hat{H}_1}$ respectively. These estimates are obtained by subtracting from $\boldsymbol{y}$ the components which lie in the signal and interference subspaces as follows:

$$\boldsymbol{\epsilon}_{\hat{H}_0} = (I - P_S)\boldsymbol{y} \tag{1.42}$$

$$\boldsymbol{\epsilon}_{\hat{H}_1} = (I - P_{HS})\boldsymbol{y}\,. \tag{1.43}$$

The detection problem can be formulated as an GLRT between H_0 and H_1. The log-likelihood ratio is given by:

$$\mathcal{L}(s) = 2\log\left[\frac{P(\boldsymbol{\epsilon}(s) \mid H_0)}{P(\boldsymbol{\epsilon}(s) \mid H_1)}\right] \tag{1.44}$$

$$= 2\log\left[\frac{\exp\left(\frac{[\boldsymbol{\epsilon}_{\hat{H}_0}]^2}{2\rho^2}\right)}{\exp\left(\frac{[\boldsymbol{\epsilon}_{\hat{H}_1}]^2}{2\rho^2}\right)}\right] \tag{1.45}$$

$$= \frac{1}{\rho^2}\boldsymbol{y}^T(P_{HS} - P_S)\boldsymbol{y}\,. \tag{1.46}$$

The signal to noise ratio (SNR) is the ratio between the signal and the noise in terms of intensity. We define the SNR as the second power of the ratio between the signal which do not lie in the interference subspace, and the standard deviation of the noise:

$$SNR = \frac{1}{\rho^2}[H\boldsymbol{\psi}]^T[I - P_S][H\boldsymbol{\psi}]\,. \tag{1.47}$$

Let u denote the rank of the anomaly subspace. $\mathcal{L}$ is a sum of squared independent normally distributed variables and therefor is chi-square distributed with u degrees of freedom:

$$\mathcal{L} \sim \begin{cases} \chi^2_u(0), & \text{under } H_0 \\ \chi^2_u(SNR), & \text{under } H_1 \end{cases}\,. \tag{1.48}$$

Under hypothesis H_1, the non-centrality parameter of the chi-square distribution of $\mathcal{L}$ is equal to the SNR [37]. The decision rule is based on thresholding the log-likelihood

ratio using the threshold $\eta_{\mathcal{L}}$:

$$\mathcal{L} \underset{H_0}{\overset{H_1}{\gtrless}} \eta_{\mathcal{L}} \,. \tag{1.49}$$

The probabilities of false alarm and detection are:

$$P_{FA} = 1 - P[\chi^2_u(0) \leq \eta_{\mathcal{L}}] \tag{1.50}$$

$$P_D = 1 - P[\chi^2_u(SNR) \leq \eta_{\mathcal{L}}] \,. \tag{1.51}$$

Kraut *et al.* [21]extended the MSD to the the case were the noise (clutter) covariance matrix is unknown. The extended detector is named adaptive subspace detector (ASD).

1.2 Motivation and Goals

The Gaussian distribution is a common basis for feature space statistical models. This is due to the mathematical tractability of the Gaussian statistics. It is of no surprise that anomaly detection is often applied using a Gaussian distribution. Consider as an example the GMRF based algorithm presented by Goldman and Cohen [13]. This algorithm is based on 2D GMRF modeling of uncorrelated layers in a multiscale representation of the image. Correlation between layers is reduced by means of the 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 Figure 1.6. The side-scan sonar images presented in this work are from the Sonar-5 database collected by the Naval Surface Warfare Center (NSWC) 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 Figure 1.6) 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 to the reverberation from the seabed. A description of the acquisition process of side-scan sonar imagery and a discussion on the various shapes of mine like objects in such imagery is presented by Reed et. al. in [35, 36]. 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 work 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 priory*, 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 Figure 1.6 results in a high false alarm rate. This high false alarm rate is demonstrated at the bottom row of Figure 1.6, 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 Figures 1.6 (a) and (b). The kurtosis (forth moment divided by the square second moment) is a measure of fat tail behavior as noted by Buccigrossi and Simoncelli [6]. The sample kurtosis of the multiresolution feature space of these two figures is 9.8 and 10.9 respectively. The expected kurtosis value for the Gaussian distribution is 3. The high kurtosis values of the feature space for these two images implies a distribution with much heavier tails than the Gaussian distribution. Second, let us examine Figure 1.6 (c). The sample kurtosis value for the feature space of this image

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 a phenomena where large changes tend to follow large changes, and small changes tend to follow small changes. In either case, the changes are typically of unpredictable sign. Clustering of innovations is clearly seen in the image itself and is also apparent in the layers of the multiresolution representation as demonstrated in Figure 1.7 (c)-(d). This phenomena is also present in Figure 1.6 (b) and in its corresponding multiresolution layers presented in Figure 1.7 (a)-(b). We note that the clustering of innovations phenomena demonstrated in Figure 1.7 appears at the same spatial location in the different multiscale representation layers. This demonstrates scale to scale dependency. These 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 Goldman and Cohen [13] and therefore call for an alternative statistical model. For that purpose we introduce the multidimensional GARCH model.

The 1D GARCH model introduced by Bollerslev [5] has been used for modeling financial time series. It is an extension to the ARCH model introduced by Engle [9]. The GARCH model was designed to capture the fat tails and volatility clustering associated with financial time series. In a GARCH model the conditional variance at every location is dependent on conditional variance values and squared field values of neighboring locations yielding clustering of innovations. The notion of conditional variance and neighboring locations is mathematically defined in Chapter 3. For details on the 1D GARCH model we refer the reader to the work by Bollerslev [5] and to the book by Hamilton [14].

We begin by demonstrating the heavy tails and volatility clustering properties of

the multidimensional GARCH model. Figure 1.8 shows seven layers of a 3D synthetic GARCH data (the anomaly planted in these layers is discussed later in Chapter 4). The sample kurtosis of the complete data set is 26.87. The kurtosis of each of the 2D layers in Figure 1.8 is shown in Table 1.1. The kurtosis values are much larger than the value of 3 characterizing the Gaussian distribution, demonstrating the heavy tails property of the GARCH model. The heavy tails of this sample data can also be viewed from the data's histogram shown in Figure 1.9. The volatility clustering property of the GARCH model, which is due to the special structure of the conditional variance, is apparent from Figure 1.8, where clustered areas of high variations in gray-scale levels are easily noticed. To further demonstrate it Figure 1.10 shows the seven layers of the conditional variance field based on the estimated model parameters. Darker areas in Figure 1.10 represent areas of high conditional variance. These darker areas appear in clusters and not as scattered pixels. The match between darker areas in Figure 1.10 and the areas of clustered variations in Figure 1.8 is obvious.

In this work we introduce the multidimensional GARCH model. Since we assume the multiscale feature space follows a GARCH distribution, we are faced with the challenge of developing an appropriate detection approach. Scharf and Frienlander [38] developed an MSD for the detection of signals in subspace interference and additive WGN. Here we introduce a set of multiscale MSDs operating in subspace interference and additive GARCH noise. Since the statistical model is not limited to 2 dimensions, our MSD utilizes the correlation within and between layers, meaning that detection at each location in the feature space may be based on feature space data from adjacent layers, and not limited to a single layer. A separate anomaly subspace is assumed for each feature space layer, thus allowing incorporation of *a priory* information into the detection process. These anomaly subspaces need not be

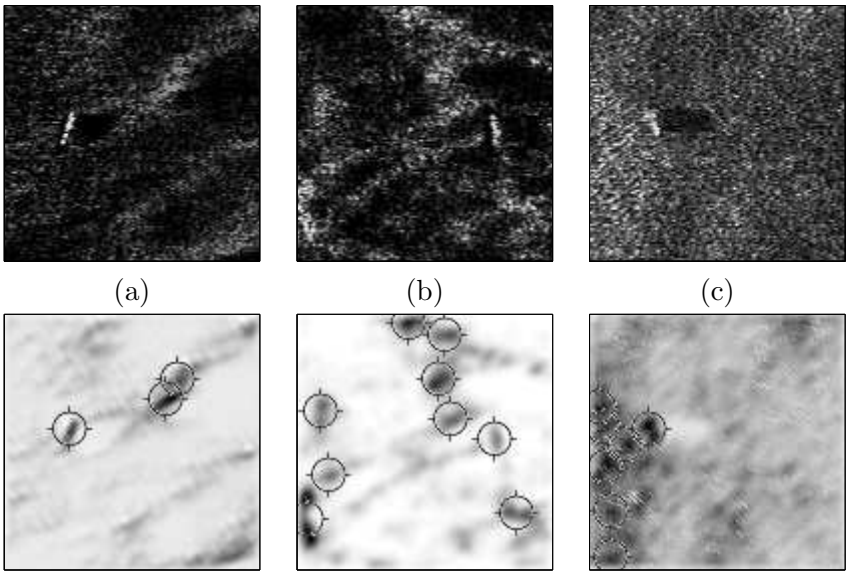
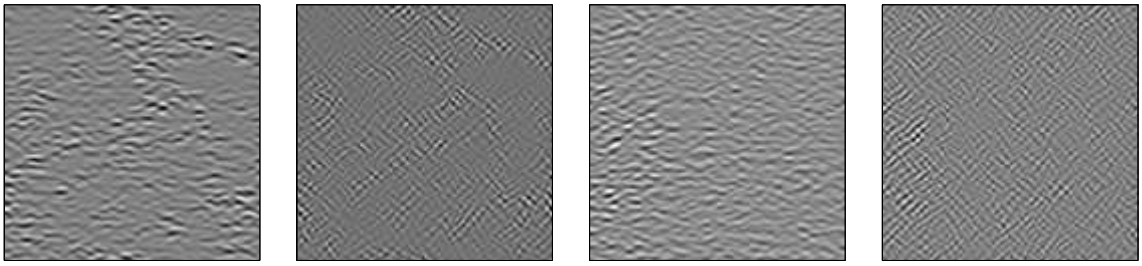


Figure 1.6: Detection results in side-scan sea mine sonar images: (top row) Original side-scan sea-mine sonar images; (bottom row) Detection Results using a GMRF based method.

of the same size thus allowing for greater adaptivity of the anomaly subspace to the characteristics of the feature space, namely: scale and orientation. Recognizing that not all feature space layers contribute evenly to the detection process, our multiscale MSD allows for a selection of the layers to use. This selection can be made *a priory*, or after some processing has been performed and intermediate results of the detection process are available.

1.3 Overview of the Thesis

The original contribution of this thesis starts from Chapter 2, where we present an anomaly detection approach for three dimensional data. In a pre-processing stage, we de-correlate the data layers using the KLT. After de-correlating the data layers, each



(a) (b) (c) (d)

Figure 1.7: Example of layers from an undecimated wavelet transform representation of sea mine sonar images: (a)-(b) Two layers from the multiresolution representation of Figure 1.6(b); (c)-(d) Two layers from the multiresolution representation of Figure 1.6(c).

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

Table 1.1: Kurtosis Values of Each of the 2D Layers in Figure 1.8

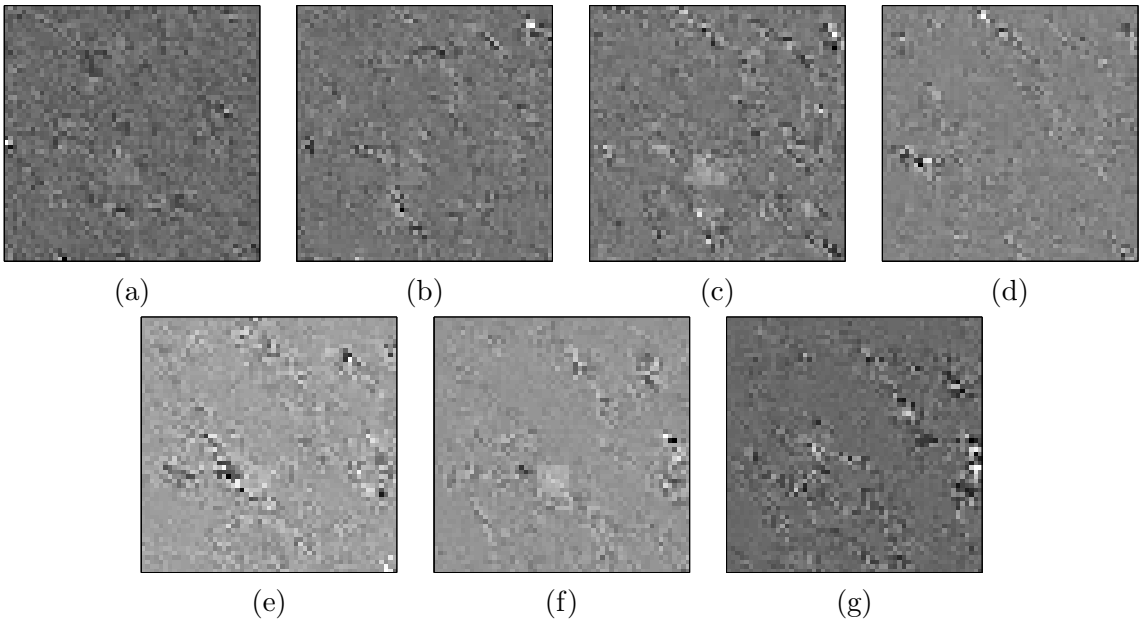


Figure 1.8: Seven layers of a GARCH synthetic image with a Gaussian shaped anomaly

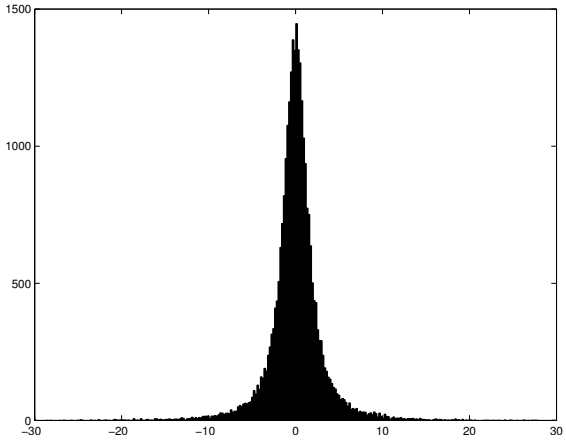


Figure 1.9: Histogram of the 3D GARCH data shown in Figure 1.8. The sample kurtosis of this data is 26.87

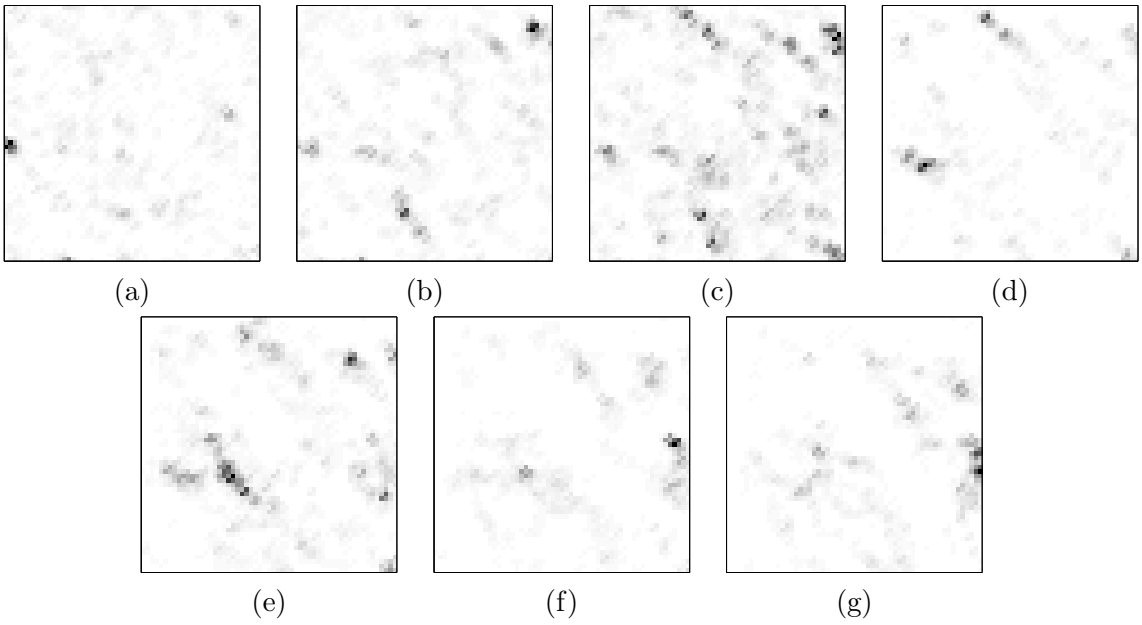


Figure 1.10: Seven layers of the conditional variance field of the synthetic GARCH data presented in Figure 1.8. Darker areas represent higher conditional variance values.

layer is modeled as a GMRF. The GMRF accounts for spatial correlation between neighboring image pixels. We present a least squares model estimation, and present a parametric form of the model inverse covariance matrix. The inverse covariance matrix is later used in the anomaly detection stage, thus saving the computational burden of covariance matrix inversion. Anomaly detection is performed by utilizing a set of MSDs. Our MSD detects anomalies in colored Gaussian noise, and interference subspace. We use a different MSD for each data layer. This approach allows us to incorporate into the anomaly subspace a-priori knowledge about the sensors used to capture the data, or a-priori information about the anomalies. Using such information would yield better detection rates, and lower false alarm rates. However, if such information is not available, detection is carried out based on a general subspace. We utilize the eigen-picture technique for creating the anomaly subspace when a

large training set of anomaly examples is available. The robustness of the proposed detection approach is demonstrated on real life data from two different applications: detection of defects in wafer images and detection of faults in 3D seismic data.

In Chapter 3, we introduce the N-Dimensional (ND) GARCH model. The ND GARCH model is a novel clutter model which is capable of taking into account important characteristics of a 3D 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 3D wavelet-based feature space. The undecimated wavelet transform has the property of translation invariance, which is important in the context of anomaly detection. A 3D GARCH model is used as the underlying statistics of the feature space.

In Chapter 4 we develop a multiscale MSD 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. For every scale and orientation in the wavelet domain (referred to as feature space layers) a separate anomaly subspace is used, thus allowing for better incorporation of *a priory* information into the process. Our multiscale MSD approach takes into 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 is application dependent and independent of the detection algorithm. We demonstrate the potential advantages of the proposed method on synthetic data and real life sea-mine side-scan sonar imagery.

We Conclude in Chapter 5 with a summary and discussion on future research directions.

1.4 Organization

The organization on this thesis is as follows. In Chapter 2, we present an anomaly detection approach for three dimensional data, based on the GMRF statistical model and an appropriate MSD anomaly detection scheme operating in colored noise. In Chapter 3 we introduce the multi-dimensional GARCH statistical model. We present a maximum likelihood model estimation. In chapter 4 we introduce an anomaly subspace detection approach operating in an undecimated wavelet domain, modeled using the multidimensional GARCH model. We analyze the performance of the proposed method and 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. Finally in Chapter 5 we conclude with a summary and discussion on future research directions.

We would like to note that Chapters 2, 3 and 4 are the detailed and extended version of our published materials. Chapter 2 is based on [30, 31], Chapter 3 is based on [30, 32] and Chapter 4 is based on [30].

Chapter 2

Anomaly Detection Based on GMRF Modeling

2.1 Introduction

Anomaly detection in three dimensional data has many practical applications including automatic target detection in multispectral and hyperspectral images, defect detection, detecting of faults in seismic data, etc. Statistical methods in this field assume distinct statistical models for both the background and the anomalies. Statistical parameters for the background are estimated from the 3D data, while statistical parameters for the anomalies are estimated based on some training set, such as in the RX algorithm used by Stein *et al.* [42]. Other methods assume a known anomaly pattern in a random clutter background, such as in the RX algorithm introduced in [34], or the adaptive matched filter detector of Manolakis and Shaw [24]. Some detection algorithms assume the anomalies are those portions of the data that have the worst fit to the background model. Examples of those algorithm are the single

hypothesis tests presented by Schweizer and Moura [40] and the iterative detection procedure of Goldman and Cohen [12]. Another example is the detector presented by Ashton [1], in which image clustering is performed in a preprocessing stage followed by spectral decorrelation and energy detection for each cluster. No knowledge about the anomaly is required. In anomaly detection, Once statistical modeling is accomplished, hypothesis testing is often used for deciding which pixels represent an anomaly and which represent the background. For a more thorough introduction to the field of anomaly detection we refer the reader to Chapter 1 in this thesis.

A major limitation of these methods is the inefficient statistical model for the anomalies. In real life the anomaly has an unknown pattern and one searches for different anomalies within a given data set. Estimating the statistical parameters for the anomalies from a given training data set yields poor detection results when the anomalies differ from the examples in the training set. Estimating these parameters based on the data itself renders the estimation unstable since the anomalies are sparse within the data and their location is unknown. Detecting anomalies based on worst fit to the background model does not allow utilization of *a-priori* information about the anomaly, if such information is available. An additional drawback of many of these methods is their high computational complexity. The computational cost comes from inefficient parameter estimation techniques and from the need to find inverse covariance matrices for the hypothesis testing stage.

In this chapter, we present an anomaly detection approach for three dimensional data, such as multispectral and hyperspectral imagery (referred to as spectral imagery for convenience), seismic data and images taken from different sensors, which are spatially registered and grouped into a 3D data format. In a pre-processing stage,

we de-correlate the data layers using the Karhunen-Loeve transform (KLT). After de-correlating the data layers, each layer is modeled as a GMRF. The GMRF accounts for spatial correlation between neighboring image pixels. We present an efficient least squares model estimation, and present a parametric form of the model inverse covariance matrix. The inverse covariance matrix is later used in the anomaly detection stage, thus saving the computational burden of covariance matrix inversion. Anomaly detection is performed by utilizing a set of matched subspace detectors. The MSD was originally developed by Scharf and Friedlander [38] for the detection of signals in subspace interference and additive WGN. Our MSD detects anomalies in colored Gaussian noise, and interference subspace. We use a different MSD for each data layer. This approach allows us to incorporate into the anomaly subspace a-priori knowledge about the sensors used to capture the data, or a-priori information about the anomalies. Using such information would yield better detection rates, and lower false alarm rates. However, if such information is not available, detection is carried out based on a general subspace. If a large set of anomaly and interference examples are available for creating of the anomaly and interference subspace a procedure known as eigen-faces is utilized. The robustness of the proposed detection approach is demonstrated on real life data from two different applications: detection of defects in wafer images and detection of faults in 3D seismic data.

It is appropriate to note that the approach presented in this chapter has some similarities with that presented by Goldman and Cohen [11, 13], however, the two methods differ in the following aspects: Goldman and Cohen suggested using a multiresolution representation of a given 2D image in order to create an appropriate feature space for natural clutter images having several periodical patterns of various period lengths. Since the GMRF model may not sufficiently describe such clutter

image, Goldman and Cohen assumed that there exists a multiresolution representation such that the GMRF model can be appropriately applied. In contrast, this work deals with true 3D data, such as seismic data or registered data taken from a number of sensors and organized in a 3D data structure. In creating a subspace for the anomalies, Goldman and Cohen overlooked the limitation of anomaly and interference subspaces ranks. If a large training set of anomaly (or interference) examples is made available, and is to be used to improve creation of the anomaly (or interference) subspace, Goldman and Cohen did not suggest any means to deal with it. If all anomaly examples are to be used in the anomaly subspace creation method they suggested, it would result in anomaly subspace, which spans the entire feature space, thus resulting in a high false alarm rate. We utilize a procedure often used in the field of face recognition, namely eigen-pictures, to overcome this limitation. In addition, the MSD suggested by Goldman and Cohen suffers from the computational burden of inverting the covariance matrix of the innovation process. We utilized a parametric form of the inverse covariance matrix, such that no matrix inversion is required.

This chapter is organized as follows: In Section 2.2, we present the GMRF model and present an efficient estimation method of its parameters. In Section 2.3, we propose our anomaly and interference subspaces creation method, which is based on the eigen-picture procedure followed by our anomaly detection approach. In Section 2.4 we demonstrate the performance of our approach by applying it to the detection of defects in wafer images and to the detection of faults in 3D seismic data. Finally in Section 2.5 we conclude.

2.2 Gauss Markov Random Field

In this section we begin by performing a pre-processing layer de-correlation stage using the KLT. We then model each transformed data layer as a GMRF, and perform an efficient model estimation method based on least squares.

2.2.1 Clutter Modeling

A 3D data set of size $K_1 \times K_2 \times K_3$ can be regarded as K_3, 2D images stacked one on top of the other. We assume that the layers are uncorrelated. Since this assumption is generally incorrect, we use a KLT of the data in the depth direction to create the uncorrelated layers. Let Y represent a 3D data set, and let $\mathbf{Y}_{ij}$ be a column vector (of size $K_3 \times 1$) representing all data layers at spatial location (i, j) in Y. Let K denote a matrix whose columns are the eigenvectors of the covariance matrix of $\mathbf{Y}_{i,j}$. The 3D data W whose layers are independent is given by:

$$\mathbf{W}_{ij} = K^T \mathbf{Y}_{ij} \,. \tag{2.1}$$

The KLT transform allows for dimensionality reduction by creating $\mathbf{W}$ from only those layers in $K^T\mathbf{Y}_{ij}$ corresponding to the larger eigenvalues of the covariance matrix of $\mathbf{Y}$. Dimensionality reduction may be important for reducing the computational complexity, reducing memory and disk space and for using only layers containing information relevant to the problem at hand. Each independent data layer in $\mathbf{W}$ is modeled as a zero mean, homogenous, first order, GMRF. The GMRF accounts for spatial correlation between neighboring image pixels. Let w_{ij} be an image pixel at location (i, j). Let β_h and β_v be the minimum mean square error (MMSE) estimation parameters of w_{ij} from its four nearest neighbors, and let ε_{ij} be the estimation error.

A first-order GMRF model is given by:

$$\begin{aligned} w_{ij} &= \beta_h \left(w_{i(j-1)} + w_{i(j+1)} \right) + \\ &\quad + \beta_v \left(w_{(i-1)j} + w_{(i+1)j} \right) + \varepsilon_{ij} \, . \end{aligned} \tag{2.2}$$

We assume a first order GMRF with zero boundary conditions for simplicity. We refer the interested reader to the work by Moura and Balram who investigated higher order fields and different boundary conditions for the GMRF model [2, 28]. We now turn to the parametric representation of the inverse covariance matrix of the GMRF model. All pixels within an image (data layer) of size $K_1 \times K_2$ are row stacked into a column vector $\boldsymbol{w}$. The estimation error pixels ε_{ij} are arranged into a column vector $\boldsymbol{\varepsilon}$ in a similar manner. Using this vector notation we can write (2.2) as:

$$A\boldsymbol{w} = \boldsymbol{\varepsilon} \, . \tag{2.3}$$

Schweizer and Moura [39] showed that the matrix A is structured and can be written in Kronecker notation as:

$$A = I_{K_1} \otimes B + H_{K_1} \otimes C \tag{2.4}$$

where

$$\begin{aligned} B &= -\beta_h H_{K_2} + I_{K_2} \\ C &= -\beta_v I_{K_2} \, . \end{aligned} \tag{2.5}$$

I_{K_1}, I_{K_2} are identity matrices of size K_1, K_2 respectively. H_{K_1}, H_{K_2} have ones on the first upper and lower diagonals and zeros everywhere else. They also proved that the error vector $\boldsymbol{\varepsilon}$ is a colored Gaussian random vector with covariance matrix $\Sigma_{\boldsymbol{\varepsilon}} = \sigma^2 A$ (where σ^2 is a positive constant). Using (2.3) and the fact that $\boldsymbol{\varepsilon}$ is Gaussian we can

write:

$$\begin{aligned} \Sigma_{\boldsymbol{w}} &= A^{-1}\Sigma_{\boldsymbol{\varepsilon}}A^{-1} \\ \Sigma_{\boldsymbol{w}} &= \sigma^2 A^{-1}\,, \end{aligned}$$

such that the inverse covariance matrix of $\boldsymbol{w}$ is given by:

$$\Sigma_{\boldsymbol{w}}^{-1} = \frac{1}{\sigma^2}A\,. \tag{2.6}$$

The three parameters β_h, β_v and σ^2 fully represent the inverse covariance matrix of $\boldsymbol{w}$. These parameters are to be estimated for each data layer, based on the data at hand as described in the following section.

2.2.2 Model Estimation

We now address the problem of model estimation. Let us note that due to the sparse presence of anomalies and interference signal in the data, the influence of the anomalies and interference signal on the parameter estimation is insignificant and therefore can be neglected. Much work has been done on the subject of GMRF model estimation. For example, Schweizer and Moura [39] explored three methods for model estimation: Maximum likelihood, approximate maximum likelihood and lease squares (minimizing the mean-square modeling error). The three methods differ in their computational complexity and in the fact that the maximum likelihood based approaches are parametric and thus take advantage of the Gaussian PDF of the clutter background. It is shown there that the least squares approach is computationally and statistically efficient. Schweizer and Moura performed least squares estimation based on averaging of mutually independent Markov windows within the image. Here we present an MMSE estimator (lease squares) applied to the entire image as a whole.

In addition Schweizer and Moura selected to solve the least squares problem using Kronecker products, we use a simpler notation by rearranging terms as described below. In a first order GMRF model every pixel w_{ij} has four neighboring pixels. Arranging these pixels in a row vector:

$$\boldsymbol{\eta}_{ij} = \begin{bmatrix} w_{i(j-1)} + w_{i(j+1)} & w_{(i-1)j} + w_{(i+1)j} \end{bmatrix} \tag{2.7}$$

and lexicographically setting these vectors as rows in a matrix: $X = \begin{bmatrix} \boldsymbol{\eta}_{11}^T & \boldsymbol{\eta}_{12}^T & \cdots & \boldsymbol{\eta}_{K_1K_2}^T \end{bmatrix}^T$ yields the following representation of (2.2):

$$\boldsymbol{w} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.8}$$

where $\boldsymbol{\theta} = \begin{bmatrix} \beta_h & \beta_v \end{bmatrix}^T$ is a vector of the unknown GMRF parameters. Our goal is to find $\boldsymbol{\theta}$ which minimizes the mean square error: $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$. The MMSE estimation of $\boldsymbol{\theta}$ is given by:

$$\hat{\boldsymbol{\theta}} = \left(X^T X\right)^{-1} X^T \boldsymbol{w}\,. \tag{2.9}$$

The third GMRF model parameter σ^2 can be estimated using the two correlation coefficients, by:

$$\hat{\sigma}^2 = \frac{1}{K_1K_2}\boldsymbol{w}^T A\boldsymbol{w} = \frac{1}{K_1K_2}\left(S_w - 2\beta_h\chi_h - 2\beta_v\chi_v\right) \tag{2.10}$$

where

$$\begin{aligned} S_w &= \sum_{i=1}^{K_1}\sum_{j=1}^{K_2}\left(w_{ij}\right)^2 \\ \chi_h &= \sum_{i=1}^{K_1}\sum_{j=1}^{K_2-1} w_{ij}w_{i(j+1)} \\ \chi_v &= \sum_{i=1}^{K_1-1}\sum_{j=1}^{K_2} w_{ij}w_{(i+1)j}\,. \end{aligned} \tag{2.11}$$

This result is based on the maximum likelihood estimation presented by Schweizer and Moura [39].

2.3 Anomaly Detection

In this section we present our anomaly detection method. We develop a set of MSDs operating in a colored Gaussian noise environment for the detection of anomalies in 3D data. The MSD was originally developed by Scharf and Friedlander [38] for the detection of signals in subspace interference and additive WGN. We use a separate MSD for each data layer thus allowing the incorporation of *a priory* information about the anomalies and the data acquisition process. If a large training set of anomalies or subspace interference is available there is a difficulty in selecting the most appropriate examples for creating a anomaly or interference subspace of an appropriate rank. For that purpose we suggest using a principal component analysis (PCA) based approach, namely eigen-faces or eigen-pictures. We utilize the parametric form of the inverse covariance matrix presented in Chapter 2.2 thus reducing the computation burden of inverting the covariance matrix.

2.3.1 Anomaly and Interference Subspaces

Since we deal with 3D data, the anomalies are also three dimensional with a spatial size $L_1 \times L_2$ which is much smaller than $K_1 \times K_2$ but larger than the GMRF neighborhood. The anomaly has the same depth dimension as the data (K_3 layers).

We assume the anomalies lie within a known subspace spanned by G image chips $o_g, g = 1, 2, \cdots, G$, each of size $(L_1 \times L_2 \times K_3)$. This 3D anomaly model allows us to represent a different pattern for the anomaly in every layer. Each 3D image chip is passed through the same KLT transform used for the 3D data. For each layer we create a matrix H_ℓ, whose columns span the anomaly subspace in that layer. This is done by row stacking layer ℓ of every image chip into a column vector and setting

these vectors as columns in a matrix: $H_\ell, \ell = 1, 2, \cdots, K_3$.

We model the interference subspace in a similar manner. We use T image chips $s_t, t = 1, 2, \cdots, T$ each of size $(L_1 \times L_2 \times K_3)$. These image chips are passed through the KLT transform and a matrix spanning the interference subspace of each layer $S_\ell, \ell = 1, 2, \cdots, K_3$ is created.

2.3.2 Dimensionality Reduction Using PCA

To generalize the discussion assume that the anomaly chip has spatial size $L_1 \times L_2$ as defined above and depth dimension $L_3 \leq K_3$. When $rank(H_\ell) \approx L_1^\ell L_2^\ell L_3^\ell$ or $rank(H_\ell) > L_1^\ell L_2^\ell L_3^\ell$, the anomaly subspace practically spans the entire space, and anomalies may be falsely detected everywhere within layer ℓ. In this case dimensionality reduction is performed by means of a technique commonly used in computer vision known as eigen-pictures or eigen-faces. This technique is based on PCA (also known as KLT). PCA techniques choose a dimensionality reducing linear projection that maximizes the scatter of all projected samples. For example, eigen-pictures are used by Wu *et al.* [46] for classification of chromosome images, and by Belhumeur *et al.* [3] for face recognition.

Let us consider the set of G_ℓ image chips o_g^ℓ. Consistent ordering of these anomaly chips into column vectors creates a set of G_ℓ column vectors $\boldsymbol{o}_g^\ell$ of size $L_1^\ell L_2^\ell L_3^\ell \times 1$. We next consider a linear transformation mapping of the original $L_1^\ell L_2^\ell L_3^\ell$ dimensional space into an M^ℓ dimensional feature space, where $M^\ell < L_1^\ell L_2^\ell L_3^\ell$. The new feature vectors $\boldsymbol{o}_g^{'\ell}$ are defined by the following linear transformation:

$$\boldsymbol{o}_g^{'\ell} = U^T \boldsymbol{o}_g^\ell \quad g = 1, 2, \ldots, G_\ell \,, \tag{2.12}$$

where U is an $L_1^\ell L_2^\ell L_3^\ell \times M^\ell$ matrix with orthonormal columns.

If the total scatter matrix $S_{\boldsymbol{O}}$ is defined as:

$$S_{\boldsymbol{O}} = \sum_{g=1}^{G_\ell} (\boldsymbol{o}_g^\ell - \boldsymbol{\mu}_{\boldsymbol{O}})(\boldsymbol{o}_g^\ell - \boldsymbol{\mu}_{\boldsymbol{O}})^T , \tag{2.13}$$

where $\boldsymbol{\mu}_{\boldsymbol{O}} = \frac{1}{G_\ell} \sum_{g=1}^{G_\ell} \boldsymbol{o}_g^\ell$ is the mean image of all image chips, then after applying the linear transformation U^T, the scatter of the transformed feature vectors $\{\boldsymbol{o}_g^{'\ell}\}_{g=1}^{G_\ell}$ is $U^T S_{\boldsymbol{O}} U$. In PCA, the projection U is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$\begin{aligned} U &= \arg\max_U |U^T S_{\boldsymbol{O}} U| \\ &= [\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_{M^\ell}] \end{aligned} \tag{2.14}$$

where $\{\mathbf{u}_i | i = 1, 2, \ldots M^\ell\}$ is the set of $L_1^\ell L_2^\ell L_3^\ell$ dimensional eigen-vectors of $S_{\boldsymbol{O}}$ corresponding to the M^ℓ largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as eigen-pictures.

2.3.3 Three Dimensional MSD

Scharf and Friedlander [38] developed an MSD for the detection of subspace signals in subspace interference and white Gaussian noise. Here, we introduce a set of MSDs for the detection of signals in subspace interference and colored Gaussian noise.

Let y_ℓ represent layer ℓ, and let $y_\ell(s)$ represent a pixel at spatial location s in y_ℓ. For each pixel $y_\ell(s)$ we create a column vector $\boldsymbol{n}_\ell(s)$ by row stacking an image chip of size $L_1 \times L_2$ centered around s.

Let $\boldsymbol{v}_\ell(s)$ be a GMRF vector of size $L_1 L_2 \times 1$, and let $\boldsymbol{\phi}_\ell(s)$, $\boldsymbol{\psi}_\ell(s)$ be the weight

vectors for the interference and anomaly subspaces respectively. We define two hypothesis:

$$\begin{aligned} H_0 : \boldsymbol{n}_\ell(s) &= S_\ell \boldsymbol{\phi}_\ell(s) + \boldsymbol{v}_\ell(s) \\ H_1 : \boldsymbol{n}_\ell(s) &= H_\ell \boldsymbol{\psi}_\ell(s) + S_\ell \boldsymbol{\phi}_\ell(s) + \boldsymbol{v}_\ell(s)\,. \end{aligned}$$

Under the two hypothesis $\boldsymbol{n}_\ell$ is distributed as:

$$\begin{aligned} H_0 : \boldsymbol{n}_\ell(s) &\sim N(S_\ell \boldsymbol{\phi}_\ell(s), \Sigma_{\boldsymbol{v}_\ell}) \\ H_1 : \boldsymbol{n}_\ell(s) &\sim N(H_\ell \boldsymbol{\psi}_\ell(s) + S_\ell \boldsymbol{\phi}_\ell(s), \Sigma_{\boldsymbol{v}_\ell})\,. \end{aligned}$$

GMRF parameter estimation (as described in Section 2.2.2) is performed for each layer. The inverse covariance matrix of $\boldsymbol{v}_\ell$ is calculated using the estimated GMRF parameters. The vectors $\boldsymbol{\psi}_\ell(s)$ and $\boldsymbol{\phi}_\ell(s)$ are estimated from the data based on maximum likelihood (ML) and are given by (see appendix A):

$$H_0 \;:\; \boldsymbol{\phi}_\ell(s) = P_0 \Sigma_{\boldsymbol{v}_\ell}^{-1/2} \boldsymbol{n}_\ell(s) \tag{2.15}$$

$$H_1 \;:\; [\boldsymbol{\phi}_\ell(s)^T, \boldsymbol{\psi}_\ell(s)^T]^T = P_1 \Sigma_{\boldsymbol{v}_\ell}^{-1/2} \boldsymbol{n}_\ell(s) \tag{2.16}$$

where

$$P_0 = (S_\ell^T \Sigma_{\boldsymbol{v}_\ell}^{-1} S_\ell)^{-1} S_\ell^T \Sigma_{\boldsymbol{v}_\ell}^{-1/2} \tag{2.17}$$

$$P_1 = ([S_\ell, H_\ell]^T \Sigma_{\boldsymbol{v}_\ell}^{-1} [S_\ell, H_\ell])^{-1} [S_\ell, H_\ell]^T \Sigma_{\boldsymbol{v}_\ell}^{-1/2}\,. \tag{2.18}$$

Following the above notation we have:

$$H_0 : \hat{v}_\ell^{(0)}(s) = (I - S_\ell P_0 \Sigma_{\boldsymbol{v}_\ell}^{-1/2}) \boldsymbol{n}_\ell(s) \tag{2.19}$$

$$H_1 : \hat{v}_\ell^{(1)}(s) = (I - [S_\ell, H_\ell] P_1 \Sigma_{\boldsymbol{v}_\ell}^{-1/2}) \boldsymbol{n}_\ell(s)\,. \tag{2.20}$$

Since $\boldsymbol{v}_\ell$ is a Gaussian vector with zero mean and a known structured inverse covariance matrix $\Sigma^{-1}_{\boldsymbol{v}_\ell}$, the log generalized likelihood ratio (GLR) (log GLR) is:

$$\begin{aligned}
\mathcal{L}_\ell(s) &= 2\ln\left[\frac{\Pr\left(\hat{\boldsymbol{v}}_\ell\left(s\right)|H_0\right)}{\Pr\left(\hat{\boldsymbol{v}}_\ell\left(s\right)|H_1\right)}\right] = \\
&= \hat{v}^{(0)}_\ell(s)^T\Sigma^{-1}_{\boldsymbol{v}_\ell}\hat{v}^{(0)}_\ell(s) - \hat{v}^{(1)}_\ell(s)^T\Sigma^{-1}_{\boldsymbol{v}_\ell}\hat{v}^{(1)}_\ell(s) = \\
&= (\Sigma^{-1/2}_{\boldsymbol{v}_\ell}\boldsymbol{n}_\ell(s))^T(\Sigma^{-1/2}_{\boldsymbol{v}_\ell}[S_\ell, H_\ell]P_1 - \\
&\quad -\Sigma^{-1/2}_{\boldsymbol{v}_\ell}S_\ell P_0)(\Sigma^{-1/2}_{\boldsymbol{v}_\ell}\boldsymbol{n}_\ell(s)) = \\
&= (\Sigma^{-1/2}_{\boldsymbol{v}_\ell}\boldsymbol{n}_\ell(s))^T(B_1 - B_0)(\Sigma^{-1/2}_{\boldsymbol{v}_\ell}\boldsymbol{n}_\ell(s)) \qquad (2.21)
\end{aligned}$$

where

$$\begin{aligned}
B_0 &= \Sigma^{-1/2}_{\boldsymbol{v}_\ell}S_\ell P_0 \\
B_1 &= \Sigma^{-1/2}_{\boldsymbol{v}_\ell}[S_\ell, H_\ell]P_1 \qquad (2.22)
\end{aligned}$$

are the projections into the subspaces spanned by the columns of $\Sigma^{-1/2}_{\boldsymbol{v}_\ell}S_\ell$ and $\Sigma^{-1/2}_{\boldsymbol{v}_\ell}[S_\ell, H_\ell]$ respectively. Due to the fact that the layers are uncorrelated, the log GLR based on K_3 image layers is:

$$\mathcal{L}(s) = \sum_{\ell=1}^{K_3}\mathcal{L}_\ell(s) \qquad (2.23)$$

$\mathcal{L}(s)$ is a sum of squared independent normally distributed variables. Muirhead [29] shows that under these conditions $\mathcal{L}(s)$ is chi-square distributed with $q = K_3 \cdot rank(H)$ degrees of freedom:

$$\begin{aligned}
H_0 : \mathcal{L}(s) &\sim \chi^2_q(0) \\
H_1 : \mathcal{L}(s) &\sim \chi^2_q(\lambda)
\end{aligned}$$

where:

$$\begin{aligned}\lambda &= \sum_{l=1}^{K_3} (\Sigma_{\boldsymbol{v}_\ell}^{-1/2} H_\ell \boldsymbol{\psi}_\ell)^T (B_1 - B_0)(\Sigma_{\boldsymbol{v}_\ell}^{-1/2} H_\ell \boldsymbol{\psi}_\ell) = \\ &= \sum_{l=1}^{K_3} (\Sigma_{\boldsymbol{v}_\ell}^{-1/2} H_\ell \boldsymbol{\psi}_\ell)^T (I - B_0)(\Sigma_{\boldsymbol{v}_\ell}^{-1/2} H_\ell \boldsymbol{\psi}_\ell)\,. \end{aligned} \tag{2.24}$$

λ is the ratio between the energy of the signal which does not lie in the interference subspace and the variance of the noise, and can therefore be regarded as a measure of SNR. If no interference subspace is assumed, such that $S = 0$, then $P_0 = 0$ and $B_0 = 0$. In this case $\lambda = \sum_{l=1}^{K_3} (H_\ell \boldsymbol{\psi}_\ell) \Sigma_{\boldsymbol{v}_\ell}^{-1} (H_\ell \boldsymbol{\psi}_\ell)^T$ and is the ratio between the signal energy and the noise variance.

The decision at every spatial location is performed by thresholding $\mathcal{L}(s)$. The threshold γ is chosen such that it would satisfy the desired detection and false alarm probabilities:

$$\begin{aligned} P_{FA} &= 1 - P[\chi_q^2(0) \le \gamma] \\ P_D &= 1 - P[\chi_q^2(\lambda) \le \gamma]\,. \end{aligned}$$

2.4 Experimental Results

In this section we demonstrate the performance of our anomaly detection approach by applying it to the detection of of defects in wafer images, and to the detection of discontinuities in 3D seismic data.

2.4.1 Defect Detection in Silicon Wafers

The wafer 3D data is composed of 3 layers. Each layer is taken from a different perspective (right angle, left angle, and top). The three layers are perfectly aligned.

The 3D data is of size $128 \times 128 \times 3$. Figures 2.1(a)-(c) show the three layers of the wafer image. The defect size is approximately 3×3 pixels. We use $L_1 = L_2 = 3$. The anomaly subspace is constructed from 3 image chips. We did not use real defects for the anomaly image chips, but rather simple bar shape structures in each layer. The results of the proposed anomaly detection algorithm (without thresholding) are presented in Figure 2.1(d). Correct detection is apparent from the image (the small target like symbol marks the detection location. To increase contrast in the detection image, we performed gamma correction on the displayed detection image. We chose to compare our results with those of Goldman and Cohen [13] and with those of the RX algorithm presented in Chapter 1.1.3. In order to implement the Goldman-Cohen algorithm we perform detection in each of the three image layers separately. A multiresolution representation of each image layer is achieved by means of an undecimated wavelet transform. The image chips used for detection are identical to those used in our algorithm for anomaly subspace creation. The results for each of the three original image layers of Figures 2.1(a)-(c) are presented in Figures 2.2(a)-(c) respectively. The gamma correction applied to image 2.1(d) is applied here as well to increase contrast. Detection is only achieved for Figure 2.2(a) and 2.2(c). In Figure 2.2(a) there are also several false alarms. In Figure 2.2(b) only false alarms are detected. Since it is not known *a priory*, which of the image layers would produce accurate results, using the Goldman-Cohen algorithm creates a problem of choosing the image to use. Heuristically trying to add the three detection images results in the detection image presented in Figure 2.2(d). Although positive detection is achieved a false alarm also occurs. This demonstrates the advantage of using all available input data (in this case all three image layers) compared to using a single image layer. We next compare our results with those of the RX algorithm presented in Chapter 1.1.3.

Unlike the Goldman-Cohen algorithm the RX algorithm was designed to deal with hyperspectral data, and thus is expected to produce adequate results for the multi-layered wafer image presented here. The RX algorithm is implemented here without a training set, such that $N_{ref} = 0$ and the clutter mean and covariance matrix are estimated from the image itself, under the assumption that the contribution of the anomalies to the calculation is negligible. Detection results for the RX algorithm are presented in Figure 2.3. Although positive detection is achieved, a false alarm also occurs. The potential advantage of the proposed method is apparent.

2.4.2 Fault Detection in Seismic Data

We use real life seismic data (from the Gulf of Mexico). The data is a 3D lattice of size $201 \times 201 \times 226$. We pre-process the data in a similar manner to that suggested by Cohen and Coifman [8]: A small analysis cube of size $4 \times 4 \times 15$ moves through every spatial location in every data layer. At each location, the analysis cube is partitioned into 4 sub-cubes each of size $2 \times 2 \times 15$. The sub-cubes are arranged in a consistent fashion into 4 column vectors. A correlation matrix of the 4 vectors is calculated. The lower triangle coefficients of the correlation matrix are formed into a column vector of size 6×1. In this manner each data layer of size $201 \times 201 \times 1$ is turned into a $201 \times 201 \times 6$ lattice. Detection is performed independently on each such lattice. We first perform the KLT transform described in Section 2.2.1. The 6 layers are highly correlated, therefor after the KLT transform has been applied, we only use the layer corresponding to the largest eigenvector, and neglect all other layers ($K_3 = 1$). We use the detection method described in Section 2.3 on the single layer data. The signal subspace is constructed from a single image chip of size 1×5, which describes a bar shape. Figure 2.4 shows a horizontal slice of the original data, and the results of the

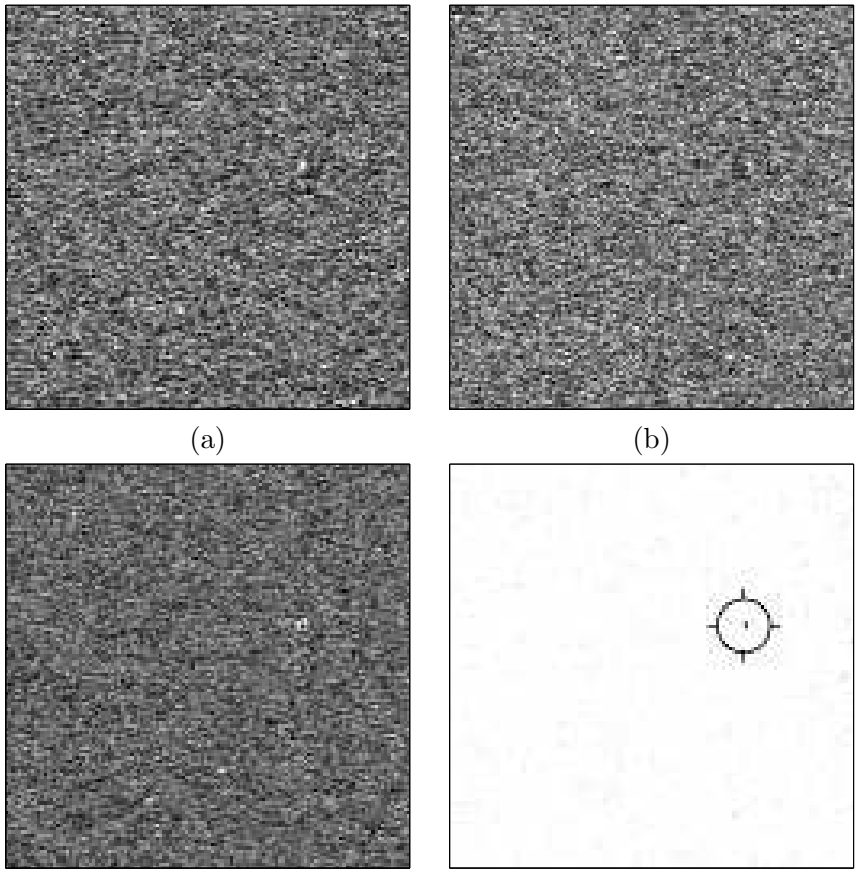


Figure 2.1: Detection results in silicon wafers: (a)-(c) The three layers of the original image; (d) Anomaly detection.

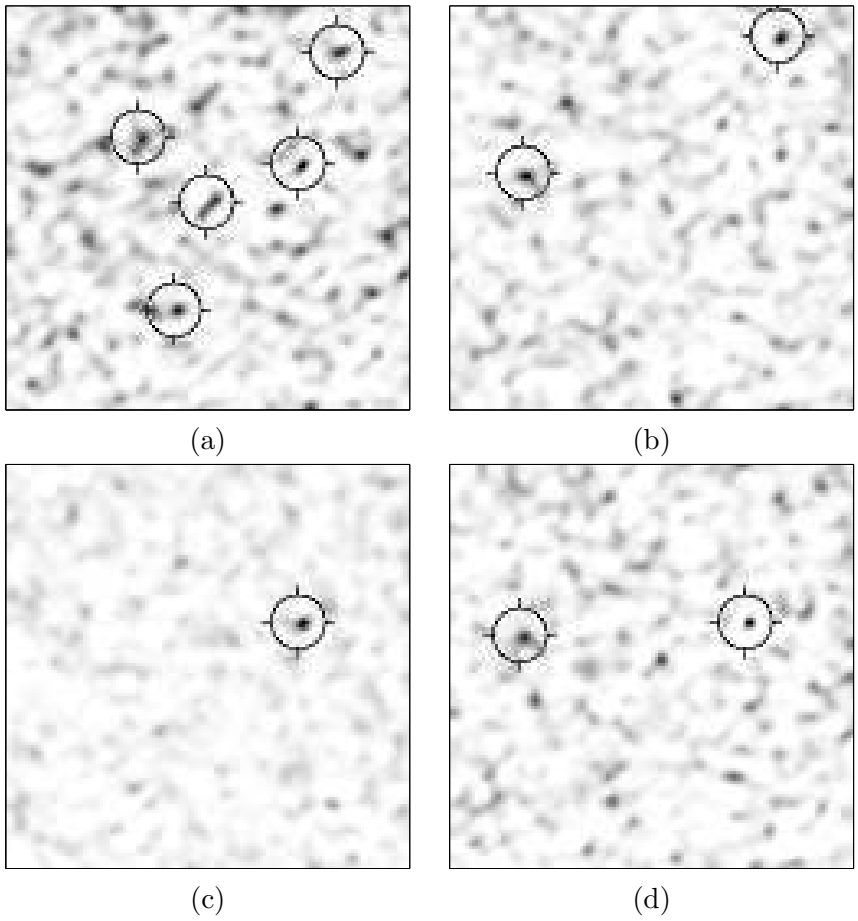


Figure 2.2: Defect detection in silicon wafers using the algorithm suggested by Goldman and Cohen: (a)-(c) detection results for layers 1-3 respectively; (d) Heuristic sum of detection results in all layers.

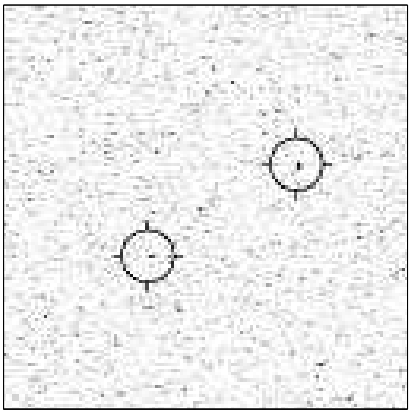
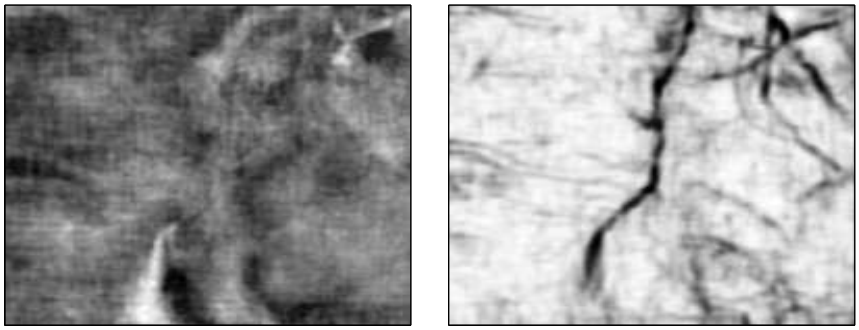


Figure 2.3: Defect detection in silicon wafers using the RX algorithm.



(a) (b)

Figure 2.4: Detection of faults in seismic data: (a) Horizontal slice of the 3D seismic data; (b) Anomaly detection.

anomaly detection algorithm (without thresholding).

2.5 Conclusion

In this chapter, we have presented an anomaly detection approach for three dimensional data. We presented a detector constructed from a set of MSDs working in a colored noise environment modeled as a GMRF. Each MSD is utilized to perform

anomaly detection in a single image layer. Using MSDs allows us to incorporate into the signal subspace a-priory information about the sensors used to capture the data and about the anomalies. We used a different MSD for each data layer, thus allowing for maximal use of *a priory* information. Incorporating such information potentially produces improved detection results. We demonstrated the proposed approach on two applications: detection of defects in wafer images, and detection of faults in real life seismic data. We compared our results with the RX algorithm designed for dealing with layered data and with the Goldman-Cohen algorithm designed for single layer data. Although our results show the potential of the proposed approach, using the GMRF may be inappropriate for some images, specifically, images characterized by a heavy tailed distribution cannot be statistically well modeled by a Gaussian distribution. In addition, if the images are characterized by clustering of innovations, that is, areas (sets of clustered pixels) in which there are large changes, and areas in which there are only minor changes may also be poorly represented by the GMRF model. This is true for both single and multi-layered data. In the following chapter we present a statistical model for multidimensional data, which is capable of capturing the above mentioned phenomena.

Chapter 3

GARCH Random Field Modeling

3.1 Introduction

Two common phenomena of often used feature spaces are 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). In particular, the wavelet transform, which is often used as a feature space in applications dealing with natural images, yields wavelet coefficients that show excess kurtosis. Spatial and scale-to-scale statistical dependencies of wavelet coefficients also exist. That is, coefficients of large magnitudes tend to appear at close spatial locations and at adjacent scales and orientations. Srivastava [41] explains that the strongest evidence of non-Gaussianity of image comes from the observed histograms: heavier tails than Gaussian (implying larger kurtosis), sharp cusps at the center and higher correlation at different scales. Willsky [44] discusses non-Gaussian models to capture the "heavy tails" nature of a wavelet representation of imagery. Mallat [23] discusses

the histogram of the detail image in the wavelet domain and presents an experimentally obtained histogram model, which is not Gaussian. Buccigrossi and Simoncelly [6] explain the non-Gaussian statistics of wavelet subband coefficients in that images typically have spatial structure consisting of smooth areas interspersed with occasional edges or other abrupt transitions. The smooth regions lead to non-zero coefficients, and the structures give occasional large-amplitude coefficients. Commonly used feature space statistical models, such as the GMRF presented in the previous chapter, may not sufficiently describe the feature space. Detection algorithms based on these models may result in high false alarm rates due to the inadequacy between the model and the data.

Bollerslev [5] introduced the 1D GARCH ,which is often used as a statistical model for time series. It is an extension to the ARCH model introduce by Engle [9]. The GARCH model allows for the conditional variance to change as a function of past squared field values and past conditional variance values. This behavior creates a heavy tailed distribution characterized by clustering of innovations. The 1D GARCH has been shown to be useful in modeling different economic phenomena. In this chapter we introduce an extension of the GARCH model to the general case of ND. This may be a more suitable statistical model for multidimensional data characterized by the two phenomenons mentioned above. For a given 2D image we propose a 3D multiresolution feature space. This feature space will be used in Chapter 4 for anomaly detection. We assume that our feature space follows a 3D GARCH model, that is, the conditional variance at every location within the feature space depends on squared field values and conditional variance values of neighboring locations, where the neighborhood is 3D.

This chapter is organized as follows: In Section 3.2 we define the ND GARCH

model. In Section 3.3 maximum likelihood GARCH model estimation is presented. In order to guarantee bounded variance for an infinite GARCH lattice, conditions for wide sense stationarity (WSS) of the GARCH model are presented. In Section 3.4 the proposed wavelet based feature space is presented and its statistical model is discussed.

3.2 N Dimensional GARCH Model Definition

Let $\mathbf{q} = (q_1, q_2, \ldots, q_N),\ q_i \geq 0,\ i = 1, \ldots, N;\ \mathbf{p} = (p_1, p_2, \ldots, p_N),\ p_i \geq 0,\ i = 1, \ldots, N$ denote the order of an ND GARCH model, and let Γ_1 and Γ_2 denote two neighborhood sets, such that:

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

Define an ND index vector $\mathbf{i} = (i_1, i_2, \ldots, i_N)$. Let $\epsilon_{\mathbf{i}}$ represent a random variable on an ND 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 ND causal neighborhood of location $\mathbf{i}$ as: $\Gamma = \Gamma(\mathbf{i}) = \{\mathbf{k} \mid k_j \leq i_j,\, j = 1, \ldots, N\}$ and let $\eta_{\mathbf{i}} \overset{iid}{\sim} N(0,1)$ be another random variable on an ND lattice independent of $\{h_{\mathbf{k}}\}_{\mathbf{k}\in\Gamma}$. An ND $GARCH(\mathbf{p};\mathbf{q})$ process is defined as:

$$\epsilon_{\mathbf{i}} = \sqrt{h_{\mathbf{i}}}\,\eta_{\mathbf{i}} \tag{3.1}$$

$$h_{\mathbf{i}} = \alpha_0 + \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} \epsilon^2_{\mathbf{i}-\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{k}} \tag{3.2}$$

and is therefore conditionally distributed as:

$$\epsilon_{\mathbf{i}} \mid \psi_{\mathbf{i}} \sim N(0, h_{\mathbf{i}})\,. \tag{3.3}$$

In order to guarantee a non-negative conditional variance the model parameters must satisfy:

$$\begin{aligned} \alpha_0 &> 0 \\ \alpha_{\mathbf{k}} &\geq 0, \quad \mathbf{k} \in \Gamma_1 \\ \beta_{\mathbf{k}} &\geq 0, \quad \mathbf{k} \in \Gamma_2 \,. \end{aligned} \tag{3.4}$$

From (3.2) we see that at every location $\mathbf{i}$, both the ND neighboring squared field values and the ND 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 $\epsilon_{\mathbf{i}}$ 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 1D GARCH model introduce by Bollerslev [5].

Note that although causality may seem an unnatural model limitation, it is a means of guaranteeing non negativity of the conditional variance in the above model (see Appendix B for more details). The causality of the model may lead to different results depending on the image orientation. This is demonstrated in Section 4.4.2. Depending on the application and on the data at hand, it may be appropriate to consider more than one image orientation when performing anomaly detection based on the casual GARCH model.

3.3 Estimation of an N-D GARCH Model

In this section we find a maximum likelihood estimate for the GARCH model. We let $\epsilon_{\mathbf{i}}$ be innovations of a linear regression on an ND lattice, where $y_{\mathbf{i}}$ is the dependent variable, $\boldsymbol{x}_{\mathbf{i}}$ a vector of explanatory variables and $\boldsymbol{b}$ a vector of unknown parameters, such that:

$$\epsilon_{\mathbf{i}} = y_{\mathbf{i}} - u_{\mathbf{i}}\,, \tag{3.5}$$

where $u_{\mathbf{i}} = \boldsymbol{x}_{\mathbf{i}}^T \boldsymbol{b}$.

Note that if $\epsilon_{\mathbf{i}}$ in (3.5) is WGN (as described in Section 3.2) the regression model is a casual GMRF. This is a special case of the GARCH process.

Using (3.5) we can write (3.2) as:

$$\begin{aligned} h_{\mathbf{i}} &= \alpha_0 + \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} (y_{\mathbf{i}-\mathbf{k}} - u_{\mathbf{i}-\mathbf{k}})^2 \\ &\quad + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{k}}\,. \end{aligned} \tag{3.6}$$

The conditional distribution of $y_{\mathbf{i}}$ is Gaussian with mean $u_{\mathbf{i}}$ and variance $h_{\mathbf{i}}$,

$$f(y_{\mathbf{i}} \mid u_{\mathbf{i}}, \psi_{\mathbf{i}}) = \frac{1}{\sqrt{2\pi h_{\mathbf{i}}}} \exp\left(-\frac{(y_{\mathbf{i}} - u_{\mathbf{i}})^2}{2h_{\mathbf{i}}}\right)\,. \tag{3.7}$$

Let $\boldsymbol{\delta}$ denote a column vector of the GARCH model parameters,

$$\boldsymbol{\delta} = [\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}]^T$$

and define four neighborhood column vectors associated with location $\mathbf{i}$: $\boldsymbol{\epsilon}_{\mathbf{i}}$, $\boldsymbol{h}_{\mathbf{i}}$, $\boldsymbol{y}_{\mathbf{i}}$, $\mathbf{u}_{\mathbf{i}}$. These vectors represent consistent ordering into a column vector of the random fields elements: $\{\epsilon_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k}\in\Gamma_1}$, $\{h_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k}\in\Gamma_2}$, $\{y_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k}\in\Gamma_1}$, $\{u_{\mathbf{i}-\mathbf{k}}\}_{\mathbf{k}\in\Gamma_1}$ respectively.

We further define

$$\begin{aligned} \boldsymbol{z}_{\mathbf{i}} &= [1, (\boldsymbol{\epsilon}_{\mathbf{i}}^T)^{\cdot 2}, \boldsymbol{h}_{\mathbf{i}}^T]^T = \\ &= [1, (\boldsymbol{y}_{\mathbf{i}}^T - \mathbf{u}_{\mathbf{i}}^T)^{\cdot 2}, \boldsymbol{h}_{\mathbf{i}}^T]^T\,, \end{aligned}$$

where $()^{\cdot 2}$ represents squaring element by element. Using the above notation the conditional variance in (3.6) can be written as:

$$h_{\mathbf{i}} \;=\; [\boldsymbol{z}_{\mathbf{i}}]^T \boldsymbol{\delta}\,. \tag{3.8}$$

Define the sample space Ω_s as an ND lattice of size $K_1 \times K_2 \times \cdots \times K_N$ such that: $\Omega_s = \{\mathbf{i} \mid 1 \le i_j \le K_j,\, j = 1, \ldots, N\}$ and let $\boldsymbol{\theta} = [\boldsymbol{b}^T, \boldsymbol{\delta}^T]^T$ be a vector of the unknown parameters. The conditional sample log likelihood is:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \sum_{\mathbf{i}\in\Omega_s} \log f(y_{\mathbf{i}} \mid \boldsymbol{x}_{\mathbf{i}}, \psi_{\mathbf{i}}) = \\
&= -\frac{1}{2}\left[(K_1 + \cdots + K_N) log(2\pi) - \sum_{\mathbf{i}\in\Omega_s} \log([\boldsymbol{z}_{\mathbf{i}}]^T \boldsymbol{\delta}) + \right. \\
&\quad \left. - \sum_{\mathbf{i}\in\Omega_s} (y_{\mathbf{i}} - u_{\mathbf{i}})^2 / ([\boldsymbol{z}_{\mathbf{i}}]^T \boldsymbol{\delta})\right]\,.
\end{aligned} \tag{3.9}$$

Equation (3.9) together with the constraints in (3.4) may seem enough to estimate model parameters. However, due to the structure of the conditional variance (3.2), 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. Bollerslev [5] proves that a sufficient condition for WSS of the 1D GARCH process is that the sum of all model parameters is smaller than 1. We [32] have extended this results for the 2D GARCH process. A similar result is obtained here for the ND case as we prove in the following theorem.

Theorem 1 *The GARCH*$(\mathbf{p};\mathbf{q})$ *process as defined in (3.1) and (3.2) is wide-sense stationary with:*

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

if and only if

$$\mathbf{1}^T(\boldsymbol{\alpha}+\boldsymbol{\beta}) \leq 1\,,$$

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.*

Proof: See Appendix C.

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 the work by Berndt *et al.* [4]). The constraints used are those presented in (3.4) and in Theorem 1. To solve the maximization problem, knowledge of $\epsilon_{\mathbf{i}}$ and $h_{\mathbf{i}}$ where $i_1,\ldots,i_N \leq 0$ is required. We set these boundaries in a similar way to that used by Bollerslev for the 1D case [5] and by us for the 2D case [32] such that:

$$\epsilon_{\mathbf{i}} = h_{\mathbf{i}} = \frac{1}{\prod_{\ell=1}^{N} K_\ell} \sum_{\mathbf{k}\in\Omega_s} (y_{\mathbf{i}} - u_{\mathbf{i}})^2;\ \forall\, i_1,\ldots,i_N \leq 0\,. \tag{3.10}$$

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 before going into the effort of estimating it. Several tests have been proposed for the 1D GARCH model (see for example Bollerslev [5] and Engle [9]). The problem of testing for GARCH and model order estimation is beyond the scope of the current work and may be the subject of future research.

3.4 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 2D image of size $K_1 \times K_2$. We use an undecimated wavelet transform into z levels to create a multiresolution representation of Y. The undecimated wavelet transform yields 4 subband images at every analysis level. These 4 subband images are labeled d^i_{LH}, d^i_{HL}, d^i_{HH} and s^i_{LL}, where the subscripts L, 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. Strickland [43] used the subband images of an undecimated wavelet transform to create a 3D 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):

$$\boldsymbol{\epsilon}_{i_1,i_2} = [d^1_{LH} + d^1_{HL}, d^1_{HH}, d^2_{LH} + d^2_{HL}, d^2_{HH}, \ldots, d^z_{LH} + d^z_{HL}, d^z_{HH}, s^z_{LL}]^T_{(i_1,i_2)} \,. \quad (3.11)$$

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 3D 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):

$$\boldsymbol{\epsilon}_{i_1,i_2} = [d^1_{LH}, d^1_{HL}, d^1_{HH}, d^2_{LH}, d^2_{HL}, d^2_{HH}, \ldots, d^z_{LH}, d^z_{HL}, d^z_{HH}, s^z_{LL}]^T_{(i_1,i_2)} \,. \quad (3.12)$$

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 (3.11) or (3.12) respectively.

As discussed in the introduction, 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. It is also argued that spatial and scale-to-scale statistical dependencies of wavelet coefficients exist. That is, coefficients of large magnitudes tend to appear at close spatial locations and at adjacent scales and orientations. We assume that there is a set of wavelet filters such that $\mathbf{Y}$ follows a 3D GARCH model.

3.5 Conclusion

The GARCH statistical model is a heavy tailed distribution characterized by clustering of innovations. It is of interest since a heavy tailed distribution and clustering of innovations are common characteristics of image multiresolution representations, and thus cannot be well modeled by often used Gaussian based statistical models such as the GMRF. In this chapter we have extended the 1D GARCH model into the multidimensional case, and proposed a maximum likelihood model estimation scheme. Conditions for WSS of the model were presented to ensure bounded conditional variance for unlimited feature space dimensions. We presented an undecimated wavelet multiresolution representation and assumed clutter modeling can be based on a 3D GARCH model, thus allowing for correlation between pixels of different spatial and depth locations within the 3D data cube. An appropriate subspace anomaly detector is developed for the proposed feature space and statistical model in the following chapter.

Chapter 4

Multiscale Anomaly Detection

4.1 Introduction

Many anomaly detection algorithms have been developed over the years. A short review of some of the most common anomaly detection algorithms is presented in Chapter 1. Markou and Singh [26] published a comprehensive review of statistical approaches to detection. There is a strong connection between the selection of feature space, clutter statistical model, anomaly model and the detection algorithm. In this chapter we develop an anomaly detector, which is based on modeling the undecimated wavelet transform image feature space (presented in Section 3.4) as a 3D causal autoregressive model with GARCH innovations. Scharf and Frienlander [38] developed an MSD for the detection of signals in subspace interference and additive WGN. Here we develop a set of multiscale MSDs operating in subspace interference and additive GARCH noise. For every scale and orientation in the wavelet domain (referred to as feature space layers) a separate anomaly subspace is used, thus allowing for better incorporation of *a priory* information into the process. Our multiscale MSD

approach takes into 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 is application dependent and independent of the detection algorithm. The potential advantages of the proposed approach are demonstrated by synthetic and real data examples.

As an input point into the development of the proposed detector let us note the following: we assume that the anomalies and interference signals are sparse within the image and therefore their influence on the model estimation and on the estimated conditional variance field is negligible. GARCH model estimation is performed as described in Section 3.3. The conditional variance field h_{i_1,i_2,i_3} is calculated based on the estimated model parameters using (3.2) and (3.10) and is later used in our detection process. We next turn to present our multiscale MSD anomaly detection approach.

4.2 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 improves incorporation of *a priory* information, thus potentially reducing the false alarm rate of the detection algorithm. The anomaly subspace for layer ℓ is spanned by a training set of G_ℓ anomaly chips. The anomaly chips are denoted: $o_g^\ell,\ g = 1, 2, \cdots, G_\ell$ 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 out around the spatial center of the anomaly and layer ℓ. 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_ℓ vectors associated with layer ℓ are arranged as columns in a matrix H_ℓ, such that the columns or H_ℓ span the anomaly subspace for layer ℓ. This procedure is performed for every layer $\ell = 1, \ldots, K_3$.

When the number of available image chips for a certain layer is high, such that $rank(H_\ell) \approx L_1^\ell L_2^\ell L_3^\ell$, the subspace practically spans the entire space, and anomalies may be falsely detected everywhere within layer ℓ. In this case dimensionality reduction using PCA is utilized (see Section 2.3.2 for more details).

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

4.3 Multiscale Matched Subspace Detection

In this section we introduce an anomaly detection approach based on an MSD and the multidimensional GARCH statistical model presented above. Scharf and Frienlander [38] developed an MSD 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 ℓ and spatial location s in the 3D lattice $\mathbf{Y}$. For each pixel $y(\ell, s)$ we create a column vector $\boldsymbol{y}(\ell, s)$ by row stacking an image chip

of size $L_1^\ell \times L_2^\ell \times L_3^\ell$ centered around (ℓ, s). Let $\boldsymbol{\epsilon}(\ell, s)$ 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 (ℓ, s). Similarly let $\mathbf{u}(\ell, s)$ be a vector representing the explanatory variable field $\left(\boldsymbol{x}_{i_1,i_2,i_3}^T \boldsymbol{b}\right)$ in the $L_1 \times L_2 \times L_3$ neighborhood of (ℓ, s). Let $\boldsymbol{\phi}(\ell, s)$, $\boldsymbol{\psi}(\ell, s)$ be vectors locating the interference and anomaly within their subspaces $< S_\ell >= span\{S_\ell\}$, $< H_\ell >= span\{H_\ell\}$ respectively. We define two hypotheses, H_0 and H_1, which represent absence and presence of an anomaly respectively:

$$\begin{aligned} H_0 : \boldsymbol{y}(\ell, s) &= S_\ell \boldsymbol{\phi}(\ell, s) + \mathbf{u}(\ell, s) + \boldsymbol{\epsilon}(\ell, s) \\ H_1 : \boldsymbol{y}(\ell, s) &= H_\ell \boldsymbol{\psi}(\ell, s) + S_\ell \boldsymbol{\phi}(\ell, s) + \mathbf{u}(\ell, s) + \boldsymbol{\epsilon}(\ell, s)\,. \end{aligned} \tag{4.1}$$

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

$$\begin{aligned} H_0 : \boldsymbol{y}(\ell, s) &\sim N(S_\ell \boldsymbol{\phi}(\ell, s) + \mathbf{u}(\ell, s), \Sigma(\ell, s)) \\ H_1 : \boldsymbol{y}(\ell, s) &\sim N(H_\ell \boldsymbol{\psi}(\ell, s) + S_\ell \boldsymbol{\phi}(\ell, s) + \mathbf{u}(\ell, s), \Sigma(\ell, s))\,. \end{aligned}$$

Note that although $\Sigma(\ell, s)$ is a diagonal matrix, the vector elements in $\boldsymbol{y}(\ell, s)$ are only conditionally uncorrelated. Unconditionally, these vector elements are correlated, such that correlation within and between layers plays a role in our detection algorithm. Define P_{S_ℓ} as the projection into the subspace spanned by the columns of S_ℓ, 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:

$$\begin{aligned} P_{S_\ell} &= S_\ell (S_\ell^T S_\ell)^{-1} S_\ell^T \\ P_{H_\ell S_\ell} &= [H_\ell S_\ell] \left([H_\ell S_\ell]^T [H_\ell S_\ell] \right)^{-1} [H_\ell S_\ell]^T\,. \end{aligned} \tag{4.2}$$

From (4.1) and (4.2) we find the GARCH innovations field under any one of the hypotheses:

$$\begin{aligned} H_0 : \boldsymbol{\epsilon}_0(\ell,s) &= \boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s) - S_\ell \boldsymbol{\phi}(\ell,s) = \\ &= (I - P_{S_\ell})[\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s)] \qquad (4.3) \\ H_1 : \boldsymbol{\epsilon}_1(\ell,s) &= \boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s) - S_\ell \boldsymbol{\phi}(\ell,s) - H_\ell \boldsymbol{\psi}(\ell,s) = \\ &= (I - P_{H_\ell S_\ell})[\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s)]\,. \qquad (4.4) \end{aligned}$$

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

$$\begin{aligned} H_0 : \ell_0(\ell,s) &= (2\pi)^{-L_1^\ell L_2^\ell L_3^\ell/2} \mid \Sigma(\ell,s) \mid^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2}\boldsymbol{\epsilon}_0(\ell,s)^T \Sigma(\ell,s)^{-1} \boldsymbol{\epsilon}_0(\ell,s)\right] \qquad (4.5) \\ H_1 : \ell_1(\ell,s) &= (2\pi)^{-L_1^\ell L_2^\ell L_3^\ell/2} \mid \Sigma(\ell,s) \mid^{-1/2} \\ &\quad \times \exp\left[-\frac{1}{2}\boldsymbol{\epsilon}_1(\ell,s)^T \Sigma(\ell,s)^{-1} \boldsymbol{\epsilon}_1(\ell,s)\right], \end{aligned}$$

where $\mid \Sigma(\ell,s) \mid$ denotes the determinant of $\Sigma(\ell,s)$.

The GLR is defined as:

$$\mathcal{L}(\ell,s) = 2\log\left(\frac{\ell_1(\ell,s)}{\ell_0(\ell,s)}\right), \qquad (4.6)$$

Substituting (4.5) into (4.6) yields:

$$\begin{aligned} \mathcal{L}(\ell,s) &= \boldsymbol{\epsilon}_0(\ell,s)^T \Sigma(\ell,s)^{-1} \boldsymbol{\epsilon}_0(\ell,s) - \boldsymbol{\epsilon}_1(\ell,s)^T \Sigma(\ell,s)^{-1} \boldsymbol{\epsilon}_1(\ell,s) = \\ &= [(P_{H_\ell S_\ell} - P_{S_\ell})(\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s))]^T \Sigma(\ell,s)^{-1} \\ &\quad \times [(P_{H_\ell S_\ell} - P_{S_\ell})(\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s))] \\ &= \left[\Sigma(\ell,s)^{-1/2}(\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s))\right]^T (P_{H_\ell S_\ell} - P_{S_\ell}) \\ &\quad \times \left[\Sigma(\ell,s)^{-1/2}(\boldsymbol{y}(\ell,s) - \mathbf{u}(\ell,s))\right]. \qquad (4.7) \end{aligned}$$

The 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 $[(H_\ell \boldsymbol{\psi}(\ell,s))(I-P_{S_\ell})]^T [(H_\ell \boldsymbol{\psi}(\ell,s))(I-P_{S_\ell})]$ and the innovations' conditional variance $\Sigma(\ell,s)$, such that:

$$SNR(\ell,s) = [(H_\ell \boldsymbol{\psi}(\ell,s))(I-P_{S_\ell})]^T \Sigma(\ell,s)^{-1} [(H_\ell \boldsymbol{\psi}(\ell,s))(I-P_{S_\ell})] \; . \tag{4.8}$$

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

$$\begin{aligned} H_0 \;\; &: \;\; \mathcal{L}(\ell,s) \sim \chi^2_{\mu_\ell}(0) \\ H_1 \;\; &: \;\; \mathcal{L}(\ell,s) \sim \chi^2_{\mu_\ell}(SNR(\ell,s)) \, . \end{aligned} \tag{4.9}$$

Under hypothesis H_1, the non-centrality parameters of the chi-square distributions of $\mathcal{L}(\ell,s)$ is equal to the SNR.

The GLR is a 3D lattice. Our goal is to unify the detection results for multiple layers into a single 2D 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 priory* 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: $\Omega \subset \{1,2,\ldots,K3\}$ such that the final detection image is:

$$D(i_1, i_2) = \sum_{k \in \Omega} \mathcal{L}(i_1, i_2, k); \; \forall i_1, i_2 \, . \tag{4.10}$$

The elements summed in (4.10) are in general statistically dependent. This is due to the 3D neighborhood used to create $\boldsymbol{y}(\ell, s)$. However, under certain conditions, these elements are conditionally statistically independent. We discuss two such private cases of the general case in (4.10). In the first case layer ℓ of the GLR is selected and used as the detection image, that is $\Omega = \{\ell\}$ such that:

$$D(i_1, i_2) = \mathcal{L}(i_1, i_2, \ell); \; \forall i_1, i_2 \, . \tag{4.11}$$

Depending on the application and available *a priori* information, different selections of ℓ 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 $\ell = \left\lfloor \frac{K_3}{2} \right\rfloor + 1$ where $\left\lfloor \frac{\cdot}{\cdot} \right\rfloor$ stands for integer division.

In the second case Ω consists of p layers, which are mutually further apart than the corresponding depth dimension of the 3D neighborhoods $L_3^\ell, \; \ell \in \Omega$. If $p = 1$ this reduces to the first case. For example, consider $K_3 = 7$ and $\{L_3^\ell = 3, \; \forall \ell\}$. Choosing $\Omega = \{2, 6\}$ would yield conditionally independent layers in (4.10).

Detection is performed by applying a threshold η to $D(i_1, i_2)$ yielding:

$$D(i_1, i_2) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \; , \tag{4.12}$$

The threshold is determined by the tradeoff between the desired conditional detection and false alarm rates. For the case where independent layers are used, these rates can be calculated by:

$$P_{FA} = 1 - P\left[\chi^2_{\sum\limits_{k \in \Omega} \mu_k}(0) \leq \eta\right] \tag{4.13}$$

$$P_D = 1 - P\left[\chi^2_{\sum_{k\in\Omega}\mu_k}\left(\sum_{k\in\Omega} SNR_k\right) \leq \eta\right] . \tag{4.14}$$

These rates cannot be easily found for the general case due to the conditional statistical dependence of the elements summed in (4.10). However, computer simulations can present receiver operating characteristics (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. Figure 4.1(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 (4.13) and (4.14) respectively. As expected the detection rate increases with the SNR. Figure 4.1 (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. Figure 4.1 (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 (4.13) and (4.14). 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 (4.10) contains dependent layers, are presented in Figure 4.1 (d). These curves are generated by computer simulations under similar assumptions to those used for generating Figure 4.1 (c) (constant SNR and identical anomaly subspace ranks for all layer). The ROC curve for the former case of 2 conditionally independent layers is also presented for comparison. To generate Figure 4.1 (d) we used $\mu_k = 3\,; \forall k \in \Omega$ such that the ROC curve representing two independent layers is based on information from 6 different layers. For the ROC curve representing 2 dependent layers we chose to use information from only 5 different layers ($\Omega = 2, 4$). This explains the apparent advantage of the 2 independent layers over the 2 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 2 (6 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) may be beneficial as seen in the ROC curves of Figure 4.1 (d). The clear advantage of using dependent layers with $p = 6$ over using independent layers with $p = 2$ is due to the fact that information from more than 6 layers is used. We also note that potentially, the anomaly subspace used for the additional dependent layers may contribute information to the detection process. Similarly, the ROC curve for the case of $p = 7$ is presented in Figure 4.1 (d) to show that the larger the number of dependent layers the better the detector's performance.

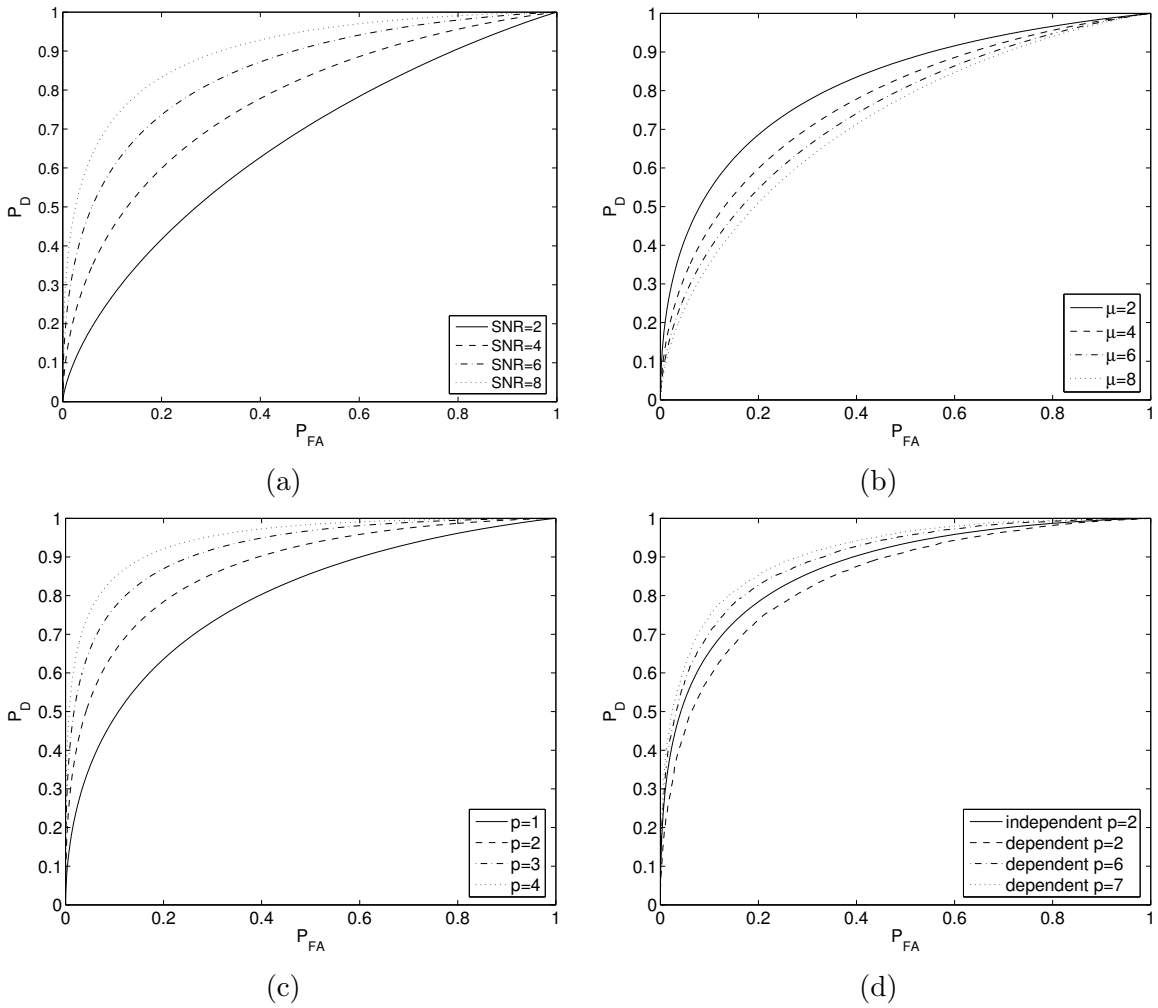


Figure 4.1: ROC curves of: (a) A single layer detector for different values of SNR; (b) A single layer detector for different anomaly subspace ranks; (c) A single layer detector and an independent layers detector; (d) A dependent layers detector vs. an independent layers detector

4.4 Experimental Results

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

4.4.1 Synthetic Data

We demonstrate the performance of the multiscale MSD on the synthetic data presented in Figure 1.8, which exactly matches our model assumptions, and qualitatively investigate the detection performance for different selections of Ω. The GARCH data in Figure 1.8 was generated using the regression and GARCH parameters shown in Tables 4.1 and 4.2, 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 parameters 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 7×7 random, Gaussian shaped anomaly, is planted to the lower left of the image center in all layers of the synthetic image as can be seen in 1.8(f) for example. Note that the anomaly does not stand out in all layers, specifically in Figures 1.8(d) and 1.8(g) it can hardly be noticed. As described above, the clutter clearly contains areas of clustered variations. These areas may generate high rate of false alarms when conventional GMRF based anomaly detection algorithms are deployed, as we have previously demonstrated for the 2D case [32]. For anomaly detection we set the anomaly size to $L_1 = L_2 = 7$, $L_3 = 3$ and create an anomaly subspace using 4 image chips. No interference subspace is assumed. We perform parameter estimation as described in Section 3.3 and anomaly detection as

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

Table 4.1: Regression Parameters Used for Generating Synthetic Image

detailed in Section 4.3. Figure 4.2 shows layers $2-6$ of the GLR. Layers 1 and 7 are not considered here since they suffer from boundary effects due to the 3D nature of the anomaly subspace. The target mark on each detection image shows the detection result when Ω contains this layer only. Note that positive detection is achieved in all layers, while layer 6 includes a false alarm. Figure 4.3 (a) shows detection results when performing detection using layers 2 and 6, that is $\Omega = \{2, 6\}$ and $L3 = 3$. These two detection layers are conditionally statistically independent and the detection image is achieved using (4.10). Once again the anomaly is clearly detected. Figure 4.3 (b) presents the results where Ω includes dependent layers $2-6$ and $L3 = 3$. The 2D detection result is achieved by means of (4.10). The detection of Figure 4.3 (b) seems clearer than that of Figure 4.3 (a), which qualitatively demonstrates the potential of using dependent layers. Figure 4.3 (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 a 7 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 $\Omega = \left\{\left\lfloor \frac{K_3}{2} \right\rfloor + 1\right\}$.

Parameter	α_0	α_{110}	α_{100}	α_{010}	α_{101}	α_{111}	α_{011}	α_{001}
value	0.002	0.1	0.2	0.2	0.2	0.02	0.02	0.02

Parameter	β_{110}	β_{100}	β_{010}	β_{101}	β_{111}	β_{011}	β_{001}
value	0.1	0.02	0.02	0.02	0.02	0.02	0.02

Table 4.2: GARCH Parameters Used for Generating Synthetic Image

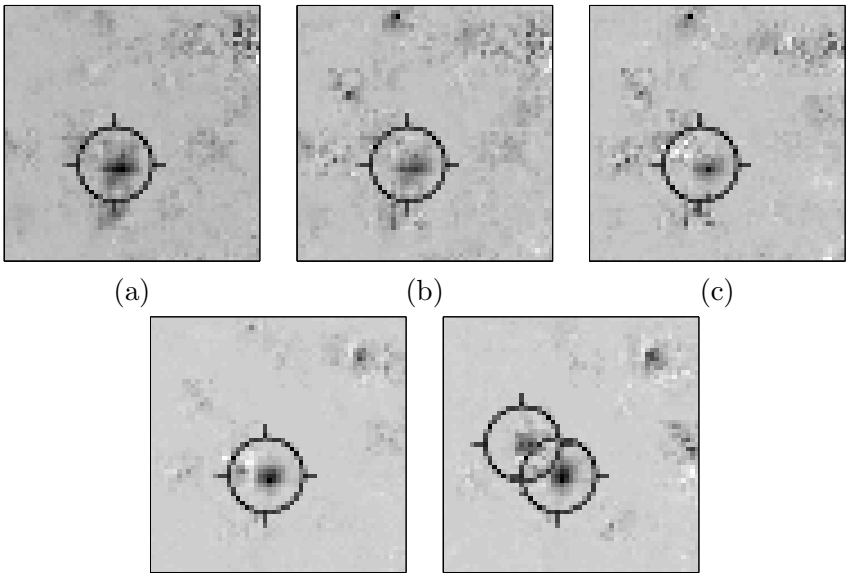


Figure 4.2: Layers 2-6 of the GLR with detected anomalies marked by a dark target sign

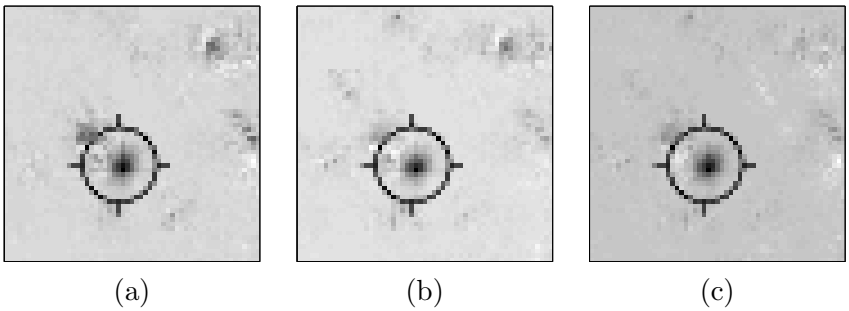


Figure 4.3: Detection using a sum of GLR layers: (a) Independent layers 2 and 6; (b) Dependent layers 2-6; (c) Layer 4 of the GLR using a 7 layers anomaly subspace.

4.4.2 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 imagery is a challenging task due to the high variability of the target and seabottom reverberation (background). An example of this variability can be seen in the top row of Figure 1.6, which shows 3 sea-mine sonar images. Mignotte *et al.* [27] present a two phase, three-class Markovian segmentation algorithm for the detection of sea mines in side-scan sonar. In the first phase the data is 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: sea-bottom reverberation and echo. Reed *et al.* [35] present a three phase procedure for detecting sea-mines in side-scan sonar data. 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. Goldman and Cohen [13] present a competing method based on 2D GMRF modeling of independent layers in a multiscale representation of the image. Independence of layers is achieved by means of the KLT. Anomaly detection is performed by using an appropriate subspace for each layer.

The proposed method has been applied to the images presented in the top row of Figure 1.6. A 5 layers feature space is created ($K_3 = 5$) for each image as described in Section 3.2 in the biorthogonal spline wavelets transform domain. We note that in our experiments, using different wavelet filters produces similar detection results. The anomaly subspace is created from arbitrarily selected 4 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 of this work. These 4 images are presented in Figure 4.4. The spatial size of the image chip is 7×7. The 4 chips in the image domain are presented in Figure 4.5. 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 4 images in Figure 4.4, in a similar manner to that used for the images in Figure 1.6. Anomaly chips of size $7 \times 7 \times 5$ are cut from the 4 feature spaces. The center of the chip is located in layer 3 at the spatial location corresponding to the center of the image chip presented in Figure 4.5. These 4 chips are consistently reordered into column vectors of size 245×1 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 image clutter. For the 1-D GARCH the GARCH(1;1) is often enough to capture characteristics of financial time series as noted by Hansen and Lunde [15]. 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 Figure 4.6. A black target like symbol marks the location of the detected anomaly. Note, that the positive detection is achieved in all 3 images (emphasis is given on the highlight region corresponding to the selected subspace). To further improve 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 Mignotte *et al.* [27] and Reed *et al.* [35] and describe above. We chose to compare our results with those of the GMRF based multiscale detection method of Goldman and Cohen [13]. Detection results of the GMRF based method on the sea-mine sonar images presented in the top row of Figure 1.6 are shown in the bottom row of Figure 1.6. 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 Figures 1.6 (a), (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 innovations 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 the images are presented in Section 1.2. 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, Figure 4.7 (a) presents a sea-mine sonar image, in which the mine object (probably a spherical object) differs from the mine objects used to create the anomaly subspace (elongated mines). Figure 4.7 (b) shows the detection results using the exact same wavelet transform and anomaly subspace used in the detection process leading to Figure 4.6.

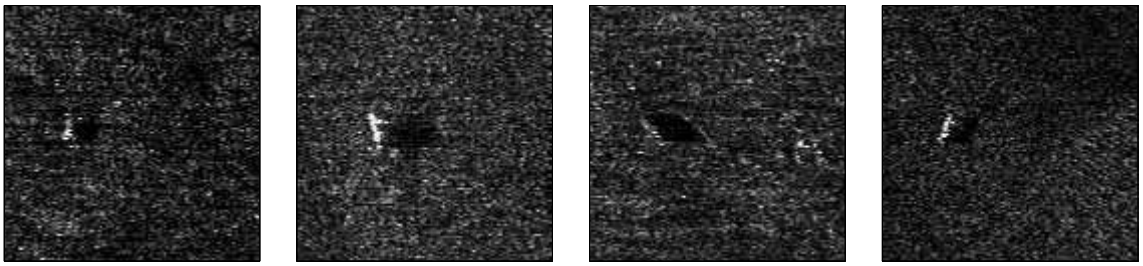


Figure 4.4: Original sea-mine sonar images from which an image chip is cut to create the anomaly subspace.

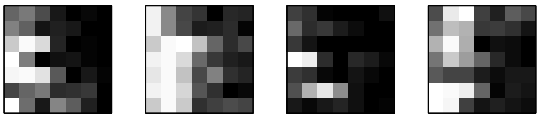


Figure 4.5: Image chips cut from the sea mine sonar images presented in Figure 4.4.

This demonstrate the potential of detecting mine like objects in sonar imagery using a subspace, which does not contain exact examples of such objects. As discussed in Section 3.2 casuality seems an unnatural model limitation. We therefor demonstrate detection results for different image orientations. The side-scan sonar image of Figure 1.6 (c) is rotated by 90^0 and 180^0 and the resulting images are presented in the top row of Figure 4.8. Detection is performed using the exact same procedures as above, only that the subspace images are rotated accordingly. Detection results are presented in the bottom row of Figure 4.8. Although positive detection without false alarms is achieved in all orientations (0^0, 90^0, 180^0) the detection images differ. In particular, Figure 4.8 (c) produces the best detection results, while Figure 4.8 (d) produces the worst detection of the three images. Due to the model casuality, it may be appropriate in some applications to consider all four possible orientations of a given image.

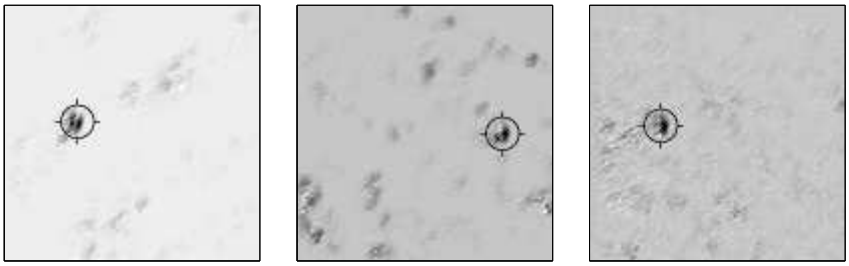
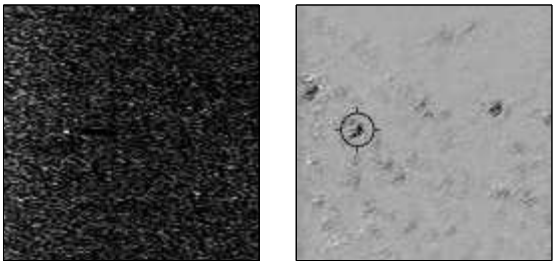


Figure 4.6: Detection results using the proposed method on the sea-mine sonar images of Figure 1.6.



(a) (b)

Figure 4.7: 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.

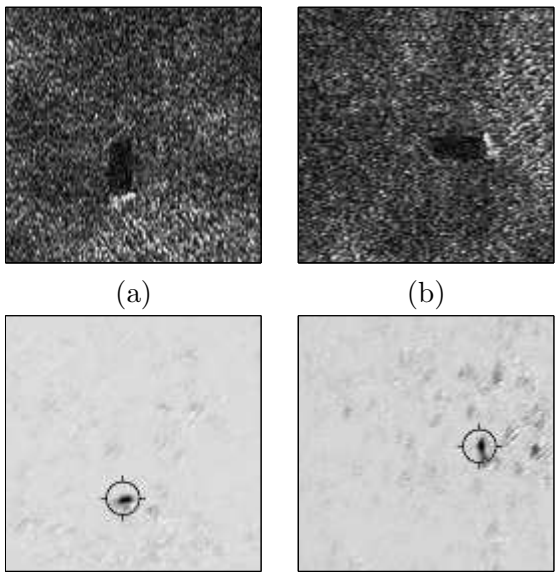


Figure 4.8: Detection results on rotated versions of the side-scan sea mine sonar image presented in Figure 1.6 (c): (a) original image rotated by 90^0; (b) original image rotated by 180^0. (c) Detection Results for Figure (a) using the proposed method; (d) Detection Results for Figure (b) using the proposed method.

4.5 Conclusion

We have developed a multiscale anomaly subspace detection method, corresponding to the multidimensional GARCH model introduced in chapter 3. The MSD enables incorporation of *a priory* 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 layers which are most significant to the detection. Layers are selected *a priory* or 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 sea-bottom reverberation. Compared with a GMRF based method, we presented improved performance, *i.e.*, a reduce false alarm rate while retaining a high detection rate.

Chapter 5

Conclusion

5.1 Summary

A novel clutter model is introduced: the multidimensional GARCH model. This statistical model can incorporate important characteristics of commonly used feature spaces, namely, heavy tailed distributions and clustering of innovations. We develope two anomaly detectors, each consists of a set of MSDs. The first detector is appropriate for detecting anomalies in three dimensional images, which can be properly modeled as a stack of two dimensional GMRFs. This approach was applied to the detection of faults is seismic data and to detection of defects in bare wafers. The second detector is suitable for feature spaces which present non-Gaussian characteristics, namely: excess kurtosis and clustering of innovations. To demonstrate such a feature space we used the undecimated wavelet transform to create a 3-D feature space for a given 2-D image. The underlying assumption is that there exists a set of wavelet filters such that the resulting feature space can be properly modeled using a 3-D GARCH model.

Our detectors operate by applying a separate MSD to each layer of the feature space. The MSD was originally developed by Scharf and Frienlander [38] for signal detection in subspace interference and WGN. Here, we formulated a set of multiscale MSDs for signal detection in subspace interference and noise which follows a GMRF model (first detector) or GARCH model (second detector). The MSDs incorporate the available *a priory* information about the targets into the detection process and thus improve the detection performance. When a large training set of anomaly examples is available *a priory* we suggest using a technique known as eigen-pictures to create an anomaly subspace of a desired rank, which can be regarded as an anomaly basis spanning the anomaly subspace. The eigen-pictures technique is based on PCA, thus choosing a dimensionality reducing linear projection, which maximizes the scatter of all projected samples in the training set. Our approach takes into consideration the fact that not all feature space layers contribute uniformly to the detection process and allows for selection of the most relevant layers, where the relevance criteria is application dependent and independent of the detection algorithm.

The performance of the two detectors is demonstrated on appropriate examples, showing the potential advantage of using the detectors, the importance of selecting an appropriate statistical model for the feature space and the ability of the GARCH model to cope with image clutter characterized by non-Gaussian statistics.

5.2 Future Research

The approach we have proposed in this thesis opens a number of interesting topics for future study:

1. Non-causal GARCH like models.

The causality assumption incorporated into the ND GARCH model introduced in Chapter 3 is unnatural for images. Developing a non-causal statistical model characterized by a heavy tailed distribution and innovation clustering may lead to an improvement in clutter modeling by reducing the dependency on image orientation. As a result, such a model, may reduce the false alarm rate for a given detection rate.

2. The multifamily likelihood ratio test.

   The proposed MSD approach assumed prior knowledge about the spatial size of the anomaly and is unable to distinguish or decide between different anomalies. Kay [18, 19] proposed a multifamily likelihood ratio test for multiple signal model detection, which may be used to alleviate the limitation of an *a-priory* known anomaly size. Extending Kay's approach and applying it with the proposed statistical model may lead to a multiple hypothesis test in which anomalies of different spatial sizes, and potentially of different characteristics are combined in a single anomaly detection scheme, such that anomaly detection and classification can be achieved in a single step.

3. Anomaly subspace detection in non-stationary clutter.

   An underlying assumption of the proposed approach is that the entire image area or region of interest (ROI) is generated by a single statistical model. This assumption is in general inaccurate as scenes tend to include different background environments (such as grass, trees, soil, etc.). Smaller spatial areas may obey a single stationary behavior. Performing image segmentation, such that each segment follows a single GARCH model and performing anomaly detection for each segment separately may improve the detector's performance.

Appendix A

Maximum Likelihood Estimation of ϕ_ℓ Under H_0

We prove the ML estimation of $\boldsymbol{\phi}_l$ under H_0. The ML solution of $\boldsymbol{\phi}_l, \boldsymbol{\psi}_l$ under H_1 is easily proved by analogy. Under H_0 we have:

$$\frac{d}{d\boldsymbol{\phi}_l}\left\{\log\left[P\left(\boldsymbol{n}_l\right)\right]\right\} = \frac{d}{d\boldsymbol{\phi}_l}\left[\left(\boldsymbol{n}_l - S_l\boldsymbol{\phi}_l\right)^T \Sigma_{\boldsymbol{v}_l}^{-1}\left(\boldsymbol{n}_l - S_l\boldsymbol{\phi}_l\right)\right] .$$

Opening the parenthesis on the right side, recalling that $\boldsymbol{n}_l^T\Sigma_{\boldsymbol{v}_l}^{-1}\boldsymbol{n}_l$ is independent of $\boldsymbol{\phi}_l$ and rearranging terms yields:

$$\frac{d}{d\boldsymbol{\phi}_l}\left\{\log\left[P\left(\boldsymbol{n}_l\right)\right]\right\} = \frac{d}{d\boldsymbol{\phi}_l}\left[\boldsymbol{\phi}_l^T S_l^T\Sigma_{\boldsymbol{v}_l}^{-1}S_l\boldsymbol{\phi}_l - 2\boldsymbol{n}_l^T\Sigma_{\boldsymbol{v}_l}^{-1}S_l\boldsymbol{\phi}_l\right] .$$

Setting the derivative to zero we have:

$$2S_l^T\Sigma_{\boldsymbol{v}_l}^{-1}S_l\boldsymbol{\phi}_l - 2(\boldsymbol{n}_l^T\Sigma_{\boldsymbol{v}_l}^{-1}S_l)^T = 0 .$$

Solving for $\boldsymbol{\phi}_l$ yields:

$$\boldsymbol{\phi}_l = (S_l^T\Sigma_{\boldsymbol{v}_l}^{-1}S_l)^{-1}S_l^T\Sigma_{\boldsymbol{v}_l}^{-1}\boldsymbol{n}_l .$$

Appendix B

The Causality Constraint

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

Let $q, p \geq 0$ denote the order of a non-causal symmetric GARCH model, and let Γ_1 and Γ_2 denote two neighborhood sets which are defined by

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

Let ϵ_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 $\eta_t \overset{iid}{\sim} N(0,1)$ be a stochastic process independent of h_k, $\forall k \neq t$. The non-causal $GARCH(p,q)$ process is defined as:

$$\epsilon_t = \sqrt{h_t}\,\eta_t \tag{B.1}$$

$$h_t = \alpha_0 + \sum_{k\in\Gamma_1} \alpha_k \epsilon_{t-k}^2 + \sum_{k\in\Gamma_2} \beta_k h_{t-k}\,, \tag{B.2}$$

and is therefore conditionally distributed as:

$$\epsilon_t \mid \psi_t \sim N(0, h_t)\,. \tag{B.3}$$

In order to guarantee a non-negative conditional variance we require that

$$h(t) \geq 0, \ \forall t \tag{B.4}$$

We need to find conditions on the parameters space $\left\{\alpha_0, \{\alpha_k\}_{k\in\Gamma_1}, \{\beta_k\}_{k\in\Gamma_2}\right\}$ such the equation (B.4) holds. Substituting (B.1) into (B.2) yields:

$$h_t = \alpha_0 + \sum_{k\in\Gamma_1} \alpha_k \eta_{t-k}^2 h_{t-k} + \sum_{k\in\Gamma_2} \beta_k h_{t-k}\,, \tag{B.5}$$

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 non-negative solution. In addition, the equation parameters include a stochastic process η_{t-k}^2, which cannot be limited in any way. This means, that causality is a necessary constraint to guarantee a non-negative conditional variance.

Appendix C

Proof of Theorem 1

In this appendix we prove Theorem 1 presented in Section 3.3.

Repeating substitutions of (3.1) into (3.2) yields:

$$\begin{aligned} h_{\mathbf{i}} &= \alpha_0 + \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} \eta^2_{\mathbf{i}-\mathbf{k}} h_{\mathbf{i}-\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{k}} \\ &= \alpha_0 + \sum_{\mathbf{r}\in\Gamma_1} \alpha_{\mathbf{r}} \eta^2_{\mathbf{i}-\mathbf{r}} \\ &\quad \times \left[\alpha_0 + \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} \eta^2_{\mathbf{i}-\mathbf{r}-\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} \right] \\ &\quad + \sum_{\mathbf{r}\in\Gamma_2} \beta_{\mathbf{r}} \left[\alpha_0 + \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} \eta^2_{\mathbf{i}-\mathbf{r}-\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} h_{\mathbf{i}-\mathbf{r}-\mathbf{k}} \right] \\ &= \alpha_0 \sum_{g=0}^{\infty} M(\mathbf{i}, g) \end{aligned} \tag{C.1}$$

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

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

for

$$\sum_{\mathbf{k}\in\Gamma_1} a_{\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} b_{\mathbf{k}} = g$$

$$\sum_{\mathbf{k}\in\Gamma_1} a_{\mathbf{k}} = n$$

and

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

Thus,

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

and in general

$$\begin{aligned}
M(\mathbf{i}, g+1) &= \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} \eta^2_{\mathbf{i}-\mathbf{k}} M(\mathbf{i}-\mathbf{k}, g) \\
&\quad + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}} M(\mathbf{i}-\mathbf{k}, g) \qquad \text{(C.2)}
\end{aligned}$$

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

$$E\{M(\mathbf{i}, g)\} = E\{M(\mathbf{k}, g)\} \ \forall \mathbf{i}, \mathbf{k}, g \tag{C.3}$$

from (C.2) and (C.3) we obtain

$$\begin{aligned} E\{M(\mathbf{i}, g+1)\} &= \left[\sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}}\right] E\{M(\mathbf{i}, g)\} \\ &= \left[\sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}}\right]^{g+1} E\{M(\mathbf{i}, 0)\} \\ &= \left[\sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} + \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}}\right]^{g+1} \end{aligned} \tag{C.4}$$

Finally by (3.1), (C.1) and (C.4),

$$\begin{aligned} E\{\epsilon_{\mathbf{i}}^2\} &= \alpha_0 E\left\{\sum_{g=0}^{\infty} M(\mathbf{i}, g)\right\} \\ &= \alpha_0 \sum_{g=0}^{\infty} E\{M(\mathbf{i}, g)\} \\ &= \alpha_0 \left[1 - \sum_{\mathbf{k}\in\Gamma_1} \alpha_{\mathbf{k}} - \sum_{\mathbf{k}\in\Gamma_2} \beta_{\mathbf{k}}\right]^{-1} \end{aligned} \tag{C.5}$$

if and only if

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

and

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

follows immediately.

References

[1] E. A. Ashton. Detection of subpixel anomalies in multispectral infrared imagery using an adaptive bayesian classifier. *IEEE Transcactions on Geoscience and Remote Sensing*, 36(2):506–517, March 1998.

[2] N. Balram and J. M. F. Moura. Noncausal Gauss Markov random fields: Parameter stracture and estimation. *IEEE Transactions on Information Theory*, 39(4):1333–1355, July 1993.

[3] P. N. Belhumeur, J .P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[4] E. K. Berndt, B. H. Hall, R. E. Hall, and J. A. Hausman. Estimation inference in nonlinear structural models. *Annals of Economics and Social Measurments*, 4:653–655, 1974.

[5] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[6] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characteristics in the wavelet domain. *IEEE Transaction on Image Processing*, 8(12):1688 – 1701, December 1999.

[7] C. I. Chang and S. S. Chiang. Anomaly detection and classification for hyperspectral imagery. *IEEE Transaction on Geoscience and Remote Sensing*, 40(6):1314 – 1325, June 2002.

[8] I. Cohen and R. R. Coifman. Local discontinuity measures for 3-D seismic data. *Geophysics*, 67(6):1933–1945, Nov.-Dec. 2002.

[9] R. F. Engle. Autoregressive conditional hetroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, July 1982.

[10] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., 1990.

[11] A. Goldman. Anomaly subspace detection based on a multi-scale Markov random field model. Master's thesis, Technion - Israel Institute of Technology, Haifa, August 2004.

[12] A. Goldman and I. Cohen. Anomaly detection based on an iterative local statistics approach. *Signal Processing*, 84(7):1225–1229, July 2004.

[13] A. Goldman and I. Cohen. Anomaly subspace detection based on a multi-scale Markov random field model. *Signal Processing*, 85(3):463–479, March 2005.

[14] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[15] P. R. Hansen and A. Lunde. A comparison of volatility models: does anything beat a GARCH(1,1)? working paper, Center for Analytical Finance, University of Aarhus, March 2001.

[16] G. G. Hazel. Multivariate Gaussian mrf for multispectral scene segmentation and anomaly detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38(3):1199–1211, May 2000.

[17] R. L. Kashyap and R. Chellappa. Estimation and choice of neighbors in spatial interaction models of images. *IEEE Transactions on Informain Theory*, IT-29(1):60–72, January 1983.

[18] S. Kay. Exponentially embedded families-new approaches to model order estimation. *IEEE Transcactions on Aerospace and Electronic Systems*, 41(1):333–345, January 2005.

[19] S. Kay. The multifamily likelihood ratio test for multiple signal model detection. *IEEE Signal Processing Letters*, 12(5):369–371, May 2005.

[20] I. G. Kazantsev, I. Lemahieu, G. I. Salov, and R. Denys. Statistical detection of defects in radiographic images in nondestructive testing. *Signal Processing*, 82(5):791–801, May 2002.

[21] S. Kraut, L. L. Scharf, and L. T. McWhorter. Adaptive subspace detectors. *IEEE Transactions on Signal Processing*, 49(1):1–16, January 2001.

[22] A. F. Laine, S. Schuler, J. Fan, and W. Huda. Mammographic feature enhancement by multiscale analysis. *IEEE Transactions on Medical Imaging*, 13(4):725–740, December 1994.

[23] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11(7):674 – 693, July 1989.

[24] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1):29–43, January 2002.

[25] D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[26] M. Markou and S. Singh. Novelty detection: a review - part 1: statistical approaches. *Signal Processing*, 83(12):2481–2497, December 2003.

[27] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy. Three-class Markovian segmentation of high-resolution sonar images. *Computer Vision and Image Processing*, 76(3):191–204, November 1999.

[28] J. M. F. Moura and N. Balram. Recursive structure of noncausal Gauss-Markov random fields. *IEEE Transactions on Information Theory*, 38(2):334–354, March 1992.

[29] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley and Sons, INC., 1982.

[30] A. Noiboar and I. Cohen. Anomaly detection based on wavelet domain GARCH random field modeling. *Submitted to IEEE Transactions on Geosience and Remote Sensing*.

[31] A. Noiboar and I. Cohen. Anomaly detection in three dimensional data based on Gauss Markov random field modeling. In *Proc. 23rd IEEE Convention of the Electrical and Electronics Engineers*, pages 448–451, Israel, Sep. 2004.

[32] A. Noiboar and I. Cohen. Two-dimensional GARCH model with application to anomaly detection. In *Proc. 13th European Signal Processing Conference*, Turkey, Sep. 2005.

[33] P. Paalanen, J. K. Kamarainen, J. Ilonen, and H. Kalviainen. Feature representation and discrimination based on Gaussian mixture model probability desities - practices and algorithms. Research Report 95, Lappeenranta University of Technology, Finland, 2005.

[34] I. S. Reed and X. Yu. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transaction on Acoustics Speech and Signal Processing*, 38(10):1760–1770, October 1990.

[35] S. Reed, Y. Petillot, and J. Bell. An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE Journal of Oceanic Engineering*, 28(1):90–105, January 2003.

[36] S. Reed, Y. Petillot, and J. Bell. Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information. *IEE Proceedings Radar, Sonar and Navigation*, 151(1):48–56, February 2004.

[37] L. L. Scharf. *Statistical Signal Processing Detection, Estimation and Time Series Analysis*. Addison-Wesley Publishing Company, 1991.

[38] L. L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Transactions on Signal Processing*, 42(8):2146–2157, August 1994.

[39] S. M. Schweizer and J. M. F. Moura. Hyperspectral imagery: Clutter adaptation in anomaly detection. *IEEE Transaction on Information Theory*, 46(5):1855–1871, August 2000.

[40] S. M. Schweizer and J. M. F. Moura. Efficient detection in hyperspectral imagery. *IEEE Transactions on Image Processing*, 10(4):584–597, April 2001.

[41] A. Srivastava. Stochastic models for capturing image variability. *IEEE Signal Processing Magazine*, 19(5):63–76, September 2002.

[42] 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 Processing Magazine*, 19(1):58–69, January 2002.

[43] R. N. Strickland and H. I. Hahn. Wavelet transform methods for object detection and recovery. *IEEE Transactions on Image Processing*, 6(5):724–735, May 1997.

[44] A. S. Willsky. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458, August 2002.

[45] J. W. Woods. Two-dimensional discrete Markovian fields. *IEEE Transaction on Information Theory*, 18(2):232–240, March 1972.

[46] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K. R. Castleman. Subspace-based prototyping and classification of chromosome images. *IEEE Transactions on Image Processing*, 14(9):1277– 1287, September 2005.

[47] X. G. Xia, C. Boncelet, and G. Arce. Wavelet transform based watermark for digital images. *Optics Express*, 3(12):497–511, December 1998.

גילוי אנומליות מבוסס על מודל שדה אקראי במרחב Wavelet GARCH

אמיר נויבואר

גילוי אנומליות מבוסס על מודל שדה GARCH אקראי במרחב Wavelet

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

מגיסטר למדעים

בהנדסת חשמל

אמיר נויבואר

הוגש לסנט הטכניון — מכון טכנולוגי לישראל

שבט תשס"ז חיפה ינואר 2007

חיבור על מחקר נעשה בהדרכת פרופסור חבר ישראל כהן

בפקולטה להנדסת חשמל

הכרת תודה

היננני אסיר תודה לפרופסור ישראל כהן על הנחייתו בכל שלבי המחקר.

ברצוני להודות לבוחנים של תיזה זו, פרופסור דוד מלאך ופרופסור אמיר אברבוך, על הערותיהם הבונות והצעותיהם המועילות.

תודה מיוחדת לאישתי ענת על אהבתה ותמיכתה.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי

מוקדש לאישתי ענת

תוכן עניינים

תקציר

גילוי אנומליות בתמונות הינו תהליך של זיהוי קבוצות של מספר קטן של פיקסלים השונים מהמאפיינים הכלליים של התמונה. סוג התמונה, המאפיינים שלה וסוג האנומליה תלויים באפליקציה. אפליקציות אופייניות כוללות: גילוי של מטרות בתמונות, גילוי של מאפייני מוקשים בתמונות סונר ימיות, גילוי פגמים בפיסות סיליקון וגילוי של גידולים ממאירים בתצלומים רפואיים. אלגוריתמים לגילוי אנומליות בתמונות מורכבים על פי רוב משלשה שלבים: בחירה של מרחב מאפיינים שבו ההבחנה בין הרקע לבין האנומליה אפשרית, בחירה של מודל סטטיסטי עבור מרחב המאפיינים המייצג את הרקע ובחירה של אלגוריתם לגילוי אנומליות. השלב האחרון כולל בתוכו בחירה של מודל האנומליה, כלומר הגדרה של האנומליה או האנומליות הרלוונטיים לאפליקציה. עבודת מחקר זו מתרכזת בשני השלבים האחרונים.

הצגות מולטירזולוציה משמשות פעמים רבות כמרחב מאפיינים עבור אפליקציות של גילוי אנומליות ממספר סיבות:

- אובייקטים בתמונות מופיעים לרוב במספר גדלים. שימוש בתצוגה מרובת רזולוציות מהווה דרך טובה לאנליזה של אובייקטים כאלו בתמונות.
- שימוש בהצגת מולטירזולוציה מאפשר עיבוד מקבילי של רמות הרזולוציה ובכך מא-פשר יעילות גבוהה יותר וניצול משאבי מיחשוב בצורה מיטבית.
- אובייקטים בתמונה יוצרים תגובה ביותר מרמת רזולוציה אחת ועל כן ניתן להשתמש

בהצגת מולטירזולוציה כמרחב מאפיינים עליו פועל גלאי.

- בתמונות מופיעות לעיתים תופעות מחזוריות עם אורך מחזור משתנה. מרחב מול-טירזולוציה מאפשר לאפיין תופעות אלו.

- מאפייני האוריינטציה והרזולוציה של מרחב ה - wavelet קשורים לאופן פעולת מערכת הראיה האנושית ועל כן שימוש בהם יכול לאפשר גילוי אנומליות באופן דומה לזה המבוצע באמצעות ההתבוננות האנושית.

הפילוג הגאוסי משמש פעמים רבות כפונקצית הפילוג הסטטיסטי של מרחב המאפיינים של הרקע לאור העובדה שניתן להגיע באמצעותו לתוצאות אנליטיות. חיסרון מרכזי של השימוש בפילוג הגאוסי נעוץ בחוסר היכולת למדל באמצעותו שתי תופעות אופיניות של מרחבי מאפיינים: "זנבות עבים" heavy tails של פונקצית צפיפות הפילוג של המאפיינים, כלומר ערך קורטוזיס (kurtosis) גבוה, ומקבצים של אי-יציבות (clustering of innovations), כלומר שינויים גדולים מופיעים במקבצים, ושינויים קטנים מופיעים במקבצים. שימוש באלגוריתמים לגילוי אנומליות המבוססים על הפילוג הגאוסי עבור מרחבי מאפיינים מסוג זה גורם לכמות גדולה של התראות שווא (false alarms) עקב חוסר ההתאמה בין המודל הסטטיסטי לבין התכונות הסטטיסטיות של הרקע. חוקרים שונים הראו שעבור מרחב ה - wavelet, שמשמש במקרים רבים כמרחב מאפיינים, מקדמי ה - wavelet מפולגים סטטיסטית עם ערך קורטוזיס גבוה. כמו כן קיים קשר סטטיסטי בין מקדמי ה - wavelet במיקומים מרחביים סמוכים ובין הרמות השונות. כלומר: מקדמים בעלי ערכים גבוהים מופיעים לרוב בקרבה מרחבית וכן ברמות ואוריינטציות סמוכות. לא ניתן למדל מאפיינים אלו באמצעות הפילוג הגאוסי ועל כן קיים צורך במציאת מודל סטטיסטי רב מימדי אלטרנטיבי.

לשם כך אנו מציגים בעבודה זו מודל generalized autoregressive conditional heteroscedasticity (GARCH) רב מימדי. מודל ה - GARCH החד מימדי משמש באופן נרחב למידול תופעות של סדרות בזמן בעיקר בתחום הכלכלה. הרחבת המודל למודל רב

מימדי יוצרת מודל מקורי למרחב המאפיינים של רקע בתמונה היכול למדל את שתי התופעות המוזכרות למעלה: "זנבות עבים" של פונקצית צפיפות הפילוג של המאפיינים ומקבצים של אי-יציבות. בעבודה זו נעשה שימוש בהתמרת מולטירזולוציה מסוג wavelet ללא דצימציה (undecimated) וניצור מרחב מאפיינים תלת מימדי עבור תמונה נתונה ע"י מיקום שכבות ההתמרה השונות זו על גבי זו ליצירת סריג תלת מימדי (קוביה). התמרת ה - wavelet ללא דצימציה הינה התמרה יתירה אולם משתמשים בה באותם המסננים כמו בהתמרת ה - wavelet הרגילה והיא ניחנת בתכונה הנוספת של אי-תלות בהזזה (translation invariance) שהינה חשובה מאוד באפליקציות של גילוי אנומליות. אנו משתמשים במודל GARCH תלת מימדי כמודל הסטטיסטי של מרחב מאפיינים זה. לאחר האיפיון הסטטיסטי של מרחב המאפיינים, אנו עומדים בפני האתגר של פיתוח שיטת גילוי מתאימה. באופן מעשי, במקרים רבים, האנומליות אותן אנו רוצים לגלות אינן מוגדרות באופן מדויק. אלגוריתמים המניחים ידע מלא על צורת האנומליה אותן מחפשים, כגון הגלאי מותאם האות (matched signal detector), אינם מתאימים. בנוסף, אלגוריתמים כגון בדיקת ההיפותיזה הבודדת (single hypothesis test) המניחים שלא קיים מידע על האנומליות כלל ומחפשים את כל מה ששונה מהרקע אינם מתאימים, מפני שעל פי רוב, דוגמאות אופייניות לאנומליות הינן זמינות. פתרון אפשרי לבעיות אלו, כלומר למצב שבו קיים ידע כלשהו על האנומליות אותן רוצים לגלות (למשל מספר דוגמאות אופייניות), ומעונינים לעשות שימוש במידע זה בתהליך הגילוי, אבל לא קיים ידע מדויק ומלא על האנומליות, הינו שימוש בגלאי מותאם תת מרחב (matched subspace detector). גלאי מותאם תת מרחב פותח לראשונה לשם גילוי של אות הנובע מתוך תת מרחב והנמצא תחת הפרעה הנובעת מתת מרחב ורעש גאוסי לבן. הגלאי מניח שהאנומליות אותן רוצים לגלות שייכות לתת מרחב מסוים וברשותנו מספיק מידע כדי לייצר בסיס הפורס את תת המרחב הזה. בצורה זו אין צורך במידע מלא על כל אנומליה אותה רוצים לגלות ויחד עם זאת, ניתן לשלב מידע אודות האנומליות ע"י הכללתו בבנית הבסיס הפורס את תת המרחב של האנומליות. החיסרון של גלאי תת המרחב המקורי

הינו העובדה שהוא פועל בסביבת רעש גאוסי לבן. החיסרון נובע מחוסר היכולת של המודל הגאוסי למדל תופעות של "זנבות עבים" ושל מקבצי אי יציבות כמוסבר למעלה. בעבודה זו נפתח שני גלאים. שניהם מבוססים על סט של גלאים מותאמי תת מרחב. הגלאים מותאמי תת מרחב שלנו פועלים ברעש גאוסי צבעוני (עבור מרחבי מאפיינים גאוסיים) וברעש GARCH (עבור מרחבי מאפיינים שמאופיינים במקבצים של אי יציבות ובפונציות פילוג שלהן "זנבות עבים"). לכל שכבה של מרחב המאפיינים (לדוגמא, במרחב ה - wavelet כל רמה ואוריינטציה יכולות להחשב כשכבה) נבחר תת מרחב של אנומליה ובצורה זו נאפשר שימוש רב במידע הקיים מראש. כאשר קיים סט אימון גדול של דוגמאות של אנומליות, נשתמש בתהליך שנקרא תמונות עצמיות (eigen-pictures) ליצירת תת מרחב של האנומליה מדרגה רצויה. כל תת מרחב כזה יכול לייצג גודל אנומליה שונה כך שניתן לייצר התאמה בין מרחב המאפיינים בשכבה הנתונה (בדגש על רמת רזולוציה ואוריינטציה) לבין תת-מרחב האנומליה. השיטה המוצעת על ידינו מאפשרת בחירה של השכבות התורמות בצורה משמעותית לתהליך הגילוי, ולהזנחה של השכבות האחרות. הקריטריון לבחירת השכבות תלוי באפליקציה.

אנו מדגימים את ביצועי סט הגלאים מותאמי התת מרחב הפועלים ברעש גאוסי צבעוני באמצעות אפליקציה של גילוי פגמים בפיסות סיליקון ובאמצעות אפליקציה של גילוי שברים וסדקים בשכבות קרקעיות המיוצגות באמצעות מידע סייסמי. בשתי האפליקציות המידע הוא תלת מימדי, ומרחב התמונה התלת מימדי משמש כמרחב המאפיינים. למרות הפוטנציאל של השיטה הנ"ל, כאשר מרחב המאפיינים הינו בעל תכונות של מקבצים של אי-יציבות ופונקציית פילוג שלה "זנבות עבים", הרי שהפילוג הגאוסי אינו מתאים והתוצאות המתקבלות אינן טובות (כמות גדולה של התראות שווא). עבור סביבות כאלו אנו מציעים את תהליך הגילוי מבוסס מודל ה - GARCH. אנו מדגימים זאת באמצעות דוגמאות על מידע סינטטי ובאמצעות תמונות סונר של מוקשים ימיים. עבור תמונות הסונר אנו יוצרים מרחב מאפיינים באמצעות התמרת ה - wavelet ללא דצימציה. פונקציית הפילוג של מרחב מאפיינים זה הינה בעלת "זנבות עבים". מרחב

זה מאופיין ע"י מקבצים של אי-יציבות. התוצאות מצביעות על הפוטנציאל של סט הגלאים מותאמי תת מרחב, על חשיבות בחירת מודל סטטיסטי מתאים לרקע ועל היתרונות של מודל ה - GARCH.