Anomaly Subspace Detection Based on a Multi-Scale Markov Random Field Model

Research Thesis

Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science in Electrical Engineering.

Arnon Goldman

Submitted to the Senate of the Technion - Israel Institute of Technology

Elul 5764 Haifa August 2004

The Research Thesis Was Done Under The Supervision of Dr. Israel Cohen in The Faculty of Electrical Engineering.

THE GENEROUS FINANCIAL HELP OF QUALCOMM IS GRATEFULLY ACKNOWLEDGED.

Contents

List of Figures

List of Tables

Abstract

Automatic target detection in natural scenes, is a challenging problem due to the large variability of background clutter and object appearance. When a typical signature of the target is available, the detection can be carried out by using a matched signal detector. Alternatively, anomaly detection methods are employed when no a priori information about the targets is available. The detection process is based on the anomalous nature of the targets with respect to the statistics of the background data. The statistics of the background data in images, is often described by random field models such as the Gaussian Markov random field (GMRF), the simultaneous auto-regressive (SAR) model, and long correlation (LC) models.

In many natural clutter images, scene elements often appear to have several periodical patterns, of various period lengths. In such cases, conventional random field models may not sufficiently describe the background clutter, and as a result, the anomaly detection is unreliable. Furthermore, in real detection problems, some partial information about the targets may be available in terms of a subspace in which the target signals lie. In such cases, a matched subspace detector (MSD), which considers the problem of detecting subspace signals in subspace interference and additive noise, may be used as part of the detection process.

In this work, we introduce multi-scale Gaussian markov random field (GMRF) models and a corresponding anomaly subspace detection algorithm. We develop the algorithm for target detection in cluttered images which follow one of the proposed models. This model is based on a multi-scale representation of the image. Using the Karhunen Loeve transform (KLT) we generate from the multi-scale representation independent layers. We assume that these independent layers can be modeled as GMRFs with different sets of parameters. The detection is subsequently carried out by using a modification of the MSD. The MSD was originally developed for signal detection in subspace interference and white Gaussian noise. Here we formulate a MSD for subspace signal detection in clutter, which follows the multi-scale GMRF model.

A quantitative performance analysis with a comparison to competing methods shows the advantages of the proposed method. The proposed model and algorithm are applied to several target detection problems: detection of airplanes in simulated cloudy backgrounds, detection of sea-mines in real sonar images, and detection of defects in wafer images. The results demonstrate the robustness and flexibility of the algorithm in adverse environments.

Glossary

Abbreviations

AR auto-regressive

DFT discrete Fourier transform

EM expectation maximization

FAR false alarm rate

GLC generalized long correlation

GLCP grey level co-occurrence probability

GLD grey level difference

GLRL grey level run length

GLRT generalized likelihood ratio test

GMRF Gaussian Markov random field

GMM Gaussian mixture model

HIP hierarchical image probability

HMT hidden Markov tree

ICA independent components analysis

IDFT inverse discrete Fourier transform

IID independently and identically distributed

KLT Karhunen-Loéve transform

LC long correlation

LPF low pass filter

LS least squares

ML maximum likelihood

MLE maximum likelihood estimate

MPL maximum pseudo-likelihood

MRF Markov random field

MSD matched subspace detector

MSP multi-scale stochastic process

NCAR noncausal autoregressive

PCA principle components analysis

pdf probability density function

ROC receiver operating characteristics

ROI region of interest

SAR simultaneous autoregressive

SNR signal-to-noise ratio

SVD singular value decomposition

Alphabetic Symbols

$*$	Convolution.
$\mathcal{A}$	A subset of Ω which contain the anomalous pixels.
$\mathcal{A}_j$	A set of pixels which have been identified as anomalous in the j-th iteration.
$\mathcal{A}_j^{(i)}$	A set of pixels which have been identified as anomalous in the i-th internal iteration of the j-th external iteration.
$\mathbf{b}$	An additive IID Gaussian Noise. $\mathbf{b} \sim \mathcal{N}(0, \rho^2 I)$.
$\mathbf{b}_\ell$	The ℓ-th layer of the innovations process of the additive noise in an image.
$\widehat{\mathbf{b}}_{H_0}^{\ell}$	The maximum likelihood estimate of the additive noise vector, $\mathbf{b}_\ell$, under H_0.
$\widehat{\mathbf{b}}_{H_1}^{\ell}$	The maximum likelihood estimate of the additive noise vector, $\mathbf{b}_\ell$, under H_1.
$\widehat{\mathbf{b}}_{H_0}$	The maximum likelihood estimate of the additive noise vector, $\mathbf{b}$, under H_0.
$\widehat{\mathbf{b}}_{H_1}$	The maximum likelihood estimate of the additive noise vector, $\mathbf{b}$, under H_1.
$\mathcal{B}$	A subset of Ω which contain the background pixels.
$\mathcal{B}_j$	A subset of $\mathcal{B}$ which contain pixels identified as anomalous in the j-th iteration.

$\mathcal{B}_j^{(i)}$	A subset of $\mathcal{B}$ which contain pixels identified as anomalous in the i-th internal iteration of the j-th external iteration.
$C(H_1\|H_0)$	Cost of false detection.
$C(H_0\|H_1)$	Cost of miss detection.
d	Differencing parameter of the LC model.
$d(\mathbf{q}(\mathbf{s}))$	Normalized distance of the feature vector $\mathbf{q}(\mathbf{s})$, from its expected vector κ.
D	Distance threshold (part of a decision rule).
$\mathcal{D}_p$	An operator which calculates the prediction error, $\widehat{\boldsymbol{\varepsilon}}(\mathbf{s})$, of the multi-scale Gaussian Markov random field (GMRF) model with p independent components.
$f(\mathbf{q}(\mathbf{s})\|H_i)$	The conditional pdf of $\mathbf{q}(\mathbf{s})$ given H_i.
$\bar{f}(i_1, i_2)$	A 2-D separable low-pass filter.
$f(i_1)$	A 1-D low-pass filter: $\bar{f}(i_1, i_2) = f(i_1)f(i_2)$.
$\mathcal{F}$	The 2-D DFT operator.
$\mathbf{g}(\mathbf{s})$	The column stack ordering of the neighborhood of $T(\mathbf{s})$.
$\mathcal{G}$	A set of multi-scale spatially invariant 2-D filters.
G_i	A spatially invariant 2-D filter.
h	A stabilization constant used in the LC model equation.
$\{\mathbf{h}_j\}$	Image chips which span the signal subspace.
H	The matrix whose columns span the signal subspace.
$\langle H \rangle$	The signal subspace.
H_ℓ	The matrix whose columns span the signal subspace (in the ℓ-th layer of the innovations process image).

$\langle H_\ell \rangle$	The signal subspace (of the ℓ-th layer of the innovations process image).
H_0	Hypothesis which indicate absence of target signal in the vector $\mathbf{n}_\ell$.
H_1	Hypothesis which indicate presence of target signal in the vector $\mathbf{n}_\ell$.
$\langle H_\ell S_\ell \rangle$	A subspace, spanned by the columns of the concatenated matrix $\begin{bmatrix} H_\ell & S_\ell \end{bmatrix}$.
$\{\mathbf{k}_j\}$	A set of N IID sample vectors having probability density function (pdf) $p_0(\cdot, \theta_0)$.
K	A matrix whose columns are the top p eigen vectors of the covariance matrix of $\mathbf{Y}(\mathbf{s})$.
$[\cdot]_\ell$	The ℓ-th layer of 3-D data.
$L(\cdot)$	Generalized likelihood ratio test.
$L_\ell(\mathbf{s})$	The log-likelihood ratio, calculated for pixel $\mathbf{s}$ based on the ℓ-th layer of the innovations process.
$L(\mathbf{s})$	The log-likelihood ratio, calculated for pixel $\mathbf{s}$ based on p layers of the innovations process.
m	The dimension of vector $\mathbf{k}_j$.
$M_x \times M_y$	The size of image T.
$n_{\mathbf{q}}$	The dimension of $\mathbf{q}(\mathbf{s})$.
$\mathbf{n}_\ell(\mathbf{s})$	The column stack ordering of an $N_x \times N_y$ pixels image-chip of $\widehat{\varepsilon}_\ell$ around $\mathbf{s}$.
N	Number of independently identically distributed (IID) sample vectors.
N_f	The length of filter $f(i_1)$ is $(2N_f - 1)$.

$N_x \times N_y$	Size of the image chips $\mathbf{h}_j$ and $\mathbf{s}_j$.
$\mathcal{N}(\boldsymbol{\mu}, \Gamma)$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance Γ.
$p_0(\cdot, \theta_0)$	The probability density function (pdf) of y under H_0.
$p_1(\cdot, \theta_0)$	The probability density function (pdf) of y under H_1.
$p_{d_2}(\zeta)$	The pdf of $d^2(\mathbf{q}(\mathbf{s}))$ under H_0.
P_S	The projection of a vector onto the subspace $\langle S \rangle$.
P_{HS}	The projection of a vector onto the subspace $\langle HS \rangle$.
$Pr(\cdot)$	Probability of an occurrence.
P_{S_ℓ}	The projection of a vector onto the subspace $\langle S_\ell \rangle$.
$P_{H_\ell S_\ell}$	The projection of a vector onto the subspace $\langle H_\ell S_\ell \rangle$.
P_D	Probability of detection.
P_{FA}	Probability of false-alarm.
$\mathcal{P}_G(\ell)$	A Gaussian image pyramid (ℓ is the layers indices).
$\mathcal{P}_L(\ell)$	A Laplacian image pyramid (ℓ is the layers indices).
$\mathbf{q}(\mathbf{s})$	Feature vector related to pixel $\mathbf{s}$.
Q	The presumed maximum number of targets in the image.
Q_j	The number of potential targets in $\mathcal{A}_j$.
$\mathbf{r}$	Relative indices of a neighbor in $\mathcal{R}$.
r_1 , r_2	The horizontal and vertical components of the indices $\mathbf{r}$.
RX$(\cdot)$	The Reed Xiaoli GLRT.
$\mathcal{R}$	A set of indices representing the neighborhood of a pixel.
$\mathcal{R}_h$	Half of the symmetric neighborhood $\mathcal{R}$.
$\mathbf{s} = (s_1, s_2)$	Indices of a pixel in the image.
$\{\mathbf{s}_j\}$	Image chips which span the interference subspace.
S	The matrix whose columns span the interference subspace.

$\langle S \rangle$	The interference subspace.
S_ℓ	The matrix whose columns span the interference subspace (in the ℓ-th layer of the innovations process image).
$\langle S_\ell \rangle$	The interference subspace (of the ℓ-th layer of the innovations process image).
T	An image.
$\mathbf{T}(\mathbf{s})$	A multi-scale representation of image Y with independent layers.
$^{\mathrm{T}}$	Transpose.
u	Rank of the signal subspace.
$\mathrm{vec}(\cdot)$	Column stack ordering of an image chip.
$V(\mathbf{s})$	A matrix of coefficients (used by the multi-resolution Markov model).
$\mathbf{w} = (w_1, w_2)$	The 2-D indices of data in the frequency domain.
$\mathbf{x}$	The target signal in the innovations process image.
$\mathbf{x}_\ell$	The target signal in the ℓ-th layer of the innovations process image.
$\mathbf{y}$	A vector to be classified.
$\mathbf{y}_o(\mathbf{s})$	An observation vector containing the values of the pixels in the neighborhood of $\mathbf{s}$.
$y(s)$	A random process.
Y	An image.
$\mathbf{Y}$	A multi-scale image generated from Y using $\mathcal{G}$.
z	Indices of the data in the z domain.
$\mathcal{Z}\{\cdot\}$	The z-transform operator.
α, β	Parameters of the Γ density function.

$\gamma(\mathbf{s})$	An operator which points to the parent of node $\mathbf{s}$ in a multi-resolution tree.
$\widehat{\Gamma}$	The sample covariance matrix of the reference data $\{\mathbf{k}_j\}$.
δ	Threshold on the relative change $\Delta_j^{(n)}$ (a stopping rule).
$\Delta_j^{(n)}$	The relative change between $\mathcal{B}_j^{(n)}$ and $\mathcal{B}_j^{(n-1)}$.
$\varepsilon(\mathbf{s})$	Innovation process.
$\boldsymbol{\varepsilon}(\mathbf{s})$	Vector of innovations in pixel $\mathbf{s}$.
$\varepsilon_\ell(\mathbf{s})$	The ℓ-th component of $\boldsymbol{\varepsilon}(\mathbf{s})$.
$\widehat{\boldsymbol{\varepsilon}}(\mathbf{s})$	Estimate of the innovations process (prediction error) in pixel $\mathbf{s}$.
$\widehat{\varepsilon}_\ell(\mathbf{s})$	The ℓ-th component of $\widehat{\boldsymbol{\varepsilon}}(\mathbf{s})$.
$\boldsymbol{\theta}$	Column stack ordering of $\{\theta(\mathbf{r}), \mathbf{r} \in \mathcal{R}\}$.
$\widehat{\boldsymbol{\theta}}$	Least-squares estimate of $\boldsymbol{\theta}$.
$\theta(\mathbf{r})$	Weight coefficient of a neighbor $\mathbf{r} \in \mathcal{R}$.
$\Theta_{\mathbf{r}}$	A diagonal matrix with weight coefficients of neighbor $\mathbf{r}$ (in the different layers of $\mathbf{T}$).
κ	Expected feature vector: $\kappa = E[\mathbf{q}(\mathbf{s})\|H_0]$.
$\nu(\mathbf{s})$	IID Gaussian random variables, with zero mean and unit variance.
ξ	Decision of the GLRT (H_0 or H_1).
ρ^2	The variance of the innovations process.
$\widehat{\rho^2}$	Least-squares estimate of ρ^2.
ρ_ℓ^2	The variance of $\mathbf{b}_\ell$.
Σ	Covariance matrix of the feature vector $\mathbf{q}(\mathbf{s})$, under H_0.
τ	τ=1,2 if the data $\{\mathbf{k}_j\}$ are real or complex valued, respectively.

φ	A confidence level - the probability of correctly deciding on H_0 given H_0 is true.
$\boldsymbol{\phi}$	Coefficients vector of the interference signal in the interference subspace.
$\boldsymbol{\phi}_\ell$	Coefficients vector of the interference signal in the interference subspace.
Φ_ℓ	The covariance matrix of $\frac{\mathbf{b}_\ell}{\rho_\ell}$.
$\chi^2_q(c)$	The chi-square probability distribution function, with q degrees of freedom and non-centrality parameter c.
$\boldsymbol{\psi}$	Coefficients vector of the target signal in the signal subspace: $\mathbf{x} = H\boldsymbol{\psi}$.
$\boldsymbol{\psi}_\ell$	Coefficients vector of the target signal in the signal subspace: $\mathbf{x}_\ell = H_\ell\boldsymbol{\psi}_\ell$.
Ω	Support of an image.

Chapter 1

Introduction

1.1 Motivation and Goals

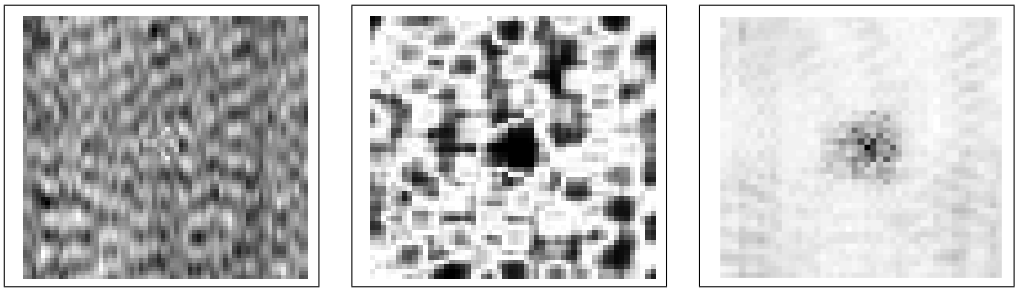
During the last decade, there has been a remarkable progress in random field models and their applications. Random field modeling has been applied extensively to texture synthesis [16], [29], image segmentation [46], [68], [72], and target detection [36], [63]. Natural clutter models are able to describe a wide range of image textures based on the interaction between neighboring pixels. Clutter models as the GMRF, simultaneous autoregressive (SAR), and generalized long correlation (GLC) model have proven to be effective in describing natural textures with periodical as well as random elements [4], [29]. In many natural clutter images, scene elements often appear to have several periodical patterns, of various period lengths. In such cases, random field models as the GMRF and the long correlation (LC) models may not sufficiently fit the clutter image. Deviations of the clutter image from the random field model influence the detection performance by increasing the false alarm rate. Furthermore, in real detection problems, some *a priori* information about the targets is often available. The information may include details about the exact nature of the target (the targets

signature) but may sometimes be more general - describing a subspace in which the target's signature may lie. Using this information in a flexible framework, for rejecting anomalies which do not resemble targets, may improve the detection performance, achieved by anomaly detection methods and matched filtering.

In this work, we introduce multi-scale Gaussian markov random field (GMRF) models and a corresponding anomaly subspace detection algorithm. Figure 1.1 presents the improvement in the detection performance, achieved by modeling an image using the multi-scale framework we propose. Figure 1.1(a) presents the original synthetic image, containing an airplane in its center. The synthetic background is a mixture of 3 different textures which contain periodic patterns with different period lengths. This image is not expected to be well described by the conventional GMRF model. Figure 1.1(b) presents the results of applying a single hypothesis test to the innovations process of the image in Fig. 1.1(a), assuming the image follows the GMRF model. The noisy results are caused by the incorrect assumption regarding the statistical model which describes the image. Figure 1.1(c) presents the likelihood ratios obtained by applying the proposed method to the image in Fig. 1.1(a), assuming the multi-scale GMRF model we propose. The results achieved by the proposed method seem to be less noisy and therefore are expected to produce less false alarms.

1.2 Overview of the Thesis

In this thesis, we introduce multi-scale GMRF models and a corresponding anomaly subspace detection algorithm. We develop the algorithm for target detection in cluttered images which follow one of the proposed models. This model is



(a) (b) (c)

Figure 1.1: A comparison between detection methods. (a) A synthetic image containing cloudy background and an airplane target in its center ; (b) Result of applying a single hypothesis test to the innovations process of the image in (a), assuming a GMRF model ; (c) Result of the proposed method applied to the image in (a), assuming a multi-scale GMRF model.

based on a multi-scale representation of the image and the Karhunen-Loéve transform (KLT). We generate from a given image, a multi-layer representation with independent layers. We assume that these independent layers can be modeled as GMRFs with different sets of parameters. The detection is subsequently carried out by applying a modification of the matched subspace detector (MSD) to the innovations process (prediction error) of the multi-scale GMRF. The MSD incorporates the available *a priori* information about the targets into the detection process and thus improves the detection performance. The MSD was originally developed for signal detection in subspace interference and white Gaussian noise [69]. Here, we formulate a MSD for signal detection in subspace interference and noise which follows the proposed multi-scale GMRF model. A quantitative performance analysis with comparison to competing methods shows the advantages of the proposed method. The proposed model and algorithm are applied to detection of airplanes in simulated cloudy backgrounds; detection of sea-mines in sonar images; and detection of defects in wafer images. The results demonstrate

the robustness and flexibility of the algorithm in adverse environments.

1.3 Organization

The structure of the thesis is as follows: In Chapter 2, we present statistical models of natural clutter. We review the SAR, GMRF and GLC models. In Chapter 3, we present anomaly detection approaches. We review the single hypothesis testing approach and the MSD. In chapters 4-8 we propose novel clutter models and detection methods. We analyze and demonstrate the method's performance using simulated and real images. In Chapter 4, we introduce an iterative approach for anomaly detection based on a single hypothesis test. In Chapter 5, we introduce multi-scale GMRF models. In section 6, we introduce an anomaly subspace detector using a multi-scale GMRF model, we analyze the performance of the proposed algorithm and compare the detection results to those obtained by using competing methods. In Chapter 7, we demonstrate the application of the proposed algorithm to automatic target detection in simulated and real imagery. Finally, in Chapter 8, we propose subjects for future research and conclude.

1.4 Background

Most detection and segmentation methods rely on the statistical characterization of the examined image. For some images, image segmentation could be easily implemented using only the intensity of each pixel. In such cases, no high order texture features are needed. But most of the time, the identification of class types in the image can not be done so easily based only on the grey levels of the pixels. Then the texture features become an helpful information in the segmentation

process. The simplest approach of texture analysis method in statistical category is known as first order methods [37]. The statistics are based on individual pixel values, rather then the relationship between pixels. Texture features such as mean, variance, standard deviation, gradient, and skewness are usually extracted from the image [37].

If an image shows regions with equal first order statistics, then second order statistics should be used. In the second order approach the texture features are extracted from an intermediate relationship matrix which is calculated using the pixels within a pixels neighborhood. There are several commonly used methods in this category [22]: grey level difference (GLD), grey level run length (GLRL), and grey level co-occurrence probability (GLCP). The GLCP method can measure textural characteristics such as homogeneity, grey level linear structure, contrast, entropy, and image complexity [20],[18],[19].

Most random field models are based on the spatial interaction of pixels in local neighborhoods. The noncausal autoregressive (NCAR) model represents each pixel as a linear combination of pixels at nearby locations, and an additive white noise variable. The Markov random field (MRF) replaces the white noise with a spatially correlated random variable. A wider discussion on the properties of these models and further details can be found in Chapter 5.2.

Model-based texture analysis is a mathematical process which can synthesize or describe the texture image. The parameters of the model are used to establish the distribution law of a pattern or simply used as features to classify the textures. The fundamental difference between the model-based methods and the statistical texture analysis methods is that the texture model has the capability to generate the texture which matches the observed texture. Using a statistical method, the

texture features are measured without a representative texture in mind, and the texture features extracted from the image, cannot be used in general to synthesize a texture image.

Random field models were developed for describing natural clutter images. Man-made objects therefore appear anomalous with respect to the random field model which describes the clutter. Anomaly detection methods make use of the anomalous appearance of such objects for their detection, but often make no *a priori* assumptions about the nature of the targets. Statistical approaches of anomaly detection, does not rely on an exhaustive statistical model of the targets, but rather on the statistics of the background data. The detection is carried out by estimating whether a test sample follows the same model as the background data. Most methods assume that the data comes from a family of known models and certain parameters are calculated to fit this model. These approaches are based on modeling the the training data and rejecting test patterns that fall in regions of low probability density. However, in order to achieve better performance, the training data itself needs to be either free of outliers or the outliers to be known.

Hazel [36] has developed an anomaly detection technique, which is based on GMRF modeling of the background in a multi-spectral image. A single hypothesis scheme is used for the detection of regions, which appear unlikely with respect to the probabilistic model of the background. A similar anomaly detection method was presented by Bello [3] for the detection of anomalous complex image pixels, using the SAR model. A completely different approach for target detection is based on a matched signal detector (matched filter). The matched signal detector is employed when a typical signature of the target is available. In many detection problems, the information about the targets is a subspace in which the targets lie.

In these applications, the matched signal detector is replaced by a MSD, a generalization of the matched filter, which was formulated by Scharf and Friedlander [69]. The MSD is used for detecting subspace signals in subspace interference and additive noise, using the principle of the generalized likelihood ratio test (GLRT). A state of the art review of anomaly detection methods can be found in Karkou and Singh [57]. The survey includes different statistical approaches for image modeling, hypothesis testing and clustering. Most of the presented methods are driven by modeling data distributions and then calculating the likelihood of test data with respect to the estimated statistical models.

Chapter 2

Modeling Natural Clutter Images

2.1 Introduction

Any analytical expression that explains the nature and extent of dependency of a pixel intensity on intensities of its neighbors can be said to be a model [13]. An image processing or analysis technique only applies well for certain kinds of images. Thus it is important to be able to model the image that is to be processed. Once the model of the image is obtained, it will serve to explain the statistical characteristics of the image and to efficiently process it. The awareness of the importance of image models was enhanced at the end of the 1970s by a workshop on image modelling held in Chicago in August, 1979 [66]. Since then, the research activities in the image modelling area have increased considerably [86].

During the last decade, there has been a remarkable progress in random field models and their applications. Random field modeling has been applied extensively to texture synthesis [16], [29], image segmentation [46], [68], [72], and target detection [36], [63]. Most random field models are based on the spatial in-

teraction of pixels in local neighborhoods [5]. The NCAR model represents each pixel as a linear combination of pixels at nearby locations, and an additive white noise variable (innovations process). Chellappa and Kashyap [16], [41] proposed an iterative estimation method and synthesis algorithm for the 2-dimensional NCAR model. They illustrated the usefulness of the NCAR models for synthesis of textures resembling several real texture images, possessing the local replication attribute. The local replication attribute is an essential ingredient of many natural textures [16].

The MRF model was first introduced by Lévi [49] in 1956. Woods [82] formulated the two-dimensional discrete MRF based on the continuous case and the mathematical foundation given by Lévi [49] and Wong [81]. The model was farther developed and studied by Geman and Geman [34], Chellappa and Chatterjee [14], and others. The discrete MRF model describes each pixel as a weighted sum of its neighboring pixels and a random variable which represents the innovations process. The difference between the MRF model and the NCAR model is that the innovations process is spatially correlated. Yue [86] investigated the texture distinguishing ability of the MRF models and the GLCP texture features. He Compared the segmentation performance of the MRF and GLCP methods for synthetic, Brodatz [9], and sea-ice images taken by a synthetic aperture radar.

A more general form of random field models is the LC model proposed by Kashyap and Lapsa [44]. The LC models can be applied to images with a correlation structure which extends over large regions using only a few model parameters. Kashyap and Eom demonstrated the applicability of the LC model by presenting a consistent estimation scheme [42] and a texture boundary detection method [43], based on this models. Eom [29],[28] proposed a LC model with circular and

elliptical correlation structure and a corresponding estimation algorithm. The LC model, proposed by Eom, has the advantage of modeling diverse real textures with less than five model parameters. Three parameters are used for defining an isotropic LC model and the other two parameters are used for describing the linear transformation (elongation and rotation) performed to the model's coordinate system. Bennett and Khotanzad [4] developed a random field model and a corresponding estimation scheme, based on a generalization of the LC model. They introduced the GLC model and showed that the NCAR and the MRF models are special cases of this model.

In this chapter we present some known clutter image models. We provide the SAR, GMRF and GLC models equations and present estimation and synthesis approaches. The organization of this chapter is as follows: In Section 2.2, we present the SAR model, in Section 2.3, we present the GMRF model, and in section 2.4 we present the GLC model.

2.2 The Simultaneous Auto-Regressive (SAR) Model

The SAR model [15], [16], [41] was one of the first random field models to be developed. It represents each pixel as a linear combination of pixels at nearby locations, and an additive white noise variable (innovations process).

Let Ω be the support of an image, and let $\mathbf{s} \in \Omega$ denote the indices of a pixel in the image. Let $\mathcal{R}$ be a given set of indices representing the neighborhood of a pixel (A simple example is the 4-neighbors set where $\mathcal{R}$={(-1,0),(1,0),(0,-1),(0,1)}). We denote the weight coefficient of a neighbor $\mathbf{r} \in \mathcal{R}$ by $\theta(\mathbf{r})$ and the

innovations process by $\varepsilon(\mathbf{s})$. The innovations process is a set of independently and identically distributed (IID) Gaussian variables with zero mean and variance ρ^2. Assuming an image y follows the SAR model, a pixel $y(\mathbf{s})$ in the image[1] is related to its neighboring pixels as follows:

$$y(\mathbf{s}) = \sum_{\mathbf{r}\in\mathcal{R}} \theta(\mathbf{r}) y\left(\mathbf{s}+\mathbf{r}\right) + \varepsilon\left(\mathbf{s}\right). \tag{2.1}$$

Chellappa and Kashyap [15] presented a procedure for synthetic generation of images which follow the SAR model. The procedure is based on the discrete Fourier transform (DFT) of (2.1):

$$\mathcal{F}\{y\}(w_1, w_2) = \frac{\rho\mathcal{F}\{\varepsilon\}(w_1, w_2)}{\lambda(w_1, w_2)} \tag{2.2}$$

where

$$\lambda(w_1, w_2) = 1 - \sum_{\mathbf{r}=(r_1,r_2)\in\mathcal{R}} \theta(\mathbf{r}) exp\left[2\pi i\left(\frac{r_1 w_1}{M_x} + \frac{r_2 w_2}{M_y}\right)\right] \tag{2.3}$$

and where M_x, M_y is the size of the image. This representation is obtained given the toroidal lattice assumption, which comfortably determine the boundaries condition:

$$y(M_x + s_1, M_y + s_2) = y(s_1, s_2) \tag{2.4}$$

The synthetic images are obtained by inverse discrete Fourier transform (IDFT) of (2.2). The SAR model is valid for all parameter values such that $\lambda(w_1, w_2) \neq 0$.

[1]For simplicity, we assume $y(\mathbf{s})$ is not in the boundaries of the image, *i.e.* $\forall\mathbf{r} \in \mathcal{R}$, $(\mathbf{s}+\mathbf{r}) \in \Omega$

2.3 The Gaussian Markov Random Field (GMRF) Model

The Markovian assumption is that the conditional probability of $y(\mathbf{s})$, given all the other values of y, depends only upon a finite group of neighboring pixels $\{y(\mathbf{s}+\mathbf{r})|\mathbf{r} \in \mathcal{R}\}$. The process is non-causal and is not driven by a white noise source. $y(\mathbf{s})$ is strict sense stationary and $\mathcal{R}$ is symmetric.

We assume that each image pixel can be represented as a weighted sum of its neighboring pixels and an additive innovations process (as described by (2.1)). The difference between the SAR model and the GMRF model is in the statistics of the innovations process. In a GMRF, the white Gaussian innovations process is replaced by a spatially correlated Gaussian noise. Woods [82] showed that under the Markovian assumption, the innovations process of a GMRF is spatially correlated with covariance given by:

$$E\left\{\varepsilon\left(\mathbf{s}\right)\varepsilon\left(\mathbf{s}+\mathbf{r}\right)\right\} = \begin{cases} \rho^2\,, & \text{if } \mathbf{r}=(0,0) \\ -\theta(\mathbf{r})\rho^2\,, & \text{if } \mathbf{r}\in\mathcal{R} \\ 0\,, & \text{otherwise.} \end{cases} \tag{2.5}$$

Kashyap and Chellappa [41] showed that the correlation structure imposes symmetry on the neighborhood set. That is, $\mathbf{r} \in \mathcal{R}$ implies $-\mathbf{r} \in \mathcal{R}$ and $\theta(\mathbf{r}) = \theta(-\mathbf{r})$. The GMRF model can also be written in terms of white noise sequence as follows [41]:

$$\mathcal{F}\{y\}(w_1,w_2) = \frac{\rho\mathcal{F}\{\varepsilon\}(w_1,w_2)}{\sqrt{\lambda(w_1,w_2)}} \tag{2.6}$$

where

$$\lambda(w_1,w_2) = 1 - 2\sum_{\mathbf{r}=(r_1,r_2)\in\mathcal{R}_h} \theta(\mathbf{r})cos\left[2\pi\left(\frac{r_1w_1}{M_x}+\frac{r_2w_2}{M_y}\right)\right] \tag{2.7}$$

and where $\mathcal{R}_h$ is half of the symmetric neighborhood $\mathcal{R}$. The difference between (2.3) and (2.7) originates from the symmetry, imposed by the correlation structure. Using (2.2) and (2.7), the procedure of synthetic generation of images which follow the GMRF model is evident.

In most detection problems, the background clutter model is unknown and therefor should be estimated. Various methods for model estimation were developed over the years, *e.g.*, [36], [41], [70], [71], [87]. A computationally efficient method for the GMRF model estimation is described in details in Kashyap and Chellappa [41]. Let vec$(\cdot)$ denote the column stack ordering of an image chip. Let the column stack ordering of the neighborhood of $y(\mathbf{s})$ be denoted by $\mathbf{g}(\mathbf{s})$:

$$\mathbf{g}(\mathbf{s}) = \text{vec}[y(\mathbf{s}+\mathbf{r}), \mathbf{r} \in \mathcal{R}] \tag{2.8}$$

and let

$$\boldsymbol{\theta} = \text{vec}[\theta(\mathbf{r}), \mathbf{r} \in \mathcal{R}]. \tag{2.9}$$

Kashyap and Chellappa showed that the least squares estimates for $\boldsymbol{\theta}$ and ρ^2 are given by [41]:

$$\widehat{\boldsymbol{\theta}} = \left[\sum_{\mathbf{s}\in\Omega} \mathbf{g}(\mathbf{s})\mathbf{g}(\mathbf{s})^{\mathrm{T}}\right]^{-1} \left[\sum_{\mathbf{s}\in\Omega} y(\mathbf{s})\mathbf{g}(\mathbf{s})\right] \tag{2.10}$$

$$\widehat{\rho^2} = \frac{1}{|\Omega|} \sum_{\mathbf{s}\in\Omega} \left(y(\mathbf{s}) - \widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{g}(\mathbf{s})\right)^2 \tag{2.11}$$

where $^{\mathrm{T}}$ denotes transpose.

2.4 Generalized Long Correlation (GLC) Models

One of the main drawbacks of the SAR and GMRF models is their inability to effectively model low frequency power in an image. The correlation function of these models decays rapidly beyond the span of the defined neighbor set. Bennett and Khotanzad [4] developed the GLC model and showed that it has an autocorrelation function which decays much more slowly and thus, enables effective modeling of clutter images with correlation of large spatial extent. They showed that the SAR and the GMRF models are special cases of the GLC model.

Let $y(s)$ be a random process which follows the one-dimensional LC model. We denote the differencing parameter by d (where $|d| < 0.5$) and the innovations process by ε. The innovations process is a white Gaussian noise sequence with zero mean and variance ρ^2. Then $y(s)$ can be described in the frequency domain by its z-transform as follows [4]:

$$(1 - hz^{-1})^d \mathcal{Z}\{y\}(z) = \rho \mathcal{Z}\{\varepsilon\}(z) \tag{2.12}$$

where $\mathcal{Z}$ is the z-transform operator and h is a stabilization constant (which moves the $z = 1$ pole into the unit circle). When $h = 1$, the model is unstable (unrealizeable). Selecting h to be slightly less than one enables a stable approximation of the model. In order to obtain the two-dimensional GLC model, (2.12) is generalized and extended as follows:

$$\left[1 - \sum_{\mathbf{r}=(r_1,r_2)\in\mathcal{R}} \theta(\mathbf{r}) z_1^{-r_1} z_2^{-r_2}\right]^d \mathcal{Z}\{y\}(z_1, z_2) = \rho \mathcal{Z}\{\varepsilon\}(z_1, z_2) \tag{2.13}$$

where the model parameters are the neighbor set $\mathcal{R}$, the correlation coefficients

$\theta(\mathbf{r})$, the differencing parameter d, and the innovations standard-deviation ρ. Let $\mathcal{F}$ be the DFT operator. Then the equivalent DFT representation of (2.13) is [4]:

$$\mathcal{F}\{y\}(w_1, w_2) = \frac{\rho \mathcal{F}\{\varepsilon\}(w_1, w_2)}{\lambda(w_1, w_2)^d} \tag{2.14}$$

where $\lambda(w_1, w_2)$ is given by (2.3). This representation is true under the toroidal lattice assumption ($y(M_x + 1, s_2) = y(1, s_2)$ and $y(s_1, M_y + 1) = y(s_1, 1)$).

For the special case of $d = 1$, (2.2) is identical to (2.14) and therefore the SAR model is equivalent to this special case of the GLC model. Similarly, the GMRF model is equivalent to the GLC model with $d = \frac{1}{2}$.

2.5 Summary

Clutter image modeling is an active field of research. During the last decade, there has been a remarkable progress in random field models and their applications. Most of the models are based on the spatial interaction of pixels in local neighborhoods. The main challenge in clutter modeling is to represent a wide range of textures using a small number of parameters and to efficiently estimate these parameters. In this chapter we present some known clutter image models. We provide the SAR, GMRF and GLC models equations and present estimation and synthesis approaches.

The SAR model represents each pixel as a linear combination of pixels at nearby locations, and an additive white noise variable (innovations process). This structure enables a simple procedure for image synthesis [4]. The more widely used GMRF model describes each pixel as a weighted sum of its neighboring pixels

and a spatially correlated innovations process. The GLC model, proposed by Bennett and Khotanzad [4], is a more general form of random field models. This model is based on a generalization of the LC model [29] and includes the SAR and the GMRF models as its special cases. It better describes textures with long correlation characteristics but lacks the simplicity which characterize the SAR and the GMRF estimation and synthesis methods.

The statistical models, presented in this chapter, were developed for describing clutter images. Regions in the image which are not part of the background clutter, may appear anomalous with respect to its model. In chapter 3 we review anomaly detection methods. These methods are used for detecting regions in the image which appear improbable with respect to the background statistical model.

Chapter 3

Anomaly Detection

3.1 Introduction

In the previous chapter, we presented statistical models of background clutter in images. The majority of work in the area of target detection has focused on detection methods, which involve statistical characterization of both targets and background [79]. Matched filters, for example, require *a priori* knowledge of a typical signature of the target [77]. In a realistic situation, however, there is a wide variety of potential targets which do not conform to a uniform model. Given the fact that we can never train a machine learning system on all possible object classes, it becomes important that it is able to distinguish between known and unknown object information during testing. Anomaly detection is the identification of new or unknown data that a learning system is not aware of during training [57]. Anomaly detection techniques are useful in applications such as fault detection [52], radar target detection [12], detection of masses in mammograms [74], e-commerce [52], and other signal processing and image analysis applications.

Statistical approaches of anomaly detection, does not rely on an exhaustive sta-

tistical model of the targets, but rather on the statistics of the background data. The detection is carried out by estimating whether a test sample comes from the same distribution as the background data. Two main approaches exist to the estimation of the pdf of the background data, parametric and non-parametric methods [25]. The parametric methods assume that the data comes from a family of known distributions (e.g. the normal distribution) and certain parameters are calculated to fit this distribution. The simplest detection method can be based on constructing a pdf for data of a known class. The detection is then carried out by computing the probability of a test sample of belonging to that class and thresholding these estimates. Another simple method is to threshold the distance of a sample from a class mean (in terms of number of standard deviations) [55], [56]. The distance measure can be Mahalanobis or some other probabilistic distance [78].

Roberts and Tarassenko [64] developed a method for anomaly detection in Gaussian mixture models (GMMs). Their method aims to minimize the number of heuristically chosen thresholds in the novelty decision process. Their anomaly detection method is similar to those presented by Barnett and Lewis [2], Bishop [8], Brotherton et. al. [10], Desforges et. al. [25], Parra et. al. [60], Tarassenko et. al. [75], Tarassenko [74], Tax et. al. [76], Yeung and Chow [85], and others. The anomaly detection is based on the estimation of the density function of the training data. The major contribution of this paper is that an automatic criterion is used for determining the number of Gaussians. The growth decision is based on measuring the smallest Mahalanobis distance between a training vector and each Gaussian within the network. This method was tested on a medical signal processing task to detect elpileptic seizures, exhibiting very high perfor-

mance rates. Van Leemput et. al. [48] presented an outlier detection algorithm for detection of multiple sclerosis lesions from multi-spectral magnetic resonance images. The algorithm is based on the MRF model. Tissue specific intensity models are estimated from the data itself and data which appear unlikely with respect to the model is suspected to be part of a lesion.

In non-parametric methods the overall form of the density function is derived from the data as well as the parameters of the model (e.g. histogram analysis). As a result non-parametric methods give greater flexibility in general systems. The parametric as well as the non-parametric approaches are based on modeling the density of the training data and rejecting test patterns that fall in regions of low density. However, in order to achieve better performance, the training data itself needs to be either free of outliers or the outliers to be known.

A different approach for anomaly detection is based on clustering - partitioning data into a number of clusters [61], [84]. Each data point can be assigned a degree of membership to each of the clusters. Thresholding the degree of membership can use for detecting data points which belong to none of the available classes.

The GLRT for unknown signals is the foundation for many anomaly detection algorithms. Kelly [45] and Reed and Yu [62] developed GLRTs for multidimensional image data assuming that the target signal and the covariance of the background are unknown. Scharf and Friedlander [69] formulated the MSD - an extension of the GLRT for unknown signals, to situations in which the target signature is assumed to be in a known subspace. The matched filter and the single hypothesis test are special cases of the MSD (generated for subspaces of rank 1 and of full rank respectively).

In Sec. 3.2 we describe the single hypothesis test, formulated based on the GLRT for unknown signals, assuming Gaussian noise. Then, in Sec. 3.3 we present the MSD, as formulated by Scharf and Friedlander [69].

3.2 Single Hypothesis Testing

Applying the GLRT to detection of unknown signals is the foundation for anomaly detection algorithms. In this section we describe the single hypothesis test, formulated based on the GLRT for unknown signals, assuming Gaussian noise.

Let $\mathbf{y}$ be a vector that is to be classified as arising from either pdf $p_0(\cdot, \theta_0)$ or $p_1(\cdot, \theta_1)$, (hypothesis H_0 or H_1 respectively). Let $\{\mathbf{k}_j \in \mathbb{C}^m | 1 \leq j \leq N\}$ be a set of N IID sample vectors having pdf $p_0(\cdot, \theta_0)$. The GLRT is as follows [73]:

$$L(\mathbf{y}) = \frac{\max\limits_{\theta_1} \left[p_1(\mathbf{y}, \theta_1) \, p_0(\{\mathbf{k}_j | 1 \leq j \leq N\}, \theta_1) \right]}{\max\limits_{\theta_0} \left[p_0(\mathbf{y}, \theta_0) \, p_0(\{\mathbf{k}_j | 1 \leq j \leq N\}, \theta_0) \right]} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{3.1}$$

where η is a threshold. Kelly [45] and Reed and Yu [62] developed a GLRT for multidimensional image data assuming the the target signal and the covariance of the background are unknown. Let $\mathcal{N}(\boldsymbol{\mu}, \Gamma)$ denote normal distribution with mean $\boldsymbol{\mu}$ and covariance Γ. Then the RX (Reed Xiaoli) GLRT between the target absent hypothesis H_0, and the target present hypothesis H_1:

$$\begin{aligned} H_0 : &\quad \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Gamma) \\ H_1 : &\quad \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_1, \Gamma) \end{aligned} \tag{3.2}$$

is as follows [62]:

$$RX(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu}_0)^{\mathrm{T}} \left(\frac{N}{N+1}\widehat{\Gamma} + \frac{(\mathbf{y}-\boldsymbol{\mu}_0)(\mathbf{y}-\boldsymbol{\mu}_0)^{\mathrm{T}}}{N+1} \right)^{-1} (\mathbf{y}-\boldsymbol{\mu}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (3.3)$$

where μ_1 and Γ are unknown and:

$$\widehat{\Gamma} = \frac{1}{N}\sum_{j=1}^{N} (\mathbf{k}_j - \boldsymbol{\mu}_0)(\mathbf{k}_j - \boldsymbol{\mu}_0)^{\mathrm{T}} \quad (3.4)$$

is the sample covariance matrix of the reference data. As $N \to \infty$ the RX GLRT converges to:

$$RX_{N\to\infty}(\mathbf{y}) = (\mathbf{y}-\boldsymbol{\mu}_0)^{\mathrm{T}} \left(\widehat{\Gamma}\right)^{-1} (\mathbf{y}-\boldsymbol{\mu}_0) \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (3.5)$$

which is the Mahalanobis distance [38] of vector $\mathbf{y}$ from $\boldsymbol{\mu}_0$ [53]. It is evident that the GLRT presented in (3.5) does not rely on any information about the targets but rather on an estimated model of the background. The test is to decide whether vector $\mathbf{y}$ belongs to the background and therefore is a single hypothesis test. Let $\chi^2_m(\lambda)$ denote the non-central chi-square distribution with m degrees of freedom and non-centrality parameter λ. Then the probability distributions of the GLRT RX, under H_0 and H_1 are given in terms of chi-square and noncentral chi-square densities as follows:

$$RX \sim \begin{cases} \chi^2_{\tau m}(0) & \text{, under } H_0 \\ \chi^2_{\tau m}\left((\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^{\mathrm{T}}\Gamma^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)\right) & \text{, under } H_1. \end{cases} \quad (3.6)$$

where τ=1,2 if the data are real or complex valued, respectively. Given (3.6) and

(3.6), the probabilities of false alarm and detection are:

$$P_{FA} = 1 - P\left[\chi^2_{\tau m}(0) \leq \eta\right] \tag{3.7}$$

$$P_D = 1 - P\left[\chi^2_{\tau m}\left((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^{\mathrm{T}}\Gamma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\right) \leq \eta\right]. \tag{3.8}$$

When the target is unknown but the subspaces in which the target and background interference signals lie are known, the GLRT can be extended to a more general form [54]. Scharf and Friedlander [69] formulated a MSD for the general problem of detecting subspace signals in subspace interference and additive white Gaussian noise. The MSD, based on the GLRT, is presented in Sec.3.3.

3.3 The Matched Subspace Detector

In this section we present the MSD, formulated by Scharf and Friedlander [69] for the general problem of detecting subspace signals in subspace interference and additive white Gaussian noise.

Let $\langle H \rangle$ denote the signal subspace, spanned by the columns of matrix H and let $\langle S \rangle$ denote the interference subspace, spanned by the columns of matrix S. We denote the additive IID Gaussian noise by $\mathbf{b}$ where $\mathbf{b} \sim \mathcal{N}(0, \rho^2 I)$. The problem is to determine whether the sample vector $\mathbf{y}$ contains a target signal. The target signal $\mathbf{x}$ can be described as a linear combination of the columns of H *i.e.*, $\mathbf{x} = H\boldsymbol{\psi}$, where $\boldsymbol{\psi}$ is a vector of coefficients. The interference signal is described similarly, using the matrix S and the coefficients vector $\boldsymbol{\phi}$. Considering the detection problem, we define two hypotheses, H_0 and H_1 which indicate,

respectively, absence and presence of target signal in the vector $\mathbf{y}$:

$$\begin{array}{l} H_0 : \mathbf{y} = S\boldsymbol{\phi} + \mathbf{b} \\ H_1 : \mathbf{y} = H\boldsymbol{\psi} + S\boldsymbol{\phi} + \mathbf{b}. \end{array} \tag{3.9}$$

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

$$P_S\mathbf{y}(\mathbf{s}) = S(S^{\mathrm{T}}S)^{-1}S^{\mathrm{T}}\mathbf{y}(\mathbf{s}) \tag{3.10}$$

and let P_{HS} denote the projection of a vector onto the subspace $\langle HS \rangle$, spanned by the columns of the concatenated matrix $\begin{bmatrix} H & S \end{bmatrix}$. The maximum likelihood estimates of the additive noise vector, $\mathbf{b}$, under H_0 and under H_1 are denoted by $\widehat{\mathbf{b}}_{H_0}$ and $\widehat{\mathbf{b}}_{H_1}$, respectively. These estimates are obtained by subtracting from $\mathbf{y}$ the components which lie in the signal and interference subspaces as follows:

$$\begin{array}{l} \widehat{\mathbf{b}}_{H_0} = (I - P_S)\mathbf{y} \\ \widehat{\mathbf{b}}_{H_1} = (I - P_{HS})\mathbf{y} \end{array} \tag{3.11}$$

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

$$\begin{aligned} L(\mathbf{s}) &= 2ln\left[\frac{Pr(\mathbf{b}(\mathbf{s})|H_0)}{Pr(\mathbf{b}(\mathbf{s})|H_1)}\right] = 2ln\left[\frac{\exp\left(\frac{\left[\widehat{\mathbf{b}}_{H_0}(\mathbf{s})\right]^2}{2\rho^2}\right)}{\exp\left(\frac{\left[\widehat{\mathbf{b}}_{H_1}(\mathbf{s})\right]^2}{2\rho^2}\right)}\right] \\ &= \frac{1}{\rho^2}\left[\left\|\widehat{\mathbf{b}}_{H_0}(\mathbf{s})\right\|_2^2 - \left\|\widehat{\mathbf{b}}_{H_1}(\mathbf{s})\right\|_2^2\right] \\ &= \frac{1}{\rho^2}\mathbf{y}^{\mathrm{T}}(P_{HS} - P_S)\mathbf{y}. \end{aligned} \tag{3.12}$$

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, as follows:

$$SNR = \frac{1}{\rho^2}\mathbf{x}^{\mathrm{T}}[I - P_{\widetilde{S}}]\mathbf{x}. \tag{3.13}$$

Let u denote the rank of the signal subspace. L is a sum of squared independent normally distributed variables and therefore is chi-square distributed with u degrees of freedom, as follows:

$$L \sim \begin{cases} \chi^2_u(0) & \text{, under } H_0 \\ \chi^2_u(SNR) & \text{, under } H_1. \end{cases} \tag{3.14}$$

Under hypothesis H_1, the non-centrality parameter of the chi-square distribution of L is equal to the SNR [69]. The decision rule is based on thresholding the log-likelihood ratio using the threshold η as follows:

$$\xi = \begin{cases} H_0 & \text{if} \quad L \leq \eta \\ H_1 & \text{if} \quad L > \eta. \end{cases} \tag{3.15}$$

Given (3.14) and (3.15), the probabilities of false-alarm and detection are:

$$P_{FA} = 1 - P\left[\chi^2_u(0) \leq \eta\right] \tag{3.16}$$

$$P_D = 1 - P\left[\chi^2_u(SNR) \leq \eta\right]. \tag{3.17}$$

3.4 Summary

We reviewed anomaly detectors for problems where no information about the targets is known, and where some *a priori* information is available. We presented the single hypothesis test, based on a statistical model of the background and no information about the targets. We described the MSD as formulated by Scharf

and Friedlander [69], and presented its performance with respect to the SNR of the examined scene. We assume that using *a priori* information about the targets in the detection process, may improve the detection performance. In chapter 4 we propose an iterative anomaly detection method which is based on a single hypothesis test. Any information we have about the targets is used in the post-processing in order to decide whether an anomalous region in the image is part of a target. In chapter 6 we propose a novel detection method, based on the MSD. Any *a priori* information about the targets is used for defining the targets subspace used by this detector.

Chapter 4

Iterative Anomaly Detection

4.1 Introduction

In this chapter, we propose an anomaly detection approach, which does not rely on an exhaustive statistical model of the targets, but rather on the local statistics of the data and possibly on some a priori information regarding the sizes and shapes of targets. An iterative procedure of feature extraction, based on local statistics and principle components analysis (PCA) is performed. The background is statistically characterized in a feature space of principle components. A single hypothesis scheme is used for the detection of anomalous pixels in a given region of interest (ROI). Subsequently, morphological operators are employed for extracting the sizes and shapes of anomalous clusters in the image domain, and identifying potential targets. We may compromise on the false alarm rate in order to achieve a high probability of detection, since each iteration gradually reduces the false alarm rate while maintaining the high probability of detection.

The chapter is organized as follows. In Sec. 4.2 we present a formulation of the problem. In Sec. 4.3 we provide a mathematical model. In Sec. 4.4 we

describe the proposed iterative anomaly detection algorithm. Finally, in Sec. 4.5 we demonstrate the application of the algorithm to sea-mine detection in sonar imagery.

4.2 Problem Formulation

Let Ω be the support of a gray scale image. The image may contain different textures as background (i.e. grass, trees, soil, water) and some targets which are anomalous with regards to the background. We would like to find a disjoint partition of Ω, such that $\Omega = \mathcal{B} \bigcup \mathcal{A}$, where $\mathcal{B}$ contains a few uniform subsets of background pixel clusters: $\mathcal{B} = \bigcup_{j=1}^{J} \mathcal{B}_j$, and $\mathcal{A}$ contains anomalous pixels. The subsets of the background pixels represent different textures of the background. For example, in a typical terrain scene, $\mathcal{B}_1$ may contain pixels which constitute grass, $\mathcal{B}_2$ may represent trees, and so on.

The proposed algorithm is based on an iterative two-category classification procedure. A set of pixels, $\mathcal{A}_{j-1}$, which have been identified as anomalous in the $(j-1)$'th iteration, is partitioned into background and anomalous subsets, $\mathcal{B}_j$ and $\mathcal{A}_j$ respectively:

$$\mathcal{A}_{j-1} = \mathcal{B}_j \bigcup \mathcal{A}_j \tag{4.1}$$

where $j \geq 1$ and $\mathcal{A}_0$ is initialized to the ROI in the image (possibly $\mathcal{A}_0 = \Omega$).

The fundamental problem of the iterative anomaly detection procedure is to partition the set of pixels, $\mathcal{A}_{j-1}$, into two subsets: One, which is relatively uniform, large and thus considered as the background subset ($\mathcal{B}_j$), and a second subset

which contains anomalous pixels ($\mathcal{A}_j$). Another problem is to determine a stopping rule for the iterative procedure. The procedure should be iterated until the number of targets pertaining to $\mathcal{A}_j$ does not exceed a given number. The maximal number of detectable targets in a given ROI is closely related to the permissible false alarm rate. A stopping rule based on such a criterion should incorporate into the procedure a low complexity routine for coarse target detection.

4.3 Mathematical Model

Let $\mathbf{s} \in \Omega$ be the coordinates of a pixel in the image. Let $\mathbf{q}(\mathbf{s})$ be a feature vector related to the pixel $\mathbf{s}$. Considering a two-category classification problem, we define two possible hypotheses:

$$H_0 : \mathbf{s} \in \mathcal{B} \qquad \text{(background pixel)}$$

$$H_1 : \mathbf{s} \in \mathcal{A} \qquad \text{(anomalous pixel)}$$

A two-category classification problem is often worked out using the Bayes decision rule for minimum cost [1]. Let $C\left(H_1|H_0\right)$ and $C\left(H_0|H_1\right)$ denote, respectively, the costs of false detection and miss detection. Then the optimal minimum cost decision rule is given by [33]:

$$f\left(\mathbf{q}(\mathbf{s})|H_0\right) P\left(H_0\right) C\left(H_1|H_0\right) \underset{H_1}{\overset{H_0}{\gtrless}} f\left(\mathbf{q}(\mathbf{s})|H_1\right) P\left(H_1\right) C\left(H_0|H_1\right) \tag{4.2}$$

where $f\left(\mathbf{q}(\mathbf{s})|H_i\right)$ is the conditional pdf of $\mathbf{q}(\mathbf{s})$ given H_i and $P(H_i)$ is the a priori probability of H_i.

In practice, the background can be well characterized (an empirical $f(\mathbf{q}(\mathbf{s})|H_0)$ can be generated) while the anomalies are of a wide variety and rare (a reliable estimate of $f(\mathbf{q}(\mathbf{s})|H_1)$ is unavailable). Therefore, the Bayes decision rule for minimum cost is inapplicable. A suitable alternative to such problems is based on a single hypothesis scheme [33]. Let $\kappa = E[\mathbf{q}(\mathbf{s})|H_0]$ denote the expected feature vector and $\Sigma = E\left[(\mathbf{q}(\mathbf{s})-\kappa)(\mathbf{q}(\mathbf{s})-\kappa)^T \middle| H_0\right]$ the covariance matrix under H_0 hypothesis. Let the normalized distance of $\mathbf{q}(\mathbf{s})$ from its expected vector, κ, be defined by:

$$d(\mathbf{q}(\mathbf{s})) = (\mathbf{q}(\mathbf{s})-\kappa)^T \Sigma^{-1} (\mathbf{q}(\mathbf{s})-\kappa) \tag{4.3}$$

Then the decision rule is given by:

$$d(\mathbf{q}(\mathbf{s})) \underset{H_1}{\overset{H_0}{\gtrless}} D \tag{4.4}$$

where D is a threshold to determine whether a given pixel is anomalous or not. This decision rule is based on the statistics of the background only. No knowledge about the anomalies statistics is taken into consideration. The threshold, D, can be determined according to a specified confidence level, φ, which is the probability of correctly deciding on H_0 given H_0 is true. The threshold, D, and the confidence level, φ, are related by:

$$\varphi \equiv \Pr(H_0|H_0) = \Pr(d(\mathbf{q}(\mathbf{s})) \leq D|H_0). \tag{4.5}$$

In case the feature vector, $\mathbf{q}(\mathbf{s})$, is a Gaussian random vector of dimension $n_{\mathbf{q}}$, the pdf of $d^2(\mathbf{q}(\mathbf{s}))$ under the H_0 hypothesis, denoted by $p_{d^2}(\zeta)$, is the gamma density function with parameters $\beta = \frac{n_{\mathbf{q}}}{2} - 1$ and $\alpha = 1/2$ [33]. Accordingly, the

relation between φ and D can be written as

$$\varphi = \int_0^{D^2} p_{d^2}(\zeta) d\zeta = \int_0^{D^2} \frac{1}{2^{n_{\mathbf{q}}/2}\Gamma(n_{\mathbf{q}}/2)} \zeta^{\frac{n_{\mathbf{q}}-2}{2}} e^{-\zeta/2} d\zeta \,. \tag{4.6}$$

The aforementioned decision rule is based on a one-dimensional measure, $d(\mathbf{q}(\mathbf{s}))$, whereas the feature vector, $\mathbf{q}(\mathbf{s})$ is multi-dimensional (size $n_{\mathbf{q}}$). It is quite clear that reducing the dimension of the feature space to a one dimensional distance space eliminates valuable information. Accordingly the false detection rate is likely to increase. However high false detection rate is acceptable, since we consider an iterative procedure, and each iteration gradually reduces the false alarm rate while maintaining a high probability of detection.

4.4 Anomaly Detection Algorithm

A block diagram of the proposed algorithm is presented in Fig. 4.1. The algorithm consists of two iterative procedures. The basic procedure is an iterative disjoint partition of $\mathcal{A}_{j-1}$ into $\mathcal{A}_j$ and $\mathcal{B}_j$. After the partitioning, the decision whether to further partition $\mathcal{A}_j$ depends on the number of potential targets found in $\mathcal{A}_j$. This procedure iterates, reducing $\mathcal{A}_j$'s population ($\mathcal{A}_j \subset \mathcal{A}_{j-1}$) until the number of potential targets, found in $\mathcal{A}_j$, is smaller than or equal to a given number, Q. This procedure is an implementation of a rejection based classification approach. This approach may be computationally efficient due to the rejection of most of the pixels population in the early stages of the process (the first iterations). The second procedure is implicit in the partition of $\mathcal{A}_{j-1}$ into $\mathcal{A}_j$ and $\mathcal{B}_j$. This step itself is carried out in an iterative manner. A first guess of background subset,

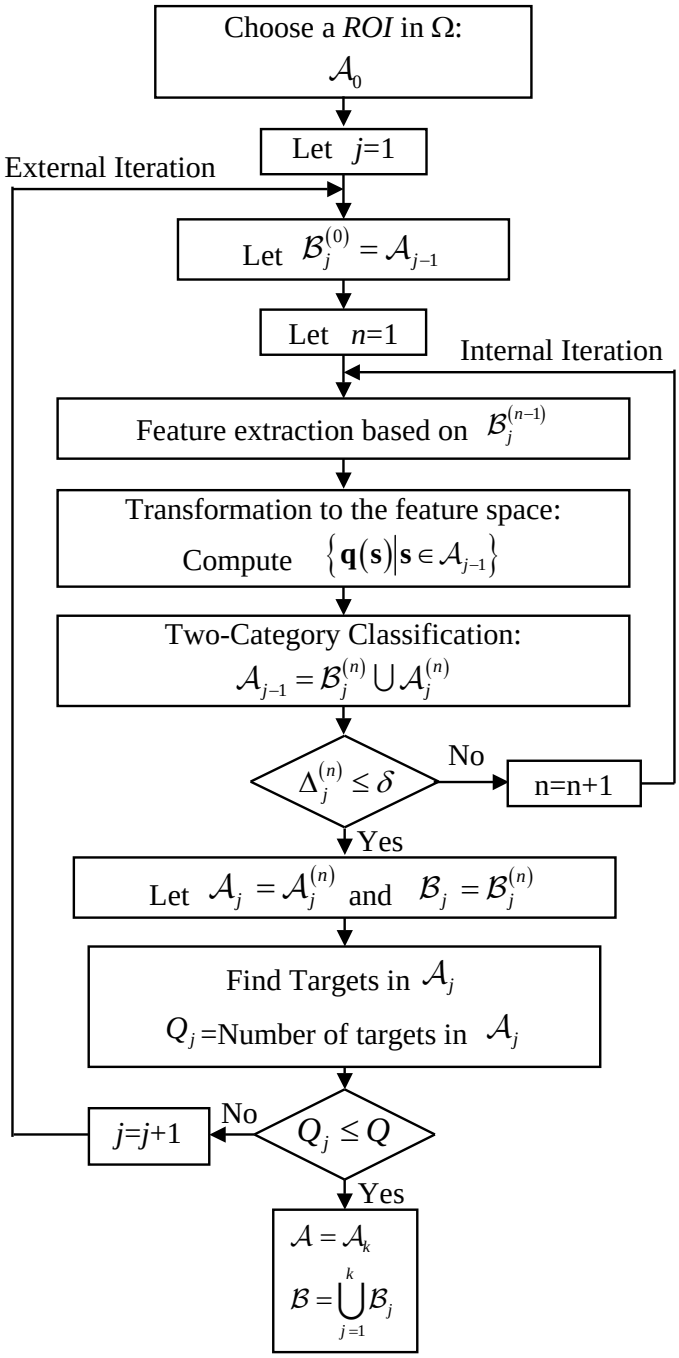


Figure 4.1: Block diagram of the iterative anomaly detection algorithm.

denoted by $\mathcal{B}_j^{(0)}$ is chosen in $\mathcal{A}_{j-1}$. For each pixel in $\mathcal{B}_j^{(0)}$, an observation vector, $\mathbf{y}_o(\mathbf{s})$, is produced (containing the values of the pixels in the neighborhood of the pixel $\mathbf{s}$). The most dominant principle components of $\{\mathbf{y}_o(\mathbf{s}) | \mathbf{s} \in \mathcal{B}_j^{(0)}\}$ are chosen

to be the features which span the feature space. A disjoint partition of $\mathcal{A}_{j-1}$ into $\mathcal{A}_j^{(1)}$ and $\mathcal{B}_j^{(1)}$ is then carried out using the single hypothesis scheme, as described in Section 3. The next step includes an update of the feature space using the first most dominant principle components of $\{\mathbf{y}_o(\mathbf{s})|\mathbf{s} \in \mathcal{B}_j^{(1)}\}$ and disjointly partitioning $\mathcal{A}_{j-1}$ into $\mathcal{A}_j^{(2)}$ and $\mathcal{B}_j^{(2)}$ in the new feature space. This procedure iterates, with partitioning $\mathcal{A}_{j-1}$ into $\mathcal{A}_j^{(n)}$ and $\mathcal{B}_j^{(n)}$ where the feature space is obtained at each iteration by PCA done on the background subset $\mathcal{B}_j^{(n-1)}$. The stopping rule for this iterative procedure relies on the relative change, $\Delta_j^{(n)}$, between $\mathcal{B}_j^{(n-1)}$ and $\mathcal{B}_j^{(n)}$. Specifically, we compute the relative change defined by:

$$\Delta_j^{(n)} = \frac{\left|\left(\mathcal{B}_j^{(n)} \cup \mathcal{B}_j^{(n-1)}\right) \setminus \left(\mathcal{B}_j^{(n)} \cap \mathcal{B}_j^{(n-1)}\right)\right|}{\left|\mathcal{B}_j^{(n-1)}\right|} \tag{4.7}$$

where $|\mathcal{B}_j^{(n)}|$ is the number of pixels in the subset $\mathcal{B}_j^{(n)}$. Then we compare the relative change to a certain threshold, δ, and stop the iterations when $\Delta_j^{(n)} \leq \delta$. The relative change is defined here as the ratio between the number of pixels in the exclusive-or set of $\mathcal{B}_j^{(n)}$ and $\mathcal{B}_j^{(n-1)}$, and the total number of pixels in $\mathcal{B}_j^{(n-1)}$. Small relative change indicates an insignificant difference between the present background subset and the previous one.

The diagram in Fig. 4.1 shows also an external loop for partitioning the anomalous subsets iteratively. Its stopping rule is based on the number of the potential targets in the anomalous subset. Specifically, the number of potential targets in $\mathcal{A}_j$ should be smaller than or equal to the presumed maximal number of possible targets in the given ROI. The number of potential targets is determined without

much *a priori* knowledge about the targets, by clustering the pixels in the spatial domain and identifying a cluster as target based on the size and shape of the cluster. Morphological tools are used for finding connected sets with areas which fit the presumed size. These are classified as potential targets and counted as such. The number of potential targets, Q_j, in the anomalous subset, is then compared to the presumed maximum number of targets, Q. The external iterative procedure proceeds until $Q_j \leq Q$.

4.5 Examples

Sea-mines detection in shallow water involves addressing the varying shape of the ocean surface and vegetation [26] which yields large variability in background clutter. Conventional methods using a preprocessing procedure and matched filters [53] for detection of mine-size regions, that closely match a typical mine signature, presume *a priori* knowledge of the mine's size and shape. Normally, mines in a sonar image include a highlight region, representing the mine's body and a dark region representing its shadow. The variability in the background clutter and mine appearance in the sonar image, leads to a high false alarm rate in conventional methods. Fig. 4.2(a) presents three sonar images with sea-mines (one in each image). These images include different clutter patterns. The left image in Fig. 4.2(a) contains a clutter pattern of highlights and shadows whereas the particular mine in this image contains no highlight. The clutter in the central image has a different pattern containing mine-like shaped blobs. The image on the right has a large flat highlight region in its left side. Conventional methods fail in detecting sea-mines in such varying environments. Furthermore, such clutters may produce a large number of false alarms. An experienced human viewer would

be able to detect the mines based on the difference from the background clutter. The proposed algorithm mimics the detection mechanism of the human viewer by detecting anomalies in the background clutter.

The application of the proposed algorithm to the three sonar images is demonstrated in Fig. 4.2. The algorithm employs the 3 most dominant principle components of 5×5 pixel neighborhoods. This neighborhood size characterizes the information which differentiates between the background texture and mines. The relative change threshold was set to 1%, the confidence level was set to 95%, the presumed maximum number of targets was set to 2. The coarse target detection procedure was based on spatial morphologic detection of anomalous clusters composed of 15 pixels or more. Fig. 4.2(b) show the result of the first external iteration, containing several potential targets (connected pixel groups of at least 15 pixels). The second iteration, as shown in Fig. 4.2(c), significantly reduces the number of potential targets. These potential targets are presented in Fig. 4.2(d) after the morphological filtering.

4.6 Summary

The proposed algorithm includes an iterative procedure of feature extraction, based on PCA of the spatial information around each pixel. The background is statistically characterized in the feature space, and a single hypothesis scheme is used for the detection of anomalous pixels. This method was successfully employed on a large data-set of sea-mines sonar images.

The method proposed in this chapter can be extended to hyper-spectral imagery [73] by incorporating the additional spectral information, related to each pixel,

into the feature vectors. Rather than considering each pixel spectra separately, the spatial information is combined with the spectral data to improve the anomaly detection performance. Multi-resolution representations [31] may as well improve the detection performance, particularly in cases where the background textures are nonuniform, and in cases where no *a priori* information regarding the sizes and shapes of targets is available.

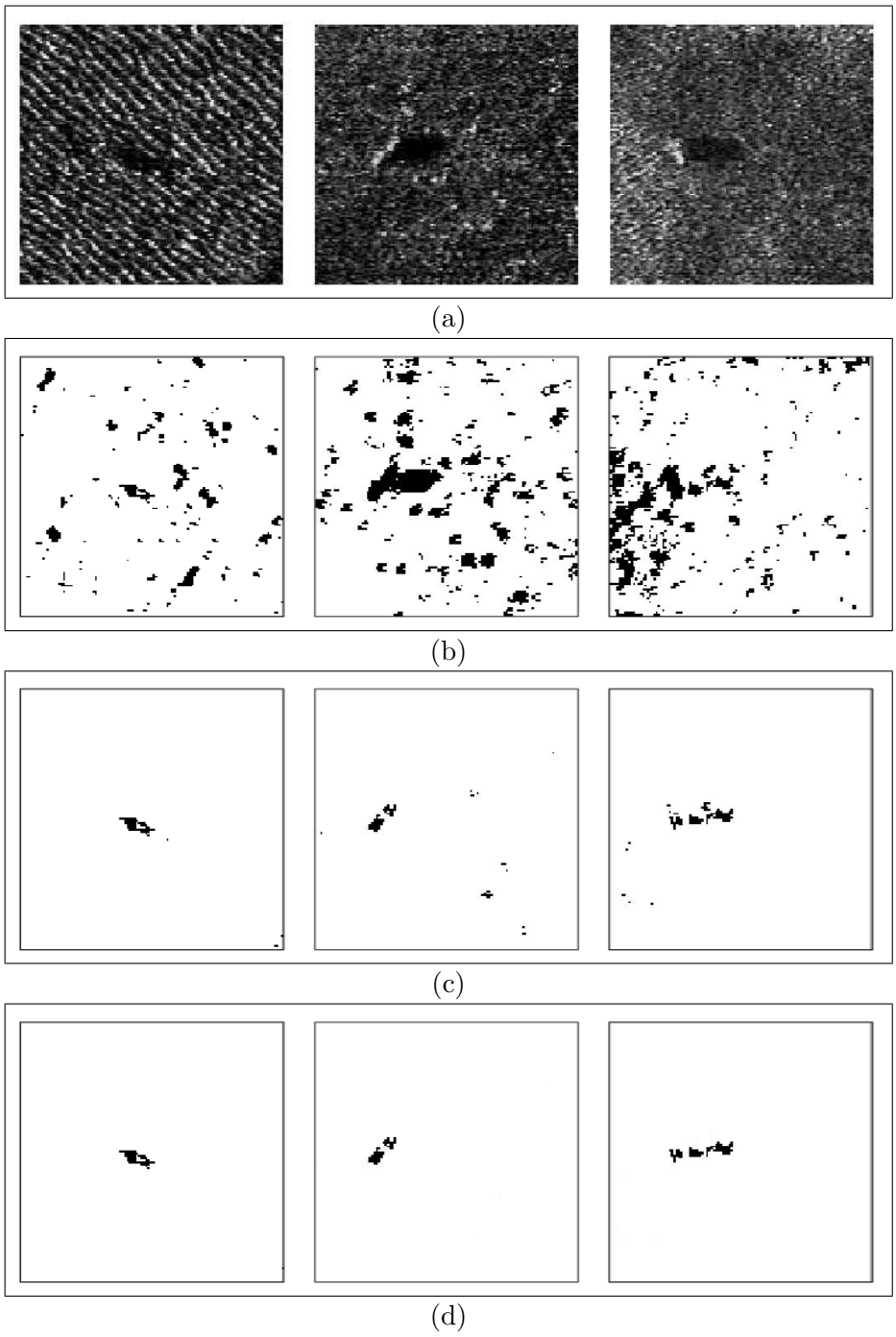


Figure 4.2: (a) Original three sonar images containing sea-mines; (b) The corresponding images of first iteration anomalies, $\mathcal{A}_1$ (black pixels); (c) The second iteration anomalies, $\mathcal{A}_2$; (d) The result of a morphological filtering for coarse target detection

Chapter 5

Multi-Scale GMRF Models

5.1 Introduction

Clutter models as the GMRF, SAR, and GLC model have proven to be effective in describing natural textures with periodical as well as random elements [4], [29]. Many physical processes (e.g. atmospheric or oceanographic phenomena) possess behavior over vast ranges of spatial or spatio-temporal scales [6], [30], [47], [50], [83]. In many natural clutter images, scene elements often appear to have several periodical patterns, of various period lengths. In such cases, random field models as the GMRF and the LC models may not sufficiently fit the clutter image. Deviations of the clutter image from the random field model influence the detection performance by increasing the false alarm rate.

In this chapter we introduce multi-scale GMRF models. The multi-scale models are based on a multi-scale representation of the image to be modelled. This representation is composed of a set of images, arranged in a pyramidal or a multi-layer structure, representing the original image in different scales. We assume each layer of this structure can be modelled as a GMRF.

A pyramidal representation of an image is based on a set of images, generated from a single image by reducing its resolution. Burt and Adelson [11] presented the Gaussian and Laplacian pyramidal representations and proposed a technique for image encoding based on the Laplacian pyramid. Figure 5.1 presents the process of generating a Gaussian pyramid $\mathcal{P}_G(\ell)$ and a Laplacian pyramid $\mathcal{P}_L(\ell)$. The Gaussian pyramid is generated by repeatedly using a reduction operation for reducing the resolution of the image. Each layer ℓ of the Gaussian pyramid is generated from the previous layer by applying a Gaussian low pass filter (LPF) and down-sampling the image by choosing every other pixel in each row and every other row. The Laplacian pyramid is generated based on the Gaussian pyramid and an expansion operation. The expansion operation enlarges the image by doubling its size (in pixels) in each axis, using interpolation. Reducing and then expanding an image preserve its size but eliminate up to half of the spatial frequencies it contains.

The pyramidal representation is generated by repeatedly applying a LPF to the original image and resampling the resulting image. Here, the multi-scale representation is generated by filtering the original image with a certain set of filters. The KLT is applied to the resulting multi-layer image for generating independent layers. We assume that the independent layers can be modeled as GMRFs with different sets of parameters.

This chapter is organized as follows: In section 5.2 we introduce the multi-scale GMRF models. We propose three different models based on the pyramidal representation and on a multi-layer representation generated using a set of scaling filters. In section 5.3 we propose an estimation scheme for the multi-scale GMRF models.

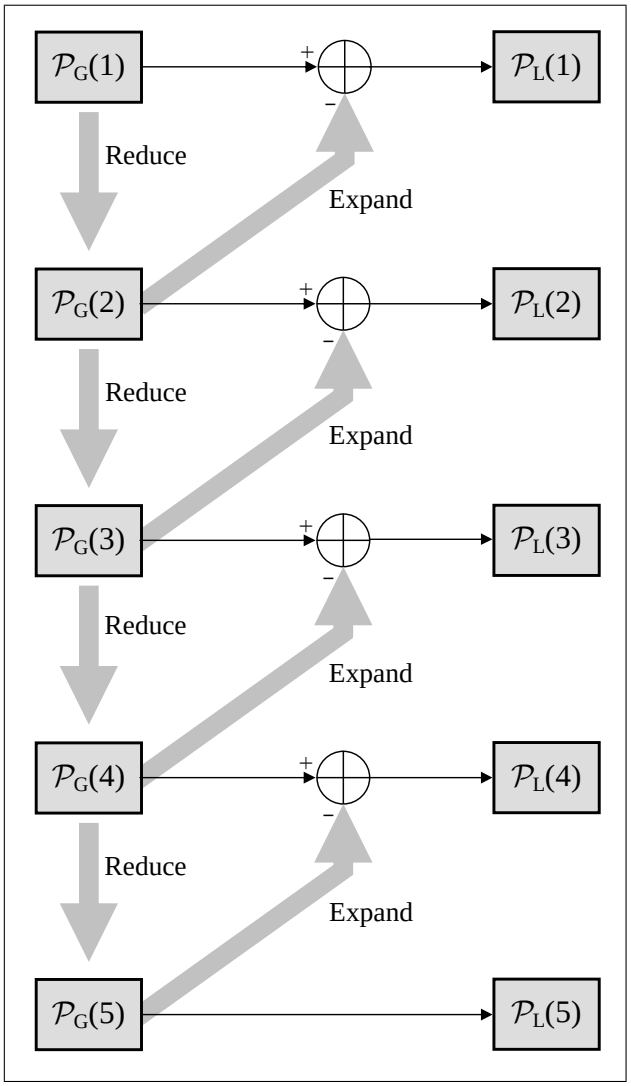


Figure 5.1: The process of generating the Gaussian and Laplacian image pyramids (Adopted from Elad [27]).

5.2 Statistical Models

The statistical models proposed in this section, are based on modeling each layer of the multi-scale representation of an image using a GMRF. We preferred it over the more general GLC model due to its simplicity and suitability to a wide range of natural textures. Another reason for using the GMRF model is its simple and practical estimation methods.

We propose 3 different models, two of which are based on the pyramidal representation.

5.2.1 Pyramidal GMRF models

The pyramidal representation of an image can be used for modeling its multi-scale characteristics. The different layers of the pyramidal representation contain information in different spatial frequency bands. In the Laplacian pyramid, for example, the spatial spectrum of the original image is divided into separate spectral bands contained in separate layers of the pyramidal structure. We assume each layer of the Laplacian pyramid can be modeled as a GMRF with a different set of parameters. In this section we propose two possible multi-scale GMRF models, based on the Laplacian pyramidal representation.

Pyramidal GMRF Model I

Figure 5.2 presents *Pyramidal GMRF Model I*. Each layer ℓ of the Laplacian pyramid $\mathcal{P}_L(\ell)$ is expanded (interpolated) to the size of the finest scale layer $\mathcal{P}_L(1)$ and then modeled as a GMRF with a different set of parameters. The innovations process of each expanded layer is estimated. Eche pixel in the original image can now be represented by a feature vector which contains the innovations, estimated for the expanded versions of the different pyramid layers.

Let $\mathbf{T}(\mathbf{s})$ denote the multi-layer image generated by expanding each layer of the pyramid $\mathcal{P}_L(\cdot)$ to the size of $\mathcal{P}_L(1)$ and let $\mathbf{T}_\ell(\mathbf{s})$ denote the ℓ-th layer of $\mathbf{T}(\mathbf{s})$. We denote the weight coefficient estimated for neighbor $\mathbf{r} \in \mathcal{R}$, and for the ℓ-th layer of $\mathbf{T}(\mathbf{s})$ by $\theta_\ell(\mathbf{r})$, and the innovations process of the ℓ-th layer by $\varepsilon_\ell(\mathbf{s})$. $\mathbf{T}(\mathbf{s})$

is then given by the following equation:

$$\mathbf{T}(\mathbf{s}) = \sum_{\mathbf{r}\in\mathcal{R}} \Theta_{\mathbf{r}} \mathbf{T}\left(\mathbf{s}+\mathbf{r}\right) + \boldsymbol{\varepsilon}\left(\mathbf{s}\right) \tag{5.1}$$

where Θ_r is the following diagonal matrix:

$$\Theta_r = \mathrm{diag}(\theta_1\left(\mathbf{r}\right), \theta_2\left(\mathbf{r}\right), ..., \theta_p\left(\mathbf{r}\right)) \tag{5.2}$$

and $\boldsymbol{\varepsilon}\left(\mathbf{s}\right)$ is a vector of the innovations in pixel $\mathbf{s}$ in the different layers of $\mathbf{T}\left(\mathbf{s}\right)$:

$$\boldsymbol{\varepsilon}\left(\mathbf{s}\right) = \left[\varepsilon_1\left(\mathbf{s}\right), \varepsilon_2\left(\mathbf{s}\right), ..., \varepsilon_p\left(\mathbf{s}\right)\right]^{\mathrm{T}}. \tag{5.3}$$

The estimation of the model parameters for each layer, is carried out using the methods described in Sec. 5.3. Subsequently, we can estimate the innovations process (prediction error), $\widehat{\boldsymbol{\varepsilon}}\left(\mathbf{s}\right)$, by:

$$\widehat{\boldsymbol{\varepsilon}}\left(\mathbf{s}\right) = \mathbf{T}(\mathbf{s}) - \sum_{\mathbf{r}\in\mathcal{R}} \widehat{\Theta_r} \mathbf{T}\left(\mathbf{s}+\mathbf{r}\right). \tag{5.4}$$

Pyramidal GMRF Model II

Figure 5.3 presents *Pyramidal GMRF Model II*. Each layer of the Laplacian pyramid is modeled as a GMRF with a different set of parameters. The innovations process of each layer is estimated and expanded (interpolated) to the size of the finest scale layer. Eche pixel in the original image can now be represented by a feature vector which contains the innovations $\widehat{\boldsymbol{\varepsilon}}\left(\mathbf{s}\right)$, estimated for the different pyramid layers.

The difference between the pyramidal models proposed here, is that in *Pyramidal GMRF model I*, each layer ℓ of the Laplacian pyramid $\mathcal{P}_L(\ell)$ is first expanded

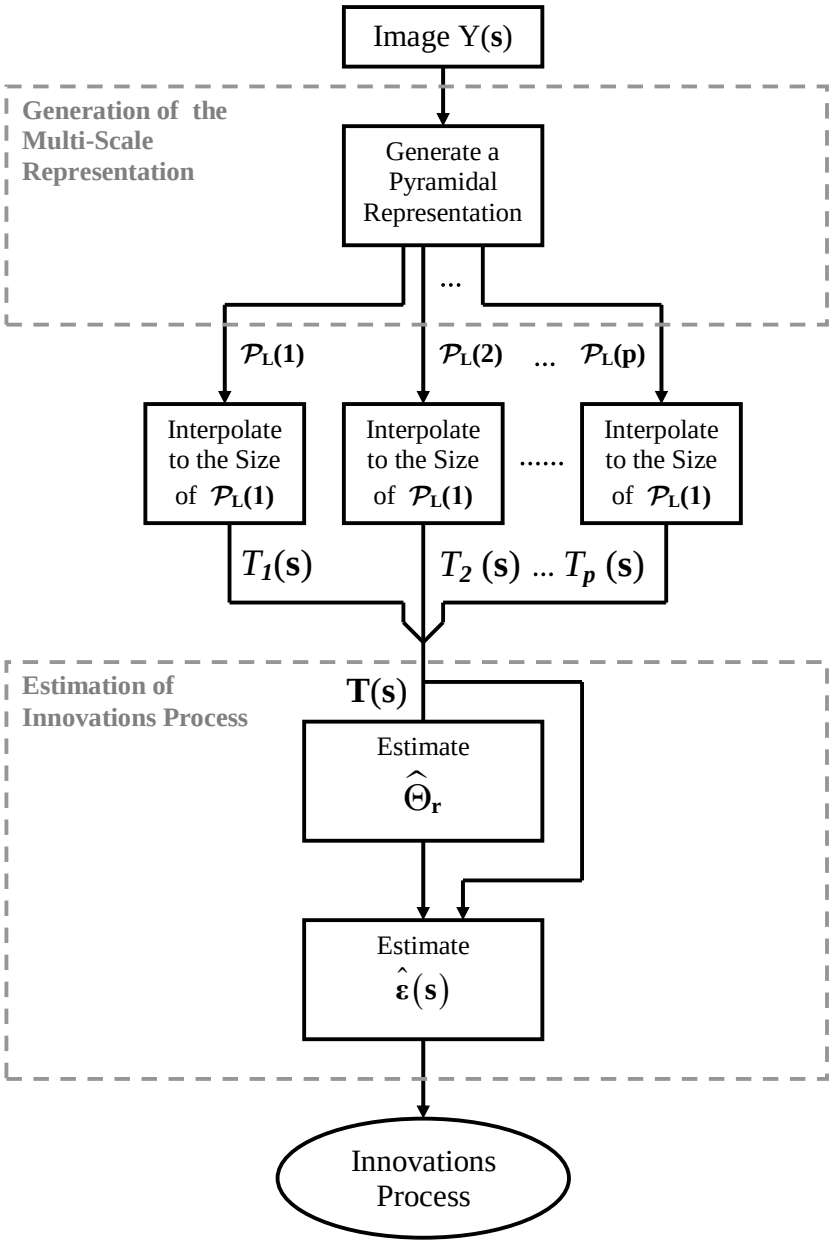


Figure 5.2: Modeling of an image according to pyramidal GMRF model I.

(interpolated) to the size of the finest scale layer $\mathcal{P}_L(1)$, and then modeled as a GMRF. In *Pyramidal GMRF model II*, each layer ℓ is first modeled as a GMRF, and only the innovations process images are expanded to the size of the finest scale layer.

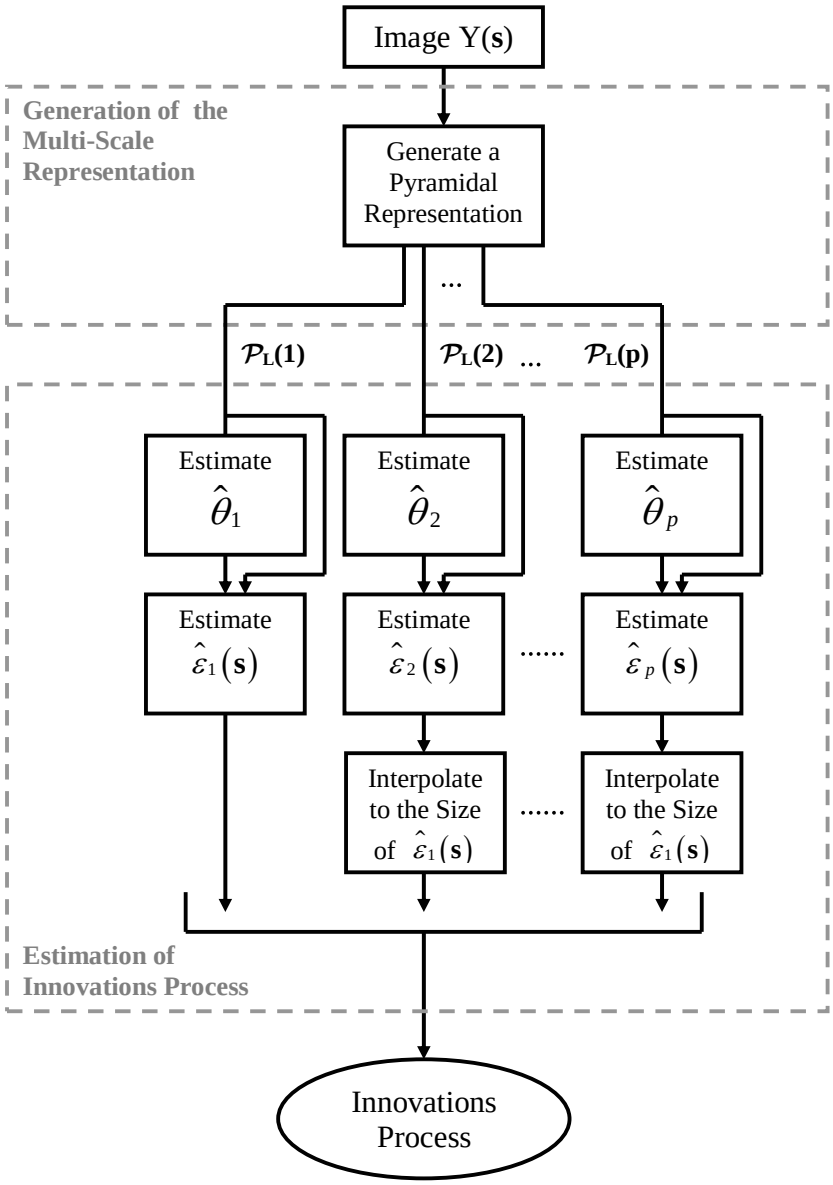


Figure 5.3: Modeling of an image according to pyramidal GMRF model II.

5.2.2 Multi-Layer GMRF model

The pyramidal representation is generated by repeatedly applying a LPF to the original image and resampling the resulting image. This approach dictates a constant scaling factor (decreasing the resolution between each following layers by 2). In order to release this constraint, we propose an alternative method for gen-

erating the multi-scale representation. This approach is based on a set of scaling filters which extend over different support sizes. The multi-scale representation is generated by filtering the original image using this set of filters. The KLT is applied to the resulting multi-layer image for generating independent layers. We assume that the independent layers can be modeled as GMRFs with different sets of parameters.

Let $Y(\mathbf{s})$ denote an image, and let $\mathcal{G} = \{G_1, G_2, ..., G_n\}$ denote a given set of multi-scale spatially invariant filters (*e.g.* scaling and wavelet filters). We generate from the image a multi-scale image, $\mathbf{Y}$, by applying the filters to the image Y and concatenating the results in the third dimension:

$$Y_i = Y * G_i \ , \quad i = 1, ..., n \tag{5.5}$$

$$\mathbf{Y}(\mathbf{s}) = [Y_1(\mathbf{s}), Y_2(\mathbf{s}), ..., Y_n(\mathbf{s})] \tag{5.6}$$

where $*$ denotes 2-dimensional convolution. The result $\mathbf{Y}$ is a 3-dimensional representation of the image, thus each pixel is now transformed to a vector. The KLT can be applied to $\mathbf{Y}(\mathbf{s})$, for generating a multi-scale image, $\mathbf{T}(\mathbf{s})$, with independent layers. $\mathbf{T}(\mathbf{s})$ has p layers representing the top p independent components of $\mathbf{Y}(\mathbf{s})$. Let K denote a matrix whose columns are the top p eigen vectors of the covariance matrix of $\mathbf{Y}(\mathbf{s})$. $\mathbf{T}(\mathbf{s})$ is then given by:

$$\mathbf{T}(\mathbf{s}) = K^{\mathrm{T}}\mathbf{Y}(\mathbf{s}). \tag{5.7}$$

We assume that there is a set of filters, $\mathcal{G}$, such that each image layer, $\mathbf{T}_\ell(\mathbf{s})$, can be modeled as a GMRF with a different set of parameters. We denote the weight coefficient estimated for neighbor $\mathbf{r} \in \mathcal{R}$, and for the ℓ-th layer of $\mathbf{T}(\mathbf{s})$ by $\theta_\ell(\mathbf{r})$, and the innovations process of the ℓ-th layer by $\varepsilon_\ell(\mathbf{s})$. $\mathbf{T}(\mathbf{s})$ is then given by (5.1) and (5.2). The innovations process $\boldsymbol{\varepsilon}(\mathbf{s})$ is estimated using (5.4).

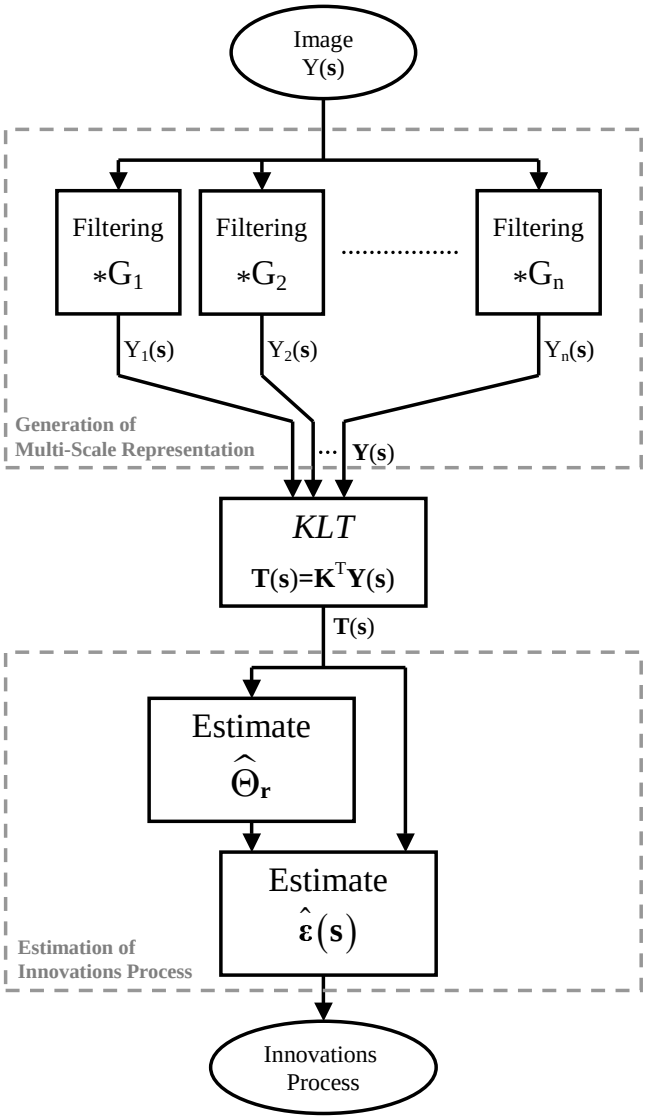


Figure 5.4: Multi-scale GMRF modeling of an image using the multi-layer representation.

Figure 5.4 presents the multi-scale GMRF modeling of an image using the multi-layer representation. The modeling process is divided to 3 stages: generation of the multi-scale representation, KLT, and estimation of the innovations process. The GMRF model is a special case of the multi-scale GMRF model, in which the set of filters $\mathcal{G}$, contain only the identity filter (a filter which causes no change to the image). In this case, the "multi-scale" representation contains the image

in one scale only and the KLT doesn't affect it. Finally, Θ_r becomes a matrix of 1×1 and thus, $\Theta_r = \theta_r$. For this special case, (5.1) reduces to the equation of the known GMRF given in (2.1).

5.3 Model Estimation

The multi-scale GMRF models estimation is based on the estimation of the GMRF models which describe each layer in the pyramidal structure or each independent layer $\mathbf{T}_\ell(\mathbf{s})$ of the multi-layer image representation $\mathbf{T}(\mathbf{s})$. The problem of GMRF model estimation for image processing received some attention in image processing literature. Various methods for the GMRF model estimation were developed over the years, *e.g.* [7], [23], [35], [36], [41], [40], [59], [70], [71], [87].

Let $\mathbf{T}_\ell$ denote the ℓ-th layer of the multi-scale representation (pyramidal layer or an independent layer from the multi-layer representation). A computationally efficient method for the GMRF model estimation is the coding method, described in details in Kashyap and Chellappa [41] and Besag [7]. Let the column stack ordering of the neighborhood of $\mathbf{T}_\ell(\mathbf{s})$ be denoted by $\mathbf{g}_\ell(\mathbf{s})$:

$$\mathbf{g}_\ell(\mathbf{s}) = \text{vec}[\mathbf{T}_\ell(\mathbf{s}+\mathbf{r}), \mathbf{r} \in \mathcal{R}] \tag{5.8}$$

and let

$$\boldsymbol{\theta}_\ell = \text{vec}[\theta_\ell(\mathbf{r}), \mathbf{r} \in \mathcal{R}]. \tag{5.9}$$

Based on Kashyap and Chellappa [41] and using (2.10), $\boldsymbol{\theta}_\ell$ and ρ_ℓ^2 are estimated

as follows:

$$\widehat{\boldsymbol{\theta}}_\ell = \left[\sum_{\mathbf{s}\in\Omega_0} \mathbf{g}_\ell(\mathbf{s})\mathbf{g}_\ell(\mathbf{s})^{\mathrm{T}}\right]^{-1}\left[\sum_{\mathbf{s}\in\Omega_0} \mathbf{T}_\ell(\mathbf{s})\mathbf{g}_\ell(\mathbf{s})\right] \tag{5.10}$$

$$\widehat{\rho_\ell^2} = \frac{1}{|\Omega|}\sum_{\mathbf{s}\in\Omega}\left(\mathbf{T}_\ell(\mathbf{s}) - \widehat{\boldsymbol{\theta}}_\ell^{\mathrm{T}}\mathbf{g}_\ell(\mathbf{s})\right)^2 \tag{5.11}$$

where Ω_0 is a subset of the support Ω with every other site of Ω skipped. The estimate of $\widehat{\Theta}_r$ is then:

$$\widehat{\Theta}_r = \mathrm{diag}(\widehat{\theta}_1(\mathbf{r}), \widehat{\theta}_2(\mathbf{r}), ..., \widehat{\theta}_p(\mathbf{r})) \tag{5.12}$$

One of the main disadvantages of the coding method is that the estimates are not efficient [59] due to the partial use of the data. Another disadvantage is that the estimate is not unique. Summing $\mathbf{s}$ in (5.10) over $\Omega - \Omega_0$ instead of Ω_0 may produce considerably different estimates. Simple averaging of these dependent estimates is not satisfactory.

The maximum likelihood (ML) estimate of $\boldsymbol{\theta}_\ell$ is consistent and efficient but does not in general have a close form solution. The ML estimate can be obtained using numerical optimization techniques on $M_x \times M_y$ equations of the form of (2.1) (where $M_x \times M_y$ is the size of the image), constrained by the requirement that (2.7) be positive. This method is computationally inefficient, since it involves the optimization of a nonlinear function with respect to a matrix valued parameter, subject to nonlinear constraints.

Let Ω_B, the boundary subset of Ω, be defined as follows:

$$\Omega_B = \{\mathbf{s}|\mathbf{s}\in\Omega \text{ and } (\mathbf{s}+\mathbf{r})\notin\Omega \text{ for at least one } \mathbf{r}\in\mathcal{R}\} \tag{5.13}$$

and let Ω_I, the interior subset of Ω, be defined as follows:

$$\Omega_I = \Omega - \Omega_B \tag{5.14}$$

A computationally efficient and asymptotically consistent estimate of $\boldsymbol{\theta}_\ell$ for a finite image can be obtained by summing $\mathbf{s}$ in (5.10) over Ω_I. Equation (5.10) can also be used for estimation in toroidal lattice representations by summing $\mathbf{s}$ over Ω instead of Ω_I.

The asymptotic consistency of this estimator has been shown in Kashyap and Chellappa [41]. Since $\mathbf{T}_\ell(\mathbf{s})$ is assumed to follow the GMRF model, we replace $\mathbf{T}_\ell(\mathbf{s})$ in 5.10 with the expression:

$$\mathbf{T}_\ell(\mathbf{s}) = \sum_{\mathbf{r}\in\mathcal{R}} \theta(\mathbf{r})\mathbf{T}_\ell(\mathbf{s}+\mathbf{r}) + \varepsilon_\ell(\mathbf{s}). \tag{5.15}$$

and summarize over Ω_I instead of Ω_0. After simplifying we receive the following equation:

$$\left[\sum_{\mathbf{s}\in\Omega_0} \mathbf{g}_\ell(\mathbf{s})\mathbf{g}_\ell(\mathbf{s})^{\mathrm{T}}\right]\left(\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}_\ell\right) = \sum_{\mathbf{s}\in\Omega_0} \mathbf{g}_\ell(\mathbf{s})\varepsilon_\ell(\mathbf{s}) \tag{5.16}$$

Since $E\left(\mathbf{g}_\ell(\mathbf{s})\varepsilon_\ell(\mathbf{s})\right) = 0$ and $\sum_{\mathbf{s}\in\Omega_0} \mathbf{g}_\ell(\mathbf{s})\mathbf{g}_\ell(\mathbf{s})^{\mathrm{T}}$ is a positive definite matrix, the estimate $\widehat{\boldsymbol{\theta}}_\ell$ is consistent.

5.4 Summary

Many physical processes possess behavior over vast ranges of spatial scales. Scene elements often appear to have several periodical patterns, of various period lengths. In such cases, random field models as the SAR, the GMRF and the

LC models may not sufficiently fit the clutter image. In this chapter, we have introduced multi-scale GMRF models. The models are based on multi-scale representations of the image, ordered in a pyramidal or a multi-layer structure. We assume each layer of the Laplacian pyramid or the multi-layer representation can be modeled as a GMRF with a different set of parameters. The model estimation scheme is based on estimating p GMRF parameters sets. We have presented the concept of the computationally inefficient ML estimator and reviewed in more details the inconsistent coding method and a consistent and computationally efficient extension of this method.

In Chapter 6, we introduce an anomaly subspace detector using the multi-scale GMRF model. We analyze the performance of the proposed algorithm and compare the detection results to those obtained by using competing methods.

Chapter 6

Anomaly Subspace Detection

6.1 Introduction

Anomaly detection methods do not rely on an exhaustive statistical model of the targets, but rather on a statistical model of the background. When *a priori* information about the targets is available, incorporating it into the detection process may improve detection performance. In most real detection problems, the information about the targets is partial. Partial information about the targets may help to define a subspace in which the targets lie. In this case the MSD may be the natural detection method to be used.

The MSD was originally developed for signal detection in subspace interference and white Gaussian noise [69]. However, noise and clutter patterns in images, do not follow the white Gaussian noise model. Here, we formulate a MSD for signal detection in subspace interference and noise which follow the multi-scale GMRF model. A quantitative performance analysis with comparison to competing methods shows the advantages of the proposed method.

This chapter is organized as follows: In section 6.2 we introduce the MSD for the

multi-scale GMRF model. In section 6.3 we analize the methods performance and compare it to competing methods. In section 6.4 we propose an implementation of the method. Finally, in section 6.5 we conclude.

6.2 Model-Based Subspace Detection

In this section, we introduce an anomaly subspace detection method based on a MSD and the multi-scale GMRF model introduced in the previous section.

Scharf and Friedlander [69] formulated a MSD for the general problem of detecting subspace signals in subspace interference and additive white Gaussian noise. However, anomaly detection in imagery, should be based on a statistical model which better describes the background clutter and the noise. For this purpose, we formulate a modified MSD for detection of subspace signals in subspace interference and additive noise which follows the multi-scale GMRF model, proposed in the previous section.

Let $\{\mathbf{h}_j | j = 1, \ldots, u\}$ and $\{\mathbf{s}_k | k = 1, \ldots, v\}$ denote two sets of image chips, which span the signal and interference subspaces of image Y, respectively. The image chips are all of the same size:$N_x \times N_y$ pixels, which is usually much larger than the size of the neighborhood $\mathcal{R}$. It should be large enough for containing shapes which span the signal and interference subspaces.

We assume that image Y contain mainly noise, which follows the multi-scale GMRF model, and that the target and interference signals are rare. Let $\mathcal{D}_p$ denote an operator which calculates the prediction error, $\widehat{\boldsymbol{\varepsilon}}(\mathbf{s})$, of the multi-scale GMRF model with p independent components. $\mathcal{D}_p$ is defined by using (5.5), (5.7),

and (5.4), as follows:

$$\widehat{\boldsymbol{\varepsilon}}(\mathbf{s}) = [\widehat{\varepsilon}_1(\mathbf{s}), \widehat{\varepsilon}_2(\mathbf{s}), ..., \widehat{\varepsilon}_p(\mathbf{s})]^{\mathrm{T}} = \mathcal{D}_p Y(\mathbf{s}) = K^{\mathrm{T}}\mathbf{Y}(\mathbf{s}) - \sum_{\mathbf{r}\in\mathcal{R}} \boldsymbol{\Theta}_{\boldsymbol{r}} K^{\mathrm{T}}\mathbf{Y}(\mathbf{s}+\mathbf{r}). \tag{6.1}$$

Let $\mathbf{n}_\ell(\mathbf{s})$ denote the column stack ordering of an $N_x \times N_y$ pixels image-chip of $\widehat{\varepsilon}_\ell$ around $\mathbf{s}$:

$$\mathbf{n}_\ell(\mathbf{s}) = \mathrm{vec}(\{\widehat{\varepsilon}_\ell(\mathbf{t}) | \mathbf{t} \in [N_x \times N_y \text{ image chip around } \mathbf{s}]\}). \tag{6.2}$$

We define H_ℓ and S_ℓ as follows:

$$\begin{aligned} H_\ell &= \left[\ \mathrm{vec}([\mathcal{D}_p\mathbf{h}_1]_\ell) \quad \mathrm{vec}([\mathcal{D}_p\mathbf{h}_2]_\ell) \quad \dots \quad \mathrm{vec}([\mathcal{D}_p\mathbf{h}_u]_\ell)\ \right] \\ S_\ell &= \left[\ \mathrm{vec}([\mathcal{D}_p\mathbf{s}_1]_\ell) \quad \mathrm{vec}([\mathcal{D}_p\mathbf{s}_2]_\ell) \quad \dots \quad \mathrm{vec}([\mathcal{D}_p\mathbf{s}_v]_\ell)\ \right] \end{aligned} \tag{6.3}$$

where $[\cdot]_\ell$ denotes the ℓ-th layer of the 3-dimensional data.

Let $\langle H_\ell \rangle$ denote the signal subspace, spanned by the columns of matrix H_ℓ and let $\langle S_\ell \rangle$ denote the interference subspace, spanned by the columns of matrix S_ℓ. We denote the additive noise by $\mathbf{b}_\ell$. The problem is to determine whether the sample vector $\mathbf{n}_\ell$ contains a target signal. The target signal $\mathbf{x}_\ell$ can be described as a linear combination of the columns of H_ℓ *i.e.*, $\mathbf{x}_\ell = H_\ell \boldsymbol{\psi}_\ell$, where $\boldsymbol{\psi}_\ell$ is a vector of coefficients. The interference signal is described similarly, using the matrix S_ℓ and the coefficients vector $\boldsymbol{\phi}_\ell$. Considering the detection problem, we define two hypotheses, H_0 and H_1 which indicate, respectively, absence and presence of target signal in the vector $\mathbf{n}_\ell$:

$$\begin{aligned} H_0 &: \mathbf{n}_\ell = S_\ell \boldsymbol{\phi}_\ell + \mathbf{b}_\ell \\ H_1 &: \mathbf{n}_\ell = H_\ell \boldsymbol{\psi}_\ell + S_\ell \boldsymbol{\phi}_\ell + \mathbf{b}_\ell. \end{aligned} \tag{6.4}$$

Let P_{S_ℓ} denote the projection of a vector onto the subspace $\langle S_\ell \rangle$:

$$P_{S_\ell}\mathbf{n}_\ell(\mathbf{s}) = S_\ell({S_\ell}^{\mathrm{T}}S_\ell)^{-1}{S_\ell}^{\mathrm{T}}\mathbf{n}_\ell(\mathbf{s}) \tag{6.5}$$

and let $P_{H_\ell S_\ell}$ denote the projection of a vector onto the subspace $\langle H_\ell S_\ell \rangle$, spanned by the columns of the concatenated matrix $\begin{bmatrix} H_\ell & S_\ell \end{bmatrix}$. The maximum likelihood estimates of the additive noise vector, $\mathbf{b}_\ell$, under H_0 and under H_1 are denoted by $\widehat{\mathbf{b}}^\ell_{H_0}$ and $\widehat{\mathbf{b}}^\ell_{H_1}$, respectively. These estimates are obtained by subtracting from $\mathbf{n}_\ell$ the components which lie in the signal and interference subspaces as follows:

$$\begin{array}{l} \widehat{\mathbf{b}}^\ell_{H_0} = (I - P_{S_\ell})\mathbf{n}_\ell \\ \widehat{\mathbf{b}}^\ell_{H_1} = (I - P_{H_\ell S_\ell})\mathbf{n}_\ell \end{array} \tag{6.6}$$

$\mathbf{b}_\ell$ is the innovations process of a GMRF and therefore is normally distributed with zero mean. We denote the covariance matrix of $\mathbf{b}_\ell$ by $\rho^2_\ell\Phi_\ell$, where ρ^2_ℓ is the variance of $\mathbf{b}_\ell$. $\rho^2_\ell\Phi_\ell$ is obtained by using (2.5).

The detection problem can be formulated as a GLRT between H_0 and H_1. The log-likelihood ratio, L_ℓ, calculated based on the ℓ-th layer of the innovations process is given by:

$$\begin{aligned} L_\ell(\mathbf{s}) &= 2ln\left[\frac{Pr(\mathbf{b}_\ell(\mathbf{s})|H_0)}{Pr(\mathbf{b}_\ell(\mathbf{s})|H_1)}\right] = 2ln\left[\frac{\exp\left(\frac{\left[\Phi_\ell^{-1/2}\widehat{\mathbf{b}}^\ell_{H_0}(\mathbf{s})\right]^2}{2\rho^2_\ell}\right)}{\exp\left(\frac{\left[\Phi_\ell^{-1/2}\widehat{\mathbf{b}}^\ell_{H_1}(\mathbf{s})\right]^2}{2\rho^2_\ell}\right)}\right] \\ &= \frac{1}{\rho^2_\ell}\left[\left\|\Phi_\ell^{-1/2}\widehat{\mathbf{b}}^\ell_{H_0}(\mathbf{s})\right\|_2^2 - \left\|\Phi_\ell^{-1/2}\widehat{\mathbf{b}}^\ell_{H_1}(\mathbf{s})\right\|_2^2\right]. \end{aligned} \tag{6.7}$$

We assume the different layers of the innovations process $\mathbf{b}_\ell(\mathbf{s})$ are normally and independently distributed. Thus, the probability function of the p components of

pixel $\mathbf{s}$ is the product of p Gaussian functions. Under the *log* function, the product in the likelihood ratio expression is translated into a sum of p log-likelihood ratios (separately calculated for each layer). The log-likelihood ratio, based on p independent layers of the innovations process is then given by $L(\mathbf{s}) = \sum_{\ell=1}^{p} L_\ell(\mathbf{s})$ as follows:

$$\begin{aligned} L(\mathbf{s}) &= \sum_{\ell=1}^{p} \frac{1}{\rho_\ell^2} \left[\left\| \Phi_\ell^{-1/2} \widehat{\mathbf{b}}_{H_0}^{\ell}(\mathbf{s}) \right\|_2^2 - \left\| \Phi_\ell^{-1/2} \widehat{\mathbf{b}}_{H_1}^{\ell}(\mathbf{s}) \right\|_2^2 \right] = \\ &= \sum_{\ell=1}^{p} \frac{1}{\rho_\ell^2} [\Phi_\ell^{-1/2} \mathbf{n}_\ell(\mathbf{s})]^{\mathrm{T}} \left(P_{H_\ell S_\ell} - P_{S_\ell} \right) [\Phi_\ell^{-1/2} \mathbf{n}_\ell(\mathbf{s})]. \end{aligned} \tag{6.8}$$

The 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, as follows:

$$SNR = \sum_{l=1}^{p} \frac{1}{\rho_\ell^2} \mathbf{x}_\ell{}^{\mathrm{T}} [I - P_{\widetilde{S}_\ell}] \mathbf{x}_\ell. \tag{6.9}$$

Let u denote the rank of the signal subspace and let $q = up$. L is a sum of squared independent normally distributed variables and therefore is chi-square distributed with q degrees of freedom, as follows:

$$L \sim \begin{cases} \chi_q^2(0) & , \text{under } H_0 \\ \chi_q^2(SNR) & , \text{under } H_1. \end{cases} \tag{6.10}$$

Under hypothesis H_1, the non-centrality parameter of the chi-square distribution of L is equal to the SNR [69]. The decision rule is based on thresholding the log-likelihood ratio using the threshold η as follows:

$$\xi = \begin{cases} H_0 & \text{if} \quad L \leq \eta \\ H_1 & \text{if} \quad L > \eta. \end{cases} \tag{6.11}$$

Given (6.10) and (6.11), the probabilities of false-alarm and detection are:

$$P_{FA} = 1 - P\left[\chi_q^2(0) \leq \eta\right] \tag{6.12}$$

$$P_D = 1 - P\left[\chi_q^2(SNR) \leq \eta\right]. \tag{6.13}$$

6.3 Performance Analysis

In this section we analyze the performance of the proposed algorithm. We present the receiver operating characteristics (ROC) of the algorithm with respect to different parameters and compare it to competing methods.

6.3.1 Receiver Operating Characteristics

The ROC of the proposed algorithm is calculated using (6.9), (6.12), and (6.13). The SNR, given by (6.9), is a function of the target's shape and intensity, the variance of the background's innovations process and the interference subspace span. The SNR increases with the norm of $\mathbf{x}$ and the angle between $\mathbf{x}$ and $\langle S \rangle$. For the synthetic example, used in these calculations, the SNR also increases with the number p of independent layers we use for detection. Large background variance results in a smaller SNR and therefore, performance degradation.

Figure. 6.1 presents the ROC of the proposed algorithm for various SNRs. This example presents the performance of the proposed anomaly detection algorithm using 3 independent component ($p = 3$). The probability of false alarm (P_{FA}) and the probability of detection (P_D) are calculated using (6.12) and (6.13) respectively. Figure 6.2 presents the ROC curves versus p given a constant target norm and background variance. For the example we use, the SNR and the probability

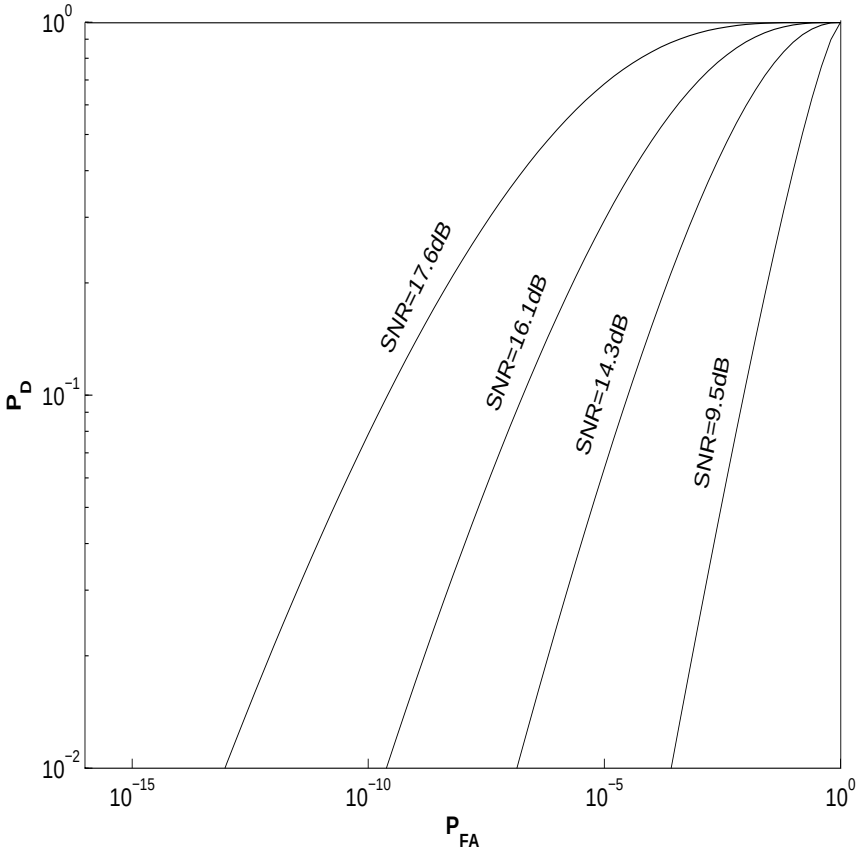


Figure 6.1: An example of ROC calculated for the proposed algorithm, using 3 principle component ($p = 3$) and various values of SNR

of detection (P_D) improve with p. The use of more independent layers improves the detection performance due to the additional information concealed in each layer. It is theoretically possible that using an independent layer which contains dim target signal (or no target signal at all) will reduce the effective SNR and cause performance reduction. In practice, we checked the influence of this parameter on the detection performance using a large number of synthetic examples and did not find any example for which the detection performance decreased with p. Another factor which may influence the performance is the dimension of the interference subspace. According to (6.9), when the dimension of S'_ℓ decreases, the SNR is lower and the performance is reduced.

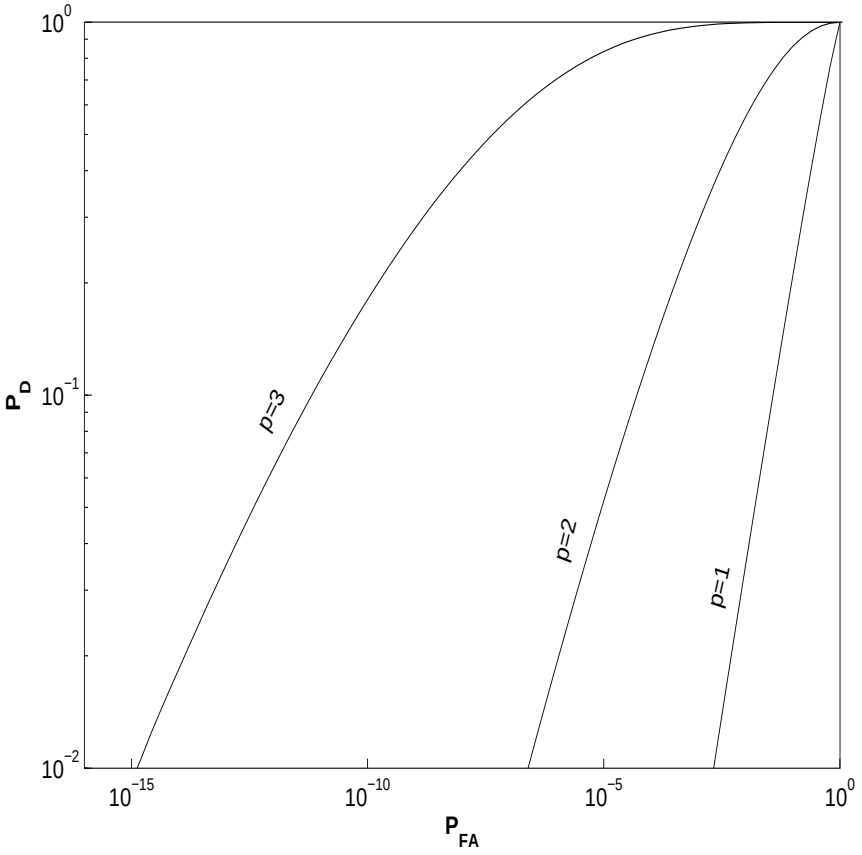


Figure 6.2: An example of ROC calculated for the proposed algorithm using p independent components. Using larger number of independent components, increases the SNR and improves the performance.

6.3.2 Performance Comparison

In order to analyze the detection performance of the proposed algorithm we compared it with those of other methods. In this section we refer to the proposed method as "*Proposed Method III*" and compare it to a competing method and to other proposed methods which are similar to *proposed method III* except of one element or step in the process. We applied the competing methods to synthetic images of an airplane on cloudy background. The synthetic images are generated by the process, described in Sec. 7.2. Figure 6.3 shows the flowcharts of the

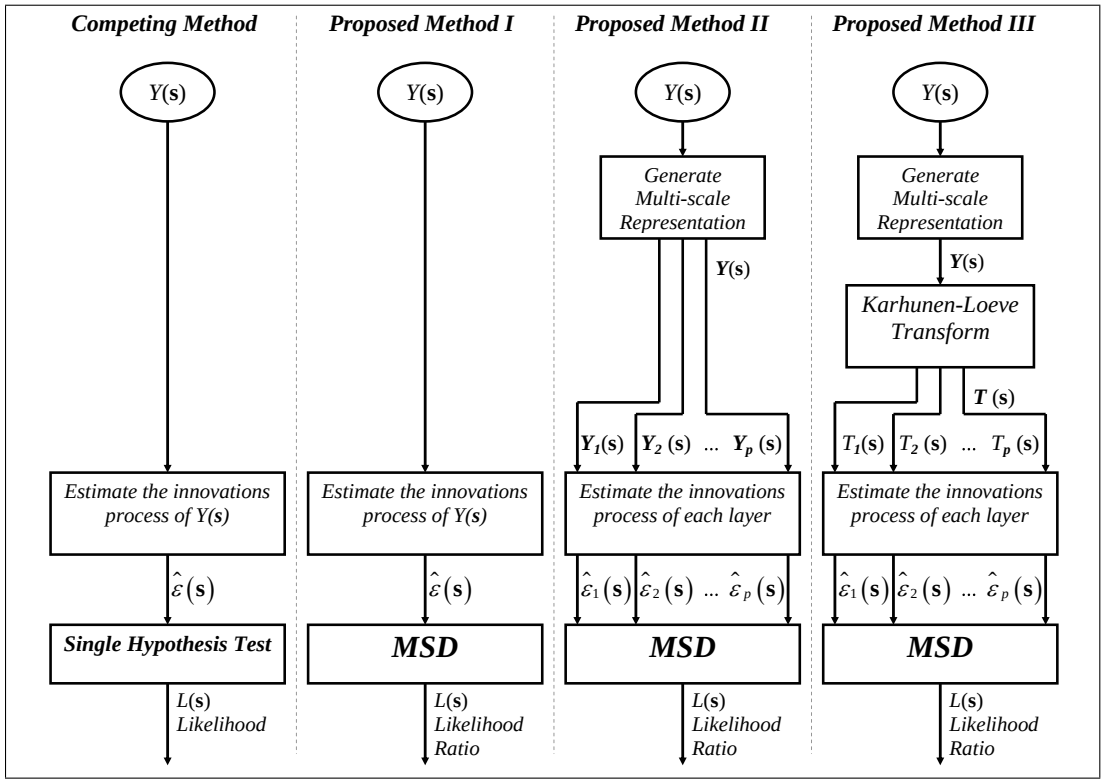


Figure 6.3: Flowcharts of the compared detection methods.

different method compared in this section:

A Competing Method: We assume the image follows the GMRF model. A single hypothesis scheme is applied to the estimated innovations process of the image for the detection of regions, which appear unlikely with respect to its normal distribution [36].

Proposed Method I: We assume the image follows the conventional GMRF model rather than the multi-scale GMRF model. The MSD, proposed in Sec. 6.2, is applied to the estimated innovations process. By analyzing the performance of this method and comparing it to *Method III* we examine the contribution of the proposed multi-scale model to the detection performance.

Proposed Method II: The multi-scale GMRF model is employed, skipping the KLT step. We assume the layers of the multi-scale representation $\mathbf{Y}(\mathbf{s})$ follow the GMRF model and estimate the innovations process of each of these layers. The MSD, proposed in Sec. 6.2, is applied to the estimated innovations process. By analyzing the performance of this method we examine the significance of the KLT in the modeling process.

Proposed Method III: This method includes all the elements of the multi-scale model proposed in Sec. 5.2.2. We assume the image follows the proposed multi-scale GMRF model. The MSD, proposed in sec. 6.2, is applied to the estimated innovations process.

Figures 6.4 and 6.5 present examples in which the proposed and competing algorithms are applied to the same synthetic images of airplanes on cloudy background. Figure 6.4 shows the synthetic images, generated using a mixture of three images of different periodical patterns (generated by summing up 3 GMRFs with 3 different parameters sets). The airplane is planted in the center of each image. Figure 6.5 (a) presents the results of the *Competing Method*, similar to the method presented by Hazel [36], applied to the images in Fig 6.4. Figure 6.5 (b) shows the resulting likelihood ratios generated by applying *Proposed Method I* to the images in Fig. 6.4. Figure 6.5 (c) presents the resulting likelihood ratios generated by applying *Proposed Method II* to the images in Fig. 6.4. Figure 6.5 (d) presents the resulting likelihood ratios generated by applying *Proposed Method III* to the images in Fig. 6.4, assuming the proposed Multi-Scale GMRF model. Figure 6.5 (d) shows the best detection results. The likelihood ratios are close to 0 everywhere except in the target's region. The results of the *Competing Method* contain high level of background noise which does not allow the targets detection

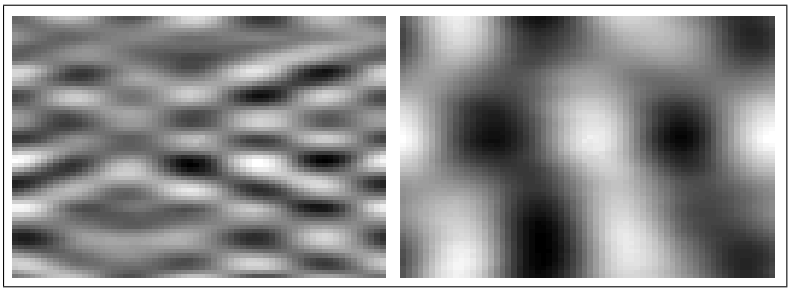


Figure 6.4: Synthetic images containing cloudy background and an airplane target in their centers

by thresholding the likelihood image. The results of *Proposed Method II* contain a noisy pattern which exceeds the likelihood level in the targets region (in the center of the image). According to these examples, it seems that skipping the KLT in the modelling process, significantly reduces the performance related to *Proposed Method III*.

Figure 6.6 presents the ROC, analytically calculated for synthetic images of an airplane on cloudy sky. The ROC curves reflect the performance of *Proposed Method II* (solid) compared to the performance of *Proposed Method I* (dashed). The images, for which the ROC's where calculated, are similar to the right example presented in Fig. 6.4. The only difference between the images is in the SNR.

The SNRs for which the ROCs where drawn are summarized in Table 6.1. The table specifies 4 different cases of background variances and target norms (L-infinity norms of the targets image). The SNRs obtained by *Proposed Method III*, which is based on the multi-scale GMRF model, are higher than those obtained by *Proposed Method I*, which is based on a conventional GMRF model, and *Proposed Method II*, which is based on a modification of the multi-scale GMRF model.

Table 6.1: Properties of the different cases for which the ROC curves in Fig. (6.6) were drawn. The SNR calculated for *Proposed Method III* is significantly higher than the SNR calculated for the other methods.

Case	**Background Variance**	**Target Norm**	**SNR[dB]**		
			Proposed Method I	***Proposed Method II***	***Proposed Method III***
(a)	188	1.3	2.8	-0.5	7.3
(b)	188	1.4	3.2	0	7.7
(c)	188	1.5	3.6	0.4	8.1
(d)	188	1.7	4.0	0.8	8.5

6.4 Implementation

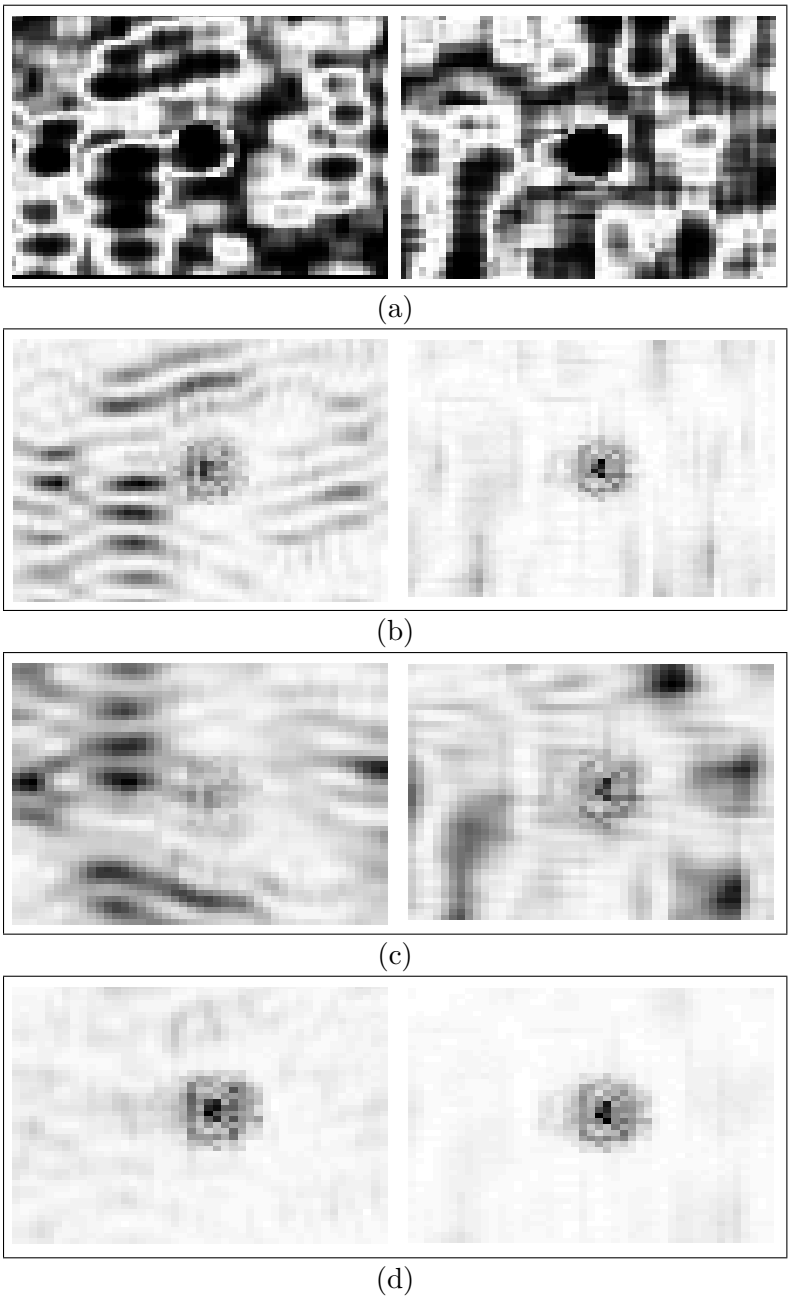
In this section, we describe an implementation of the proposed anomaly detection algorithm and calculate its computational complexity. Figure 6.7 presents a flow chart with the main steps of the algorithm:

1. **Generation of a Multi-Scale Representation:** The image Y is filtered by a set of spatial filters, $\mathcal{G}$, using (5.5), in order to create its multi-scale representation, $\mathbf{Y}$.

2. **Independent Components Analysis:** The Karhunen-Loéve transform is applied to the vectors of the multi-scale representation, $\mathbf{Y}$, using (5.7). The result is a multi-scale representation, $\mathbf{T}$, with independent layers.

3. **Estimation of the Innovations Process:** The GMRF parameters set is separately estimated for each layer of $\mathbf{T}$. The sample innovations, $\widehat{\varepsilon}_\ell(\mathbf{s})$, are calculated for each layer, ℓ, of $\mathbf{T}$ using (6.1) and the estimated parameters.

4. **Matched Subspace Detector:** S_ℓ and H_ℓ are calculated using (6.3). A



(a)

(b)

(c)

(d)

Figure 6.5: A comparison between detection methods. (a) Results of the *Competing Method* applied to the images in Fig. 6.4 ; (b) Results of *Proposed Method I* applied to the images in Fig. 6.4 ; (c) Results of *Proposed Method II* applied to the images in Fig. 6.4 ; (d) Results of *Proposed Method III* applied to the images in Fig. 6.4. The images in (d) seems to have the lowest false alarm rate (FAR).

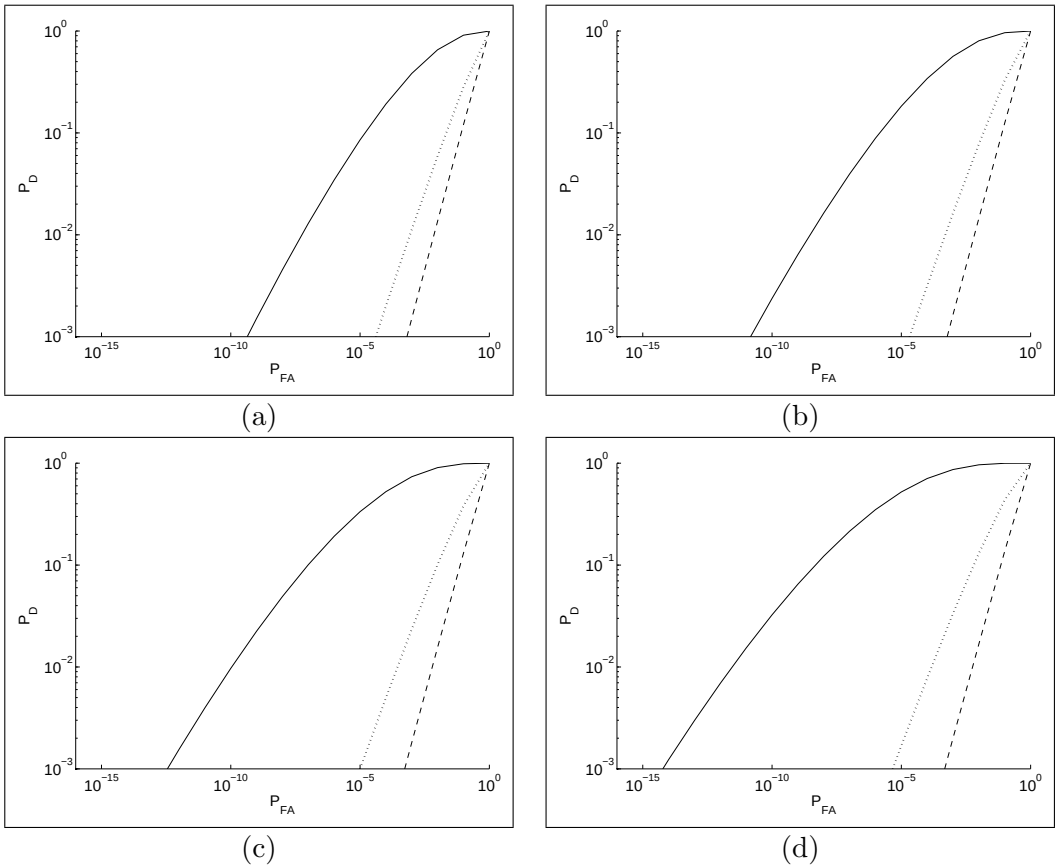


Figure 6.6: Performance of the anomaly detection based on *Proposed Method III* (solid), *Proposed Method II* (dashed), and *Proposed Method I* (dotted). (a)-(d) correspond to different parameter settings as specified in Table 6.1.

matched subspace detector is formed and the log-likelihood ratio, L, is calculated for each pixel using (6.8).

5. **Decision Rule (Thresholding):** The decision rule given in (6.11) is applied to the log-likelihood ratio, L, in order to determine whether a pixel $\mathbf{s}$ belongs to a target. The threshold, η, is determined by the admissible false alarm rate (FAR) using (6.12).

The computational complexity of the proposed algorithm is a function of the size of the image ($M_x \times M_y$), the rank of the space in which the signal and in-

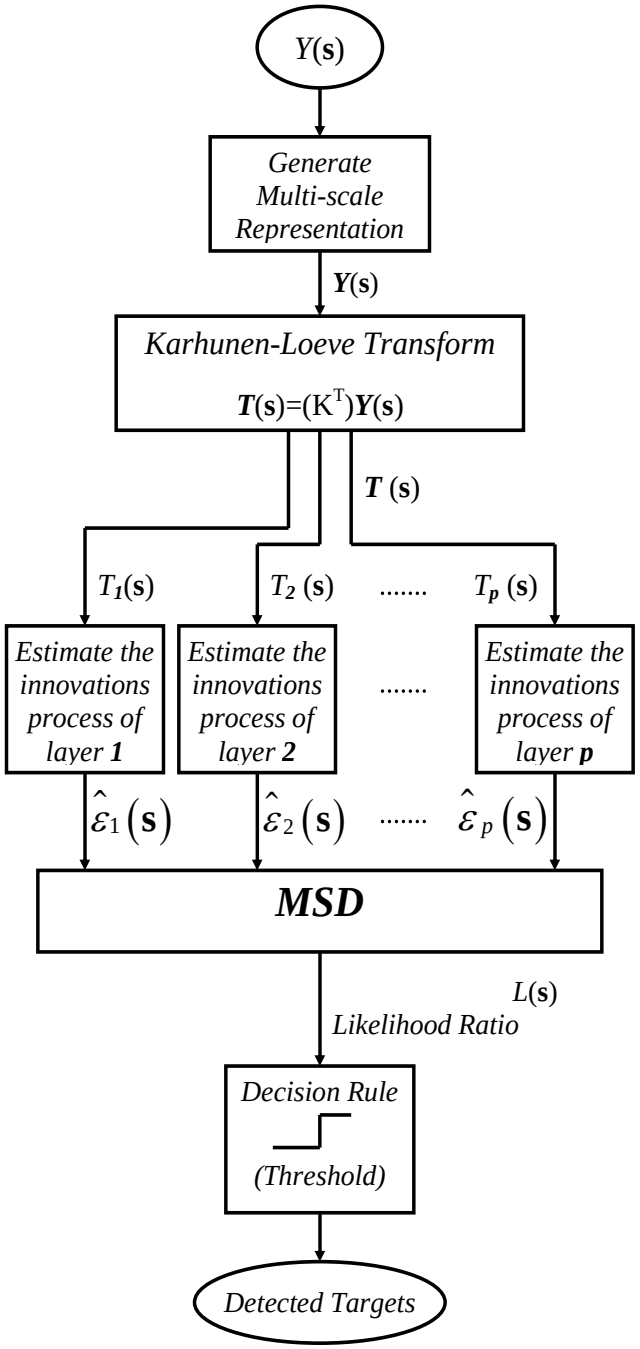


Figure 6.7: Flow chart of the proposed algorithm implementation.

terference subspaces lie $(N_x \times N_y)$, and the number of independent components p employed for the detection. The computational complexity of the multi-scale representation generation is $O(nN_xN_yM_xM_y)$ where n is the number of multi-

scale filters. Applying the KLT to the multi-scale representation of the image using the covariance matrix of the data is [32]: $O(N_x^2 N_y^2 M_x M_y + N_x^3 N_y^3)$. Using the singular value decomposition (SVD)-based approach, reduces the computational complexity of the KLT to $O(N_x^2 N_y^2 M_x M_y)$ [39]. The estimation of the innovations process is $O(|\mathcal{R}| M_x M_y)$ and the MSD is $O(M_x M_y N_x N_y p)$. Thus, the total computational complexity of the proposed algorithm is $O(N_x^2 N_y^2 M_x M_y + N_x N_y M_x M_y n + |\mathcal{R}| M_x M_y)$.

6.5 Summary

We introduced an anomaly subspace detection method based on a multi-scale GMRF model. The detection is carried out by applying a MSD to the innovations process estimated assuming the image follows the multi-scale GMRF model. The MSD incorporates available *a priori* information about the targets into the detection process and thus potentially improves the detection performance. Performance analysis was carried out by investigating the influence of different parameters on the detection performance, and comparing the performance of the proposed method to those of competing methods. The analysis demonstrate the advantages of the proposed method related to other methods which do not rely on the multi-scale GMRF model. Finally, an implementation of the method was proposed and its computational complexity was computed. In chapter 7 we present some experimental results which support the analytical results presented in this chapter. We demonstrate the proposed method using real images from different sources.

Chapter 7

Experimental Results

7.1 Introduction

In the previous chapter we proposed an anomaly subspace detection method and analyzed its performance. In this chapter, we present the results of applying the proposed model and algorithm to synthetic and real images from different sources. The algorithm is applied to: 1) simulated images of airplanes on cloudy background; 2) Sonar images of sea-mines on sea-bottom background; and 3) Detection of defects in wafer images. The different examples and applications demonstrate the robustness and flexibility of the algorithm.

The chapter is organized as follows: In section 7.2 we present the application of the proposed model to synthetic images. We first describe the process of generating synthetic clutter images and planting targets, then we present some detection results. In section 7.3.1 we present the results of sea-mine detection in sonar images. In section 7.3.2 we present the results of applying the proposed method to defect detection in wafers production.

7.2 Synthetic Imagery

In this section we present the application of the proposed method to target detection in synthetic images of airplanes on cloudy background. We first present a method for synthetic image generation, then we show some examples of likelihood ratio images, calculated using the proposed anomaly subspace detection method.

7.2.1 Generating Synthetic Images

The synthetic examples presented in this subsection contain airplanes on cloudy background. The synthetic cloudy background is created using random images which follow the GMRF model. These random images, are generated based on the expression of the GMRF model in terms of white noise. Let $T(\mathbf{s})$ be an image of $M_x \times M_y$ pixels which follows the GMRF model, let $\mathcal{F}$ be the DFT operator, and let $\mathbf{w} = (w_1, w_2)$ be the 2-dimensional indices of the data in the frequency domain. Then the equivalent DFT representation of (2.1) for image $T(\mathbf{s})$ is as follows [41]:

$$\mathcal{F}\{T\}(\mathbf{w}) = \frac{\rho \mathcal{F}\{\nu(\mathbf{s})\}}{\sqrt{\lambda(\mathbf{w})}} \tag{7.1}$$

where $\lambda(\mathbf{w})$ is given by:

$$\lambda(\mathbf{w}) = 1 - 2 \sum_{\mathbf{r} \in \mathcal{R}_h} \theta(\mathbf{r}) \cos\left(2\pi \left[\frac{r_1 w_1}{M_x} + \frac{r_2 w_2}{M_y}\right]\right) \tag{7.2}$$

$\{\nu(\mathbf{s})|\mathbf{s}\}$ are IID Gaussian random variables, with zero mean and unit variance, $\mathcal{R}_h$ is half of the symmetric neighborhood $\mathcal{R}$, and $\mathbf{r} = (r_1, r_2)$. For the model

to exist and be stable, the following expression must be true for every $\mathbf{w}$ in the support of the image:

$$\lambda(\mathbf{w}) > 0. \tag{7.3}$$

From the expression in (7.1), the procedure for synthetic generation of random fields obeying this model is evident. Further details can be found in [41].

Figure 7.1 presents a flow chart of the procedure used for generating synthetic images of airplanes on cloudy background. The synthetic examples are generated as follows:

1. Three random images are generated based on the GMRF model. Each image is obtained by using (7.1) and (7.2)with different sets of parameters ($\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, and $\boldsymbol{\theta}^{(3)}$).

2. A weighted sum of the three images is calculated (using the weight coefficients ω_1, ω_2, and ω_3). The result contains several periodical patterns with different period lengths.

3. A small airplane image is planted in the background image in a random position and orientation.

The results of this procedure contain mixtures of periodical patterns with different period lengths. This procedure does not describe an accurate synthesis of images which follow the multi-scale GMRF model. The process only claims to generate images that may be better described by the multi-scale GMRF model rather than the conventional GMRF model. The results presented in Sec. 6.3.2 demonstrate the correctness of this argument. The effective SNR achieved by the MSD under the multi-scale GMRF model assumption (*Proposed Method II*) is higher than

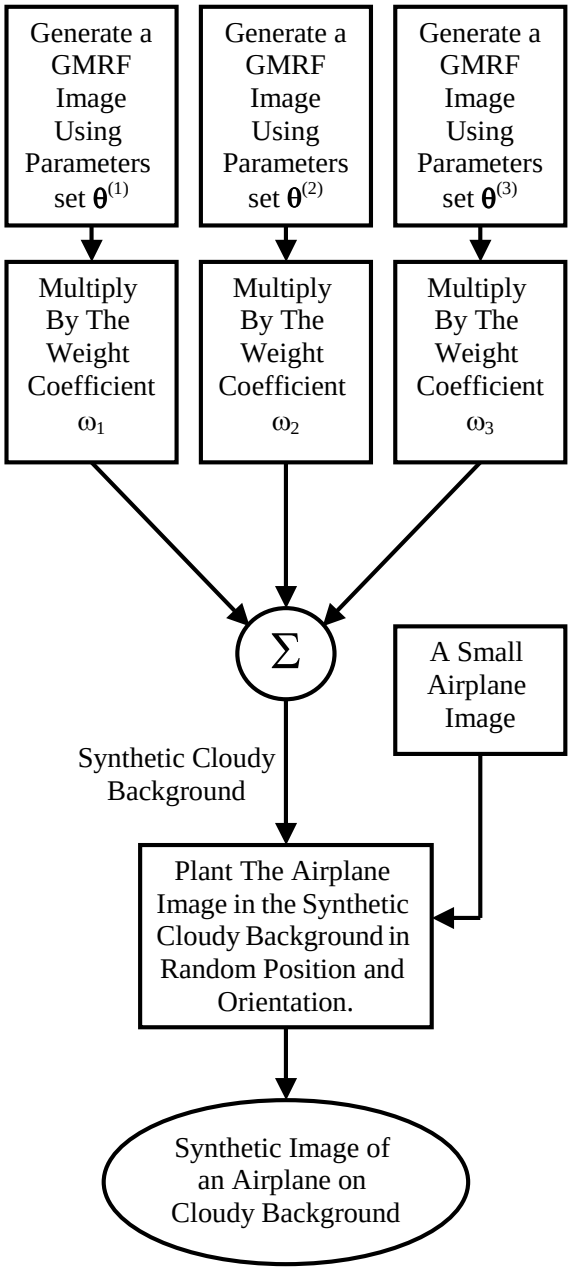


Figure 7.1: Flow chart of the procedure for synthetic examples generation.

the SNR achieved under the conventional GMRF model assumption (*Proposed Method I*).

7.2.2 Target Detection

Figure 7.2 presents examples of synthetic images generated as described above. A multi-scale representation of each image is obtained by applying undecimated wavelet transform with 2 scale levels to the image. Accordingly, the layers of the multi-scale representation are the result of convolving the image with the wavelet basis images. We employ a signal subspace that is constructed from the span of 4 image chips of 11×11 pixels. The image chips contain bar shapes in different orientations: $0°$, $45°$, $90°$, and $135°$ which resemble the fuselage of airplane targets. Figure. 7.3 shows the likelihood ratio (in gray-scale), calculated using (6.7). Black regions denote high likelihood ratio. The target detection is carried out by thresholding the likelihood images. The threshold is determined by the predefined admissible level of FAR. The detected targets are marked by circles (in Fig. 7.3). This example demonstrates the robustness of the algorithm in presence of different patterns of background. The image chips which span the signal subspace (target subspace) are simple and generally do not require detailed information about the targets.

Figures 7.4 and 7.5 show an example of target detection using the proposed algorithm with 3 independent components ($p = 3$). Figure 7.4 shows a synthetic image of cloudy sky with an airplane in its middle. The airplane is unnoticeable by a human viewer due to its weak signature. Figures. 7.5(a)-(c) show the images of the 3 top independent components generated by the algorithm (as detailed in Sec. 5.2). The target is clearly revealed in the 3rd independent component. Figure 7.5(d) shows the likelihood ratio calculated by the proposed algorithm. This image, unlike the images of the independent components, lacks the back-

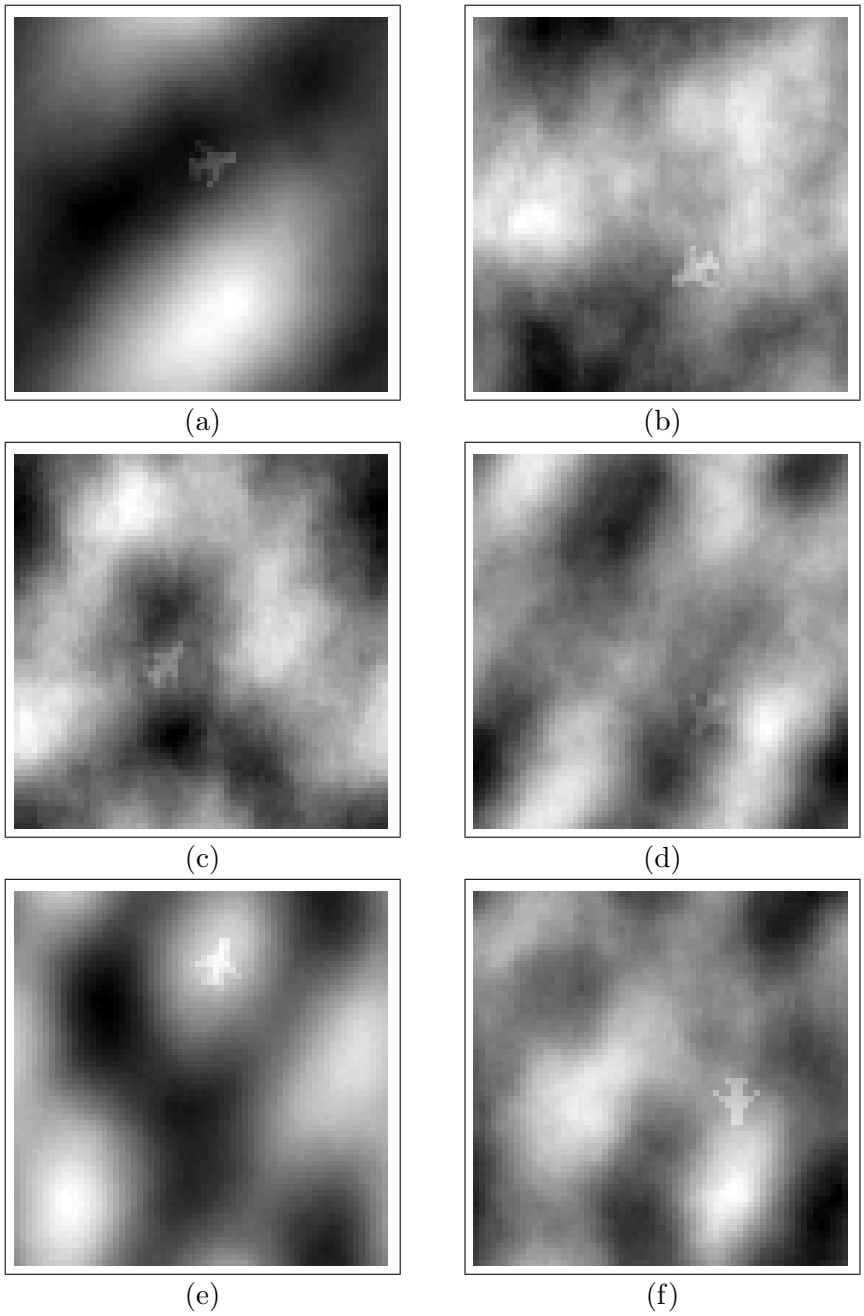


Figure 7.2: Synthetic images of cloudy sky with airplane images planted in random places and orientations.

ground patterns, which are rejected by the innovations noise and likelihood ratio calculations. Thus, the target is more clearly revealed.

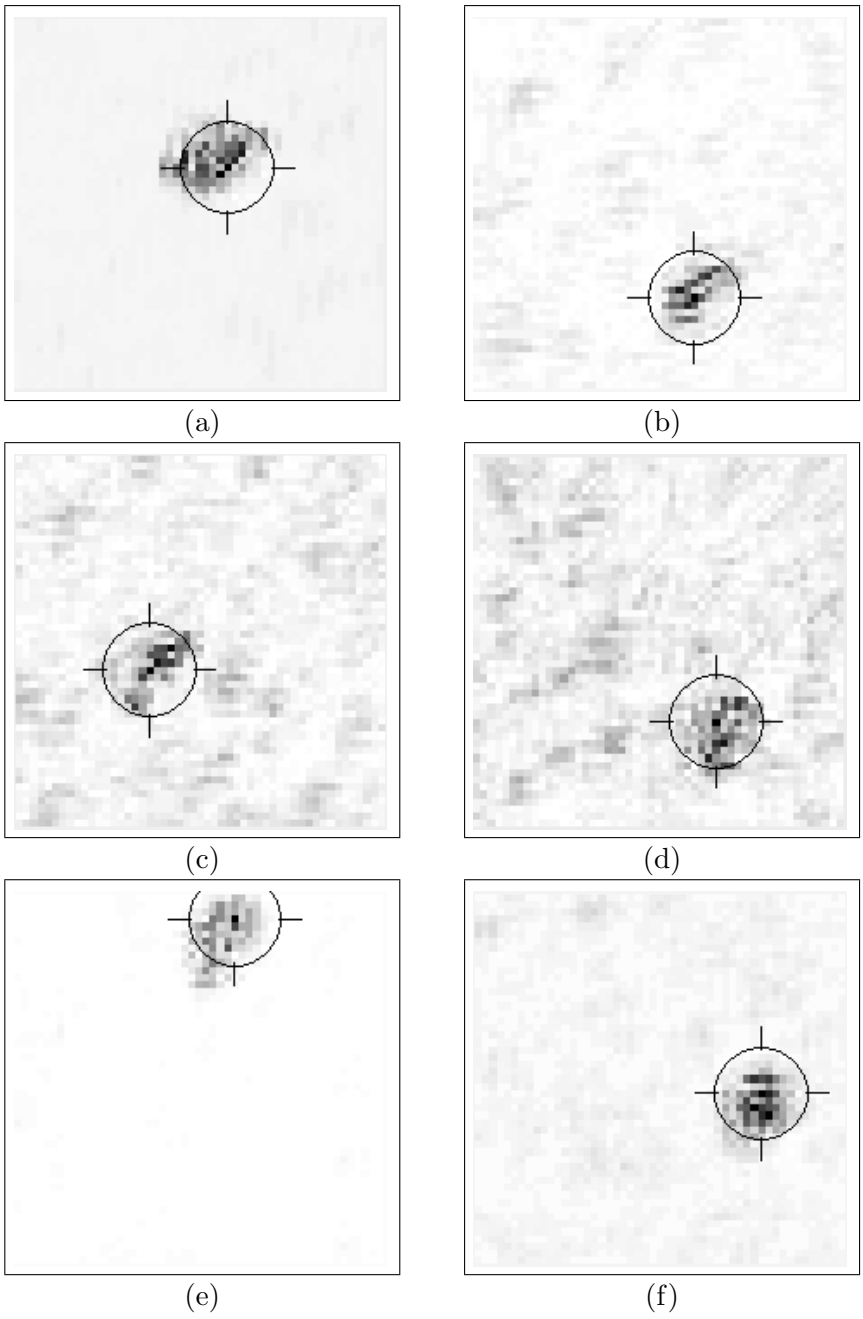


Figure 7.3: Results of anomaly detection applied to the images in Fig.7.2. The gray-scale represents the degree of local anomality around a given pixel. The circles indicate regions where the local anomality is above a predetermined threshold.

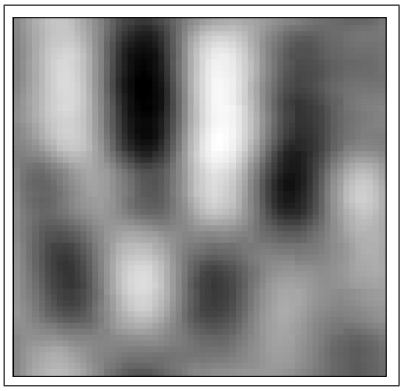


Figure 7.4: A synthetic image of cloudy sky with an airplane in its middle. The airplane is unnoticeable by a human viewer due to its weak signature.

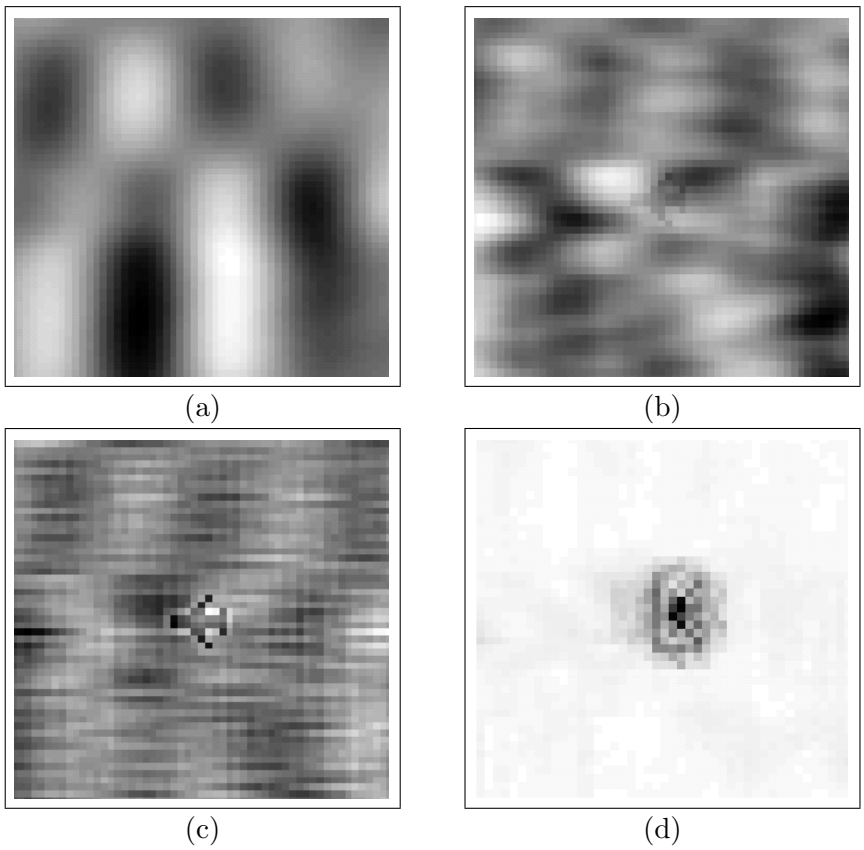


Figure 7.5: Anomaly detection applied to the image in Fig. 7.4. (a) First, (b) second, and (c) third independent components. (d) Likelihood ratio calculated by the proposed algorithm.

7.3 Real Imagery

In this section we present the application of the proposed method to target detection in real imagery. We first show the results of applying the proposed method to sea-mine detection in sonar images, then we present some examples of wafer defects detection for quality assurance purpose. These two different examples demonstrate the capability of the proposed model and algorithm to cope with variety of background clutter patterns.

7.3.1 Detection of Sea-Mines in Sonar Images

The proposed method is demonstrated on real images from a database of sea-mine sonar images. A sea-mine appears in the sonar images as a bar shaped object-highlight accompanied by a shadow which represents the hiding of the seabottom-reverberation by the sea-mine [63]. Mignotte and Collet [58] presented 3-class Markovian segmentation method for the detection of sea-mines in sonar images. The sea-mine images were segmented to 3 kinds of regions: echo, shadow, and sea-bottom reverberation areas, based on different MRF models, estimated for the different classes. Dobeck *et al.* [26] implemented a matched filter, K-nearest neighbor neural network classifier, and a discriminatory filter classifier to detect such mine-like objects in sonar images. The classification process employs up to 45 features for every possible mine-like object. The detection in [26] is based on a large collection of mine-like objects signatures. In the example presented here, no real signature examples are used for defining the signal subspace. Figure 7.6 shows 6 sonar images. Each image contains one sea-mine on highly cluttered seabottom background. The background patterns are diverse.

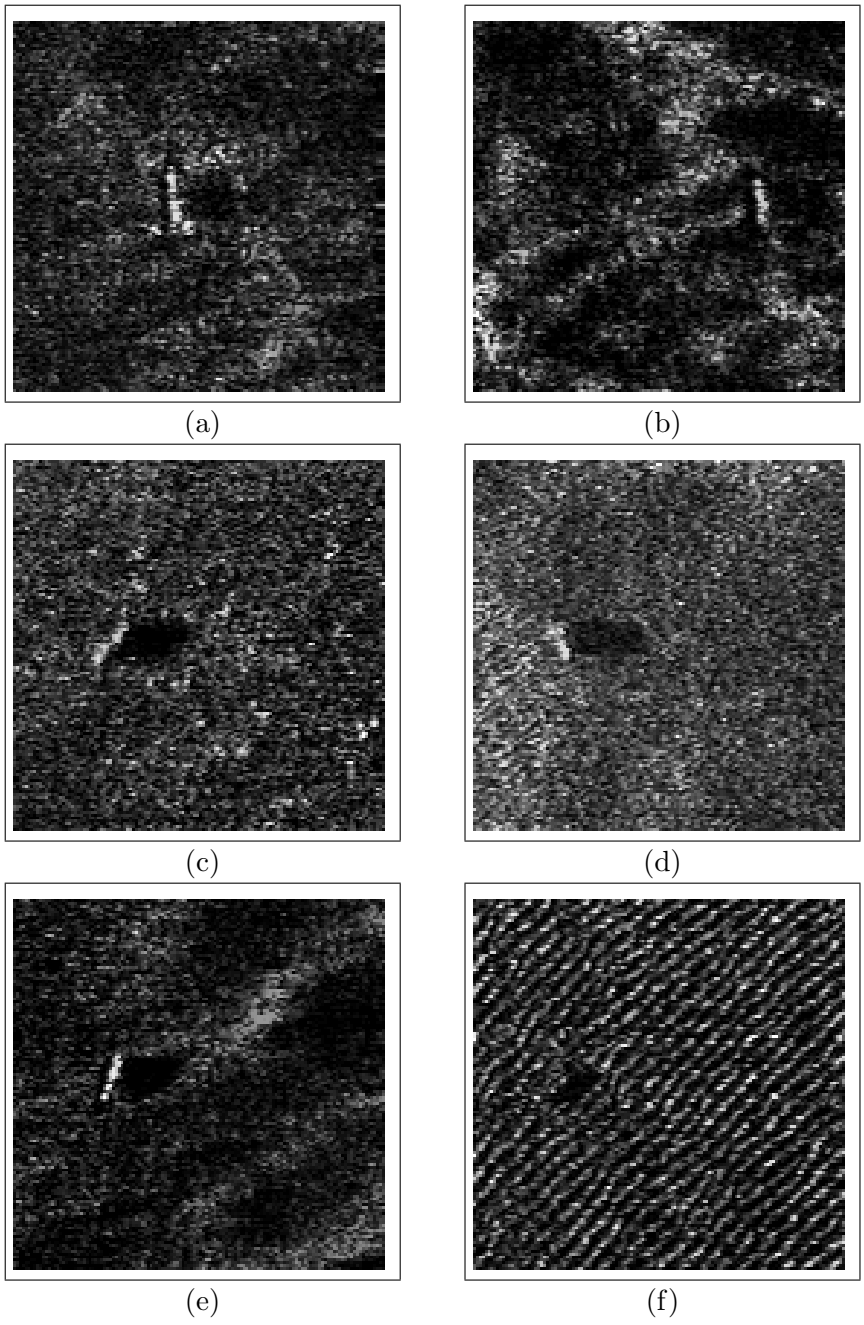


Figure 7.6: Examples of sea-mine sonar images: Sea-mines appear in the sonar images as a bar shaped object-highlight accompanied by a shadow which represents the hiding of seabottom-reverberation by the sea-mine [63].

Figures 7.6(a)-(e) contain relatively slow changing backgrounds while Fig. 7.6(f) contains background with a dominant periodical pattern. The sea-mine's high-

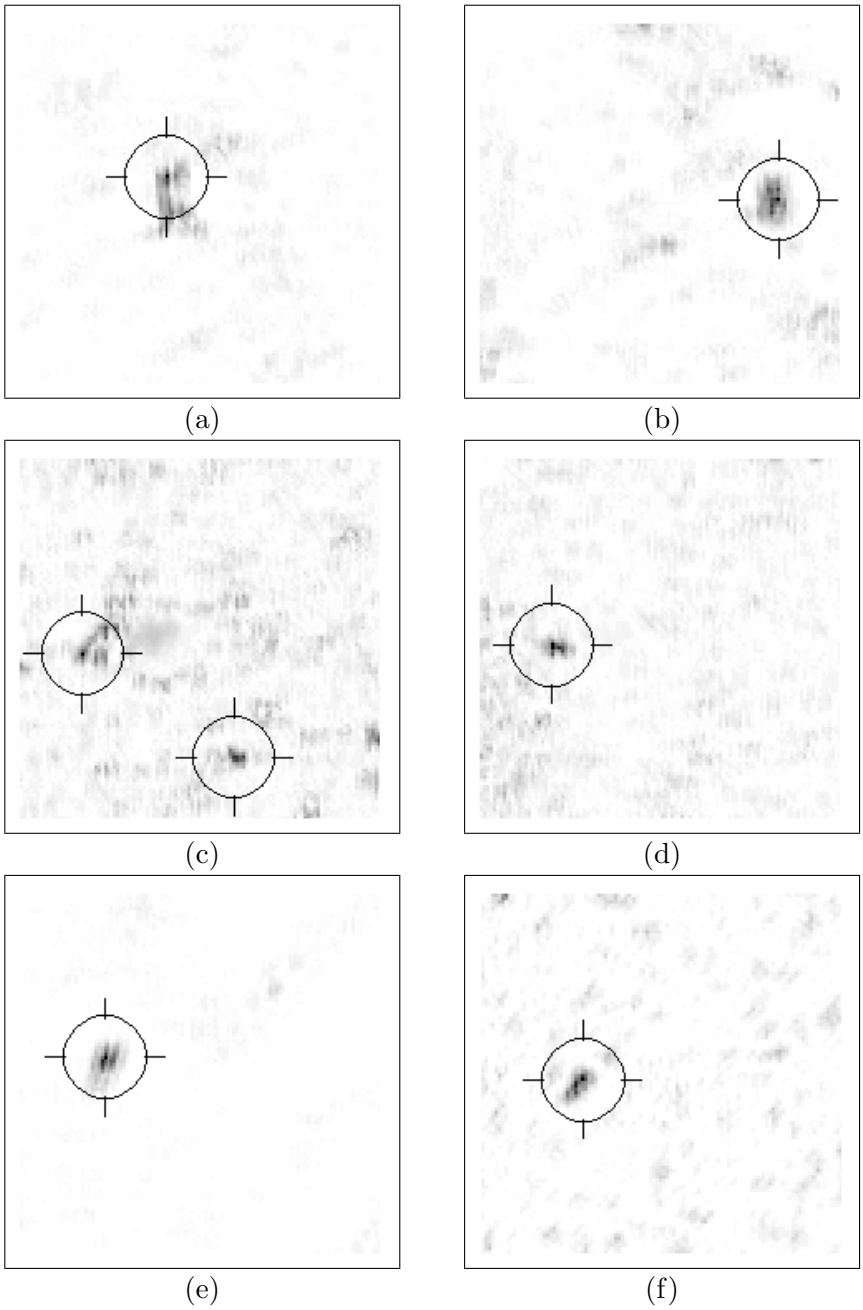


Figure 7.7: Results of the anomaly detection applied to the images in Fig.7.6. The sea-mines are detected by thresholding the gray-scale values which represent the degree of local anomality around a given pixel.

light in Fig. 7.6(f) is unnoticeable while its shadow clearly appears as a dark region. The proposed method is applied to these images for detecting sea-mines.

The multi-scale representations of the images are generated by applying undecimated wavelet transform with 3 scale levels to the images. The signal subspace is formed from the span of 4 image-chips of highlighted bars with dark shadows, in different orientations. The result of the proposed anomaly detection, applied to the sonar images, is shown in Fig. 7.7. The sea-mine in Fig. 7.6(f) is detected despite the absence of sea-mine highlight, due to its dominant shadow. The lower right circle in Fig. 7.7(c) marks a false alarm. This false alarm might result from a mine-like highlight in the background pattern.

The detection results presented here, demonstrate the capability of the proposed model and algorithm to cope with variety of background clutter patterns, using the same filters set and signal subspace. All the sea-mines in these examples are detected. The false alarm in Fig. 7.7(c) may be prevented by a more specific definition of the signal subspace.

7.3.2 Detection of Wafer Defects

The proposed algorithm is applied to detection of defects in wafers images for quality assurance. Figure 7.8 shows examples of wafer images. Each image contains a defect whose diameter is smaller than 3 pixels. The MSD is set to detect circles of 3 pixels diameter and linear shapes of 3 pixels length. The multi-scale representation is generated in the same way as described in the synthetic example. The likelihood images and the detected targets are presented in Fig. 7.9. The results are less impressive than those obtained in the previous applications. However, we are still able to detect the defects with a manageable rate of false alarms. The cause of the performance degradation may be explained by the low correlation of the background patterns. Images with weakly correlated patterns

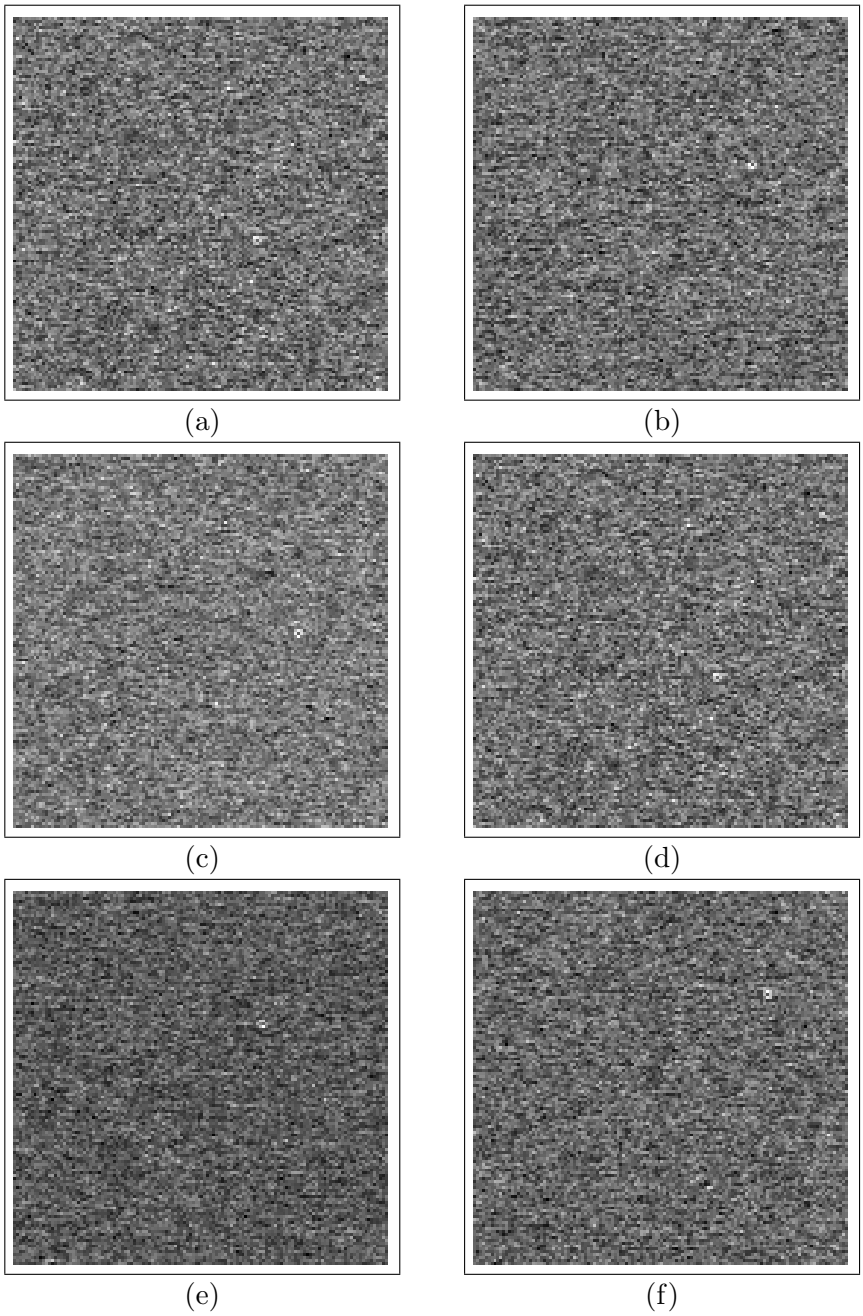


Figure 7.8: Example of wafer images. The 128×128 images include small round defects of about 3×3 pixels.

are not well described by the multi-scale GMRF model.

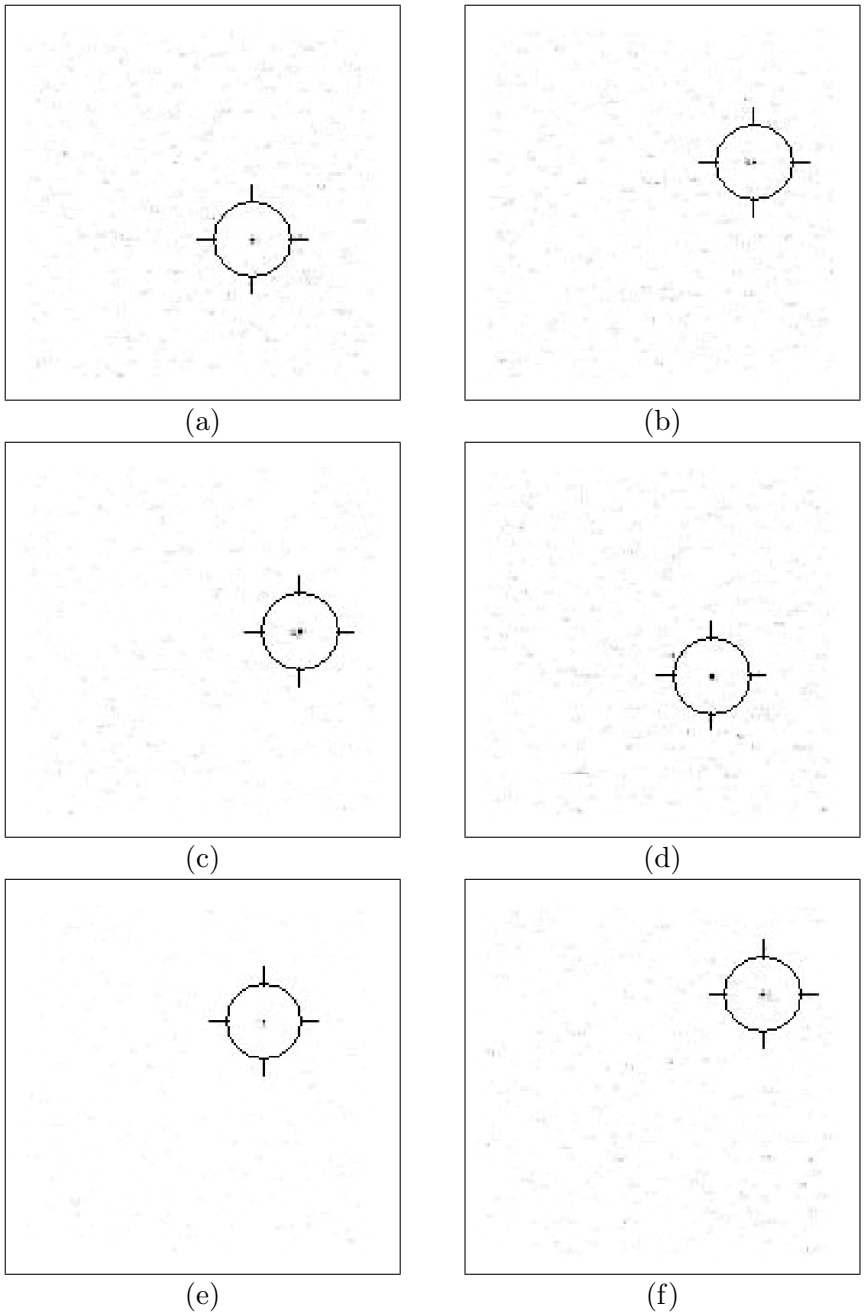


Figure 7.9: Results of the anomaly detection applied to the images in Fig.7.8.

7.4 Summary

We demonstrated the performance of the proposed algorithm with application to automatic target detection. We proposed a procedure for generating synthetic

images of airplanes on cloudy background, then we presented the results of applying the proposed algorithm to target detection in these synthetic images. The proposed method was also applied to sea mine detection in real side-scan sonar images, and to detection of defects in wafer images. The filter sets, as well as the target and interference subspaces were intuitively chosen. We presented an example in which targets, almost unnoticeable by a human viewer, were detected by the proposed method. The results demonstrate the capability of the proposed model and algorithm to cope with variety of targets and background clutter patterns.

Chapter 8

Conclusion

8.1 Summary

We have presented some models of natural clutter images and some known anomaly detection methods. Then we introduced an iterative anomaly detection method for feature extraction. The iterative method was applied to sea-mine detection in sonar images. We have introduced multi-scale GMRF models and a corresponding anomaly subspace detection algorithm. We developed the algorithm for target detection in cluttered images which follow one of the proposed models. The model was based on a multi-scale representation of the image and the KLT.

We assumed that independent layers in the multi-scale representation can be modeled as GMRFs with different sets of parameters. The detection is carried out by applying a MSD to the innovations process (prediction error) of the multi-scale GMRF. The MSD incorporates the available *a priori* information about the targets into the detection process and thus improves the detection performance. The MSD was originally developed for signal detection in subspace interference

and white Gaussian noise [69]. Here, we have formulated a MSD for signal detection in subspace interference and noise which follows the proposed multi-scale GMRF model.

The performance of the algorithm was demonstrated with application to automatic target detection in synthetic images, side-scan sonar data and wafer images. The results show the capability of the proposed model and algorithm to cope with variety of targets and background clutter patterns. Performance analysis was carried out by investigating the influence of different parameters on the detection performance, and comparing the performance of the proposed method to those of similar methods. The analysis as well as the experimental results demonstrate the advantages of the proposed method.

8.2 Future Research

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

1. Multi-scale stochastic processes modeling.

   The detection method proposed in this thesis, is carried out using independent image layers. Each independent layer is separately modeled. Models of multi-scale stochastic processes (MSPs) use a pyramidal image structure for modeling images in different scales [21], [24], [51], [65], [67], [80]. Figure 8.1 presents the pyramidal structure of a multi-resolution tree, presented by Willsky [80]. The operator γ transforms node $\mathbf{s}$ to its parent $\mathbf{s}\gamma$ in a coarser scale. The pyramid-based multi-scale stochastic process (MSP) models describe the Markovian relation between different levels of resolution. The

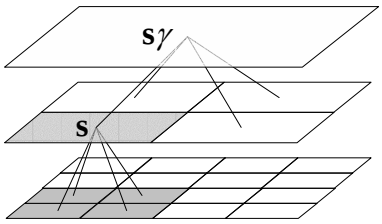


Figure 8.1: A multi-resolution tree. Node $\mathbf{s}\gamma$ is the parent of node $\mathbf{s}$ in a coarser scale.

features at a pixel in one pyramid level condition the features at each of several child pixels at the next higher-resolution pyramid level as follows:

$$y(\mathbf{s}) = V(\mathbf{s})y(\mathbf{s}\gamma) + \varepsilon(\mathbf{s}) \tag{8.1}$$

where $V(\mathbf{s})$ is a matrix and $\varepsilon(\mathbf{s})$ is the innovations process.

Future research may deal with the problem of subspace detection based on a MSP model. The improvement in detection performance, which can be achieved using the statistical relations between different resolution levels of an image, should be examined and analyzed.

2. Definition of the target and interference subspaces within a multi-scale framework.

   The target subspace examples presented in chapter 7 are determined based on target characteristics in a specific scale. The target subspace is generated using the image in its original resolution. The subspace is then transformed to the multi-layer representation in order to be used by the MSD. Defining the subspaces within a multi-resolution framework, should be based on information about the targets in different scales. Defining the subspace separately in each resolution level may incorporate more information about the targets, and may improve the detection performance due to the difference

between the information about the targets, revealed by each resolution level. Future research may deal with the performance improvement which can be achieved by defining the subspaces within a multi-resolution framework.

3. Anomaly subspace detection in presence of non-stationary background.

   When the background pattern changes in space, the stationarity assumption of the proposed model may limit the detection performance. When the change in the model parameters is slow, small areas may be well described by a stationary approximation. Combining the background modeling and the detection process with a segmentation procedure may help to better describe the background patterns and improve the detection performance. Future research may deal with non-stationary background patterns which can be locally approximated by stationary pattern models. A comparative performance analysis may demonstrate the possible advantages in segmentation of the image and separately modeling each of its segments. Dividing the image into smaller areas may justify the stationary approximation. Smaller areas may also reduce the population of pixels, used for estimating the statistical models for the different segments, and therefore may increase the estimation error. This tradeoff should also be examined in future work.

Bibliography

[1] E. A. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive bayesian classifier," *IEEE Trans. Geoscience and Remote Sensing*, vol. 36, pp. 506–517, March 1998.

[2] V. Barnett and T. Lewis, *Outliers in statistical data*. NY, USA: Wiley, 1994.

[3] M. G. Bello, "A random-field model-based algorithm for anomalous complex image pixel detection," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 186–196, April 1992.

[4] J. Bennet and A. Khotanzad, "Modeling textured images using generalized long correlation models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1365–1370, December 1998.

[5] J. W. Bennett, "Modeling and analysis of gray tone, color, and multispectral texture images by random field models and their generalizations," Ph.D. dissertation, Electrical Engineering Department, Southern Methodist Univ., Methodist Univ. Dallas, Tex, May 1997.

[6] L. M. Berliner, C. K. Wikle, and N. Cressie, "Long-lead prediction of pacific sst's via bayesian dynamic modeling," *J. Climate*, vol. 13, pp. 3953–3968, 2000.

[7] I. E. Besag, "Spatial interaction and statistical analysis of lattice systems," *J. Royal Stat. Sot., Ser. B*, vol. B-36, pp. 192–236, 1974.

[8] C. Bishop, "Novelty detection and neural network validation," in *Proceedings of the IEE Conference on Vision and Image Signal Processing*. IEE, 1994, pp. 217–222.

[9] P. Brodatz, *Textures : A Photographic Album for Artists and Designers.* New-York: Dover, 1966.

[10] T. Brotherton, T. Johnson, and G. Chadderdon, "Classification and novelty detection using linear models and class dependent elliptical basis function neural network," in *Proceedings of the IJCNN Conference*. Anchorage: IJCNN, May 1998.

[11] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. COM-31, no. 4, pp. 532–540, April 1983.

[12] G. A. Carpenter, M. A. Rubin, and W. W. Streilein, "Artmap-fd: familiarity discrimination applied to radar target recognition," in *Proceedings of the International Conference on Neural Networks*, vol. 3, Houston, TX, 1997, pp. 1459–1464.

[13] R. Chellappa, "Two-dimensional discrete gaussian markov random field models for image processing," *Progress in Pattern Recognition*, vol. 2, pp. 79–112, 1985.

[14] R. Chellappa and S. Chatterjee, "Classification of textures using gaussian markov random fields," *IEEE Trans. ASSP*, vol. 33, pp. 959–963, 1985.

[15] R. Chellappa and R. L. Kashyap, "Synthetic generation and estimation in random field models of images," in *Proc. IEEE CS Conf. Pattern Recognition and Image Processing*, Dallas Tex., August 1981, pp. 577–582.

[16] ——, "Texture synthesis using 2-d noncausal autoregressive models," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 1, pp. 194–203, Februaly 1985.

[17] H. Cheng and C. A. Bouman, "Multiscale bayesian segmentation using a trainable context model," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 511–525, April 2001.

[18] D. A. Clausi, "Texture segmentation on sar ice imagery," Ph.D. dissertation, University of Waterloo, Canada, 1996.

[19] ——, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.

[20] D. A. Clausi and B. Yue, "Comparing cooccurrence probabilities and markov random fields for texture analysis of sar sea ice imagery," *IEEE trans. on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 215–228, January 2004.

[21] H. Coi and R. G. Baraniuk, "Multiscale image segmentation using waveletdomain hidden markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1309–1321, 2001.

[22] C. A. H. R. W. Conners, "Theoretical comparison of texture algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 3, pp. 204–222, May 1980.

[23] G. R. Cross and A. K. Jain, "Markov random field texture models," in *Proceedings of the IEEE Comput. Sot. Conf. Pattern Recog. Image Proc.* Dallas, TX: IEEE, August 1981, pp. 597–601.

[24] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[25] M. Desforges, P. Jacob, and J. Cooper, Eds., *Applications of probability density estimation to the detection of abnormal conditions in engineering*, vol. 212. Institute of Mechanical Engineers, 1998.

[26] G. J. Dobeck, J. C. Hyland, and L. Smedley, "Automated detection/classification of sea mines in sonar imagery," in *Proc. SPIE*, vol. 3079, Orlando, Florida, April 1997, pp. 90–110.

[27] M. Elad, *Image Processing and Analysis*, Technion city, Haifa, Israel, 2003, textbook of an image processing course Ch.11, pp. 235–256.

[28] K. Eom, "Long-correlation image models for textures with circular and elliptical correlation structures," in *ICIP01*, 2001, pp. I: 197–200.

[29] K. B. Eom, "Long-correlation image models for texture with circular and elliptical correlation structures," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1047–1055, July 2001.

[30] P. W. Fieguth, W. C. Karl, A. S.Willsky, and C.Wunsch, "Multiresolution optimal interpolation and statistical analysis of topex/poseidon satellite altimetry," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 280–292, March 1995.

[31] C. H. Fosgate, H. Krim, W. W. Irving, W. C. Karl, and A. S. Willsky, "Multiscale segmentation and anomaly enhancement of sar imagery," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 7–20, January 1997.

[32] D. Fradkin and D. Madigan, "Experiments with random projections for machine learning," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press New York, NY, USA, 2003, pp. 517–522.

[33] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Academic Press, 1990, ch. 3, pp. 66–77.

[34] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Trans. PAMI*, vol. 6, no. 6, pp. 194–207, November 1984.

[35] M. Hassner and J. Sklansky, "The use of markov random fields as models of textures," *Comput. Graphics Image Proc.*, vol. 12, pp. 357–370, April 1980.

[36] G. G. Hazel, "Multivariate gaussian GMRF for multispectral scene segmantation and anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1199–1211, May 2000.

[37] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspetive*, 2nd ed. Prentice Hall, Inc., 1996.

[38] R. A. Johnson and D. W. Wichem, *Applied Multi-variate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1998.

[39] K. V. R. Kanth, D. Agrawal, A. E. Abbadi, and A. Singh, "Dimensionality reduction for similarity searching in dynamic databases," University of California, Technical report 1998-10, May 1998.

[40] R. L. Kashyap, *Image Modeling*. New York: Academic, 1980, ch. Univariate and multivariate random field models for images.

[41] R. L. Kashyap and R. Chellappa, "Estimation and choice of neighbors in spatial-interaction models of images," *IEEE Transactions on Information Theory*, vol. IT-29, no. 1, pp. 60–72, January 1983.

[42] R. L. Kashyap and K. B. EOM, "A consistent estimation of parameters in the long memory time series model," *Journal of Time Series Analysis*, vol. 9, no. 1, pp. 35–41, 1988.

[43] ——, "Texture boundary detection based on long correlation model," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. 11, no. 1, pp. 58–67, 1989.

[44] R. L. Kashyap and P. M. Lapsa, "Sythesis and estimation of random fields using long correlation models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 800–809, Nov. 1984.

[45] E. J. Kelly, "Adaptive detection and parameter estimation for multidimensional signal models," MIT Lincoln LAboratory, Lexington, MA, Tech. Rep. 848, April 1989.

[46] C. Kervrann and F. Heitz, "A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics," *IEEE Transactions on Image Processing*, vol. 4, no. 6, pp. 856–862, June 1995.

[47] P. Kumar, "A multiple scale state-space model for characterizing subgrid scale variability of near-surface soil moisture," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 182–197, 1999.

[48] K. V. Leemput, F. Maes, D. Vandermeulen, A. Colchester, and P. Suetens, "Automated segmentation of multiple sclerosis lesions by model outlier detection," *IEEE Trans. on Medical Imaging*, vol. 20, no. 8, pp. 677–688, August 2001.

[49] P. Lévi, "A special problem of brownian motion, and a general theory of gaussian random functions," in *3rd Berkeley Symp. Mathematical Statistics and Probability*, vol. 2, Berkeley California. California Press, 1956.

[50] S. Lovejoy and D. Schartzer, "Generalized scale invariance in the atmosphere and fractal models of rain," *Water Resources Res.*, vol. 21, pp. 1233–1250, August 1985.

[51] M. R. Luettgen and A. S. Willsky, "Likelihood calculation for a class of multiscale stochastic models, with application to texture discrimination," *IEEE Trans. Image Proc.*, vol. 4, no. 2, pp. 194–207, 1995.

[52] C. Manikopoulos and S. Papavassiliou, "Network intrusion and fault detection: a statistical anomaly approach," *IEEE Communications Magazine*, vol. 40, pp. 76–82, October 2002.

[53] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, pp. 29–43, January 2002.

[54] D. Manolakis, G. Shaw, and N. Keshava, "Comparative analysis of hyperspectral adaptive matched filter detectors," in *SPIE, Algorithms for Multispectral, Hyperspectral, and Ultrasepctral Imagery VI*, S. Shen and M. Descour, Eds., vol. 4049. SPIE, April 2000, pp. 2–17.

[55] G. Manson, G. Pierce, and K. Worden, "On the long tern stability of normal condition for damage detection in a composite panel," in *Proceedings of the 4th international conference on damage assessment of Structures (DAMAS)*. Cardiff, UK, June 2001, pp. 359–369.

[56] G. Manson, S. G. Pierce, K. Worden, T. Monnier, P. Guy, and K. Atherton, "Long-term stability of normal condition data for novelty detection," in *Smart Structures and Materials 2000: Smart Structures and Integrated Systems*, N. M. Wereley, Ed., vol. 3985. SPIE, June 2000, pp. 323–334.

[57] M. Markou and S. Singh, "Novelty detection: A review - part 1: Statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, July 2003.

[58] M. Mignotte and C. Collet, "Three-class markovian segmentation of high-resolution sonal images," *Computer Vision and Image Understanding*, vol. 76, no. 3, pp. 191–204, December 1999.

[59] P. A. P. Moran and J. E. Besag, "On the estimation and testing of spatial interaction in gaussian lattices," *Biometrika*, vol. 62, no. 3, pp. 555–562, 1975.

[60] L. Parra, G. Deco, and S. Miesbach, "Statistical independence and novelty detection with information preserving non-linear maps," *Neural Comput.*, vol. 2, no. 8, pp. 260–269, 1995.

[61] N. J. Pizzi, R. A. Vivanco, and R. L. Somorjai, "Evident: a functional magnetic resonance image analysis system," *Artificial Intelligence in Medicine*, vol. 21, no. 1-3, pp. 263–269, January-March 2001.

[62] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 38, pp. 1760–1770, October 1990.

[63] 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*, vol. 28, no. 1, pp. 90–105, January 2003.

[64] S. Roberts and L. Tarassenko, "A probabilistic resource allocating network for novelty detection," *Neural Comput.*, no. 6, pp. 270–284, 1994.

[65] J. K. Romberg, H. Coi, and R. G. Baraniuk, "Bayesian tree-stuctured image modeling using wavelet domain hidden markov models," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, 2001.

[66] A. Rosenfeld, "Image modeling during the 1980's: a brief overview. in markov random fields: Theory and applications," vol. 2, 1993.

[67] P. Sajda, C. Spence, and L. Parra, "A multi-scale probabilistic network model for detection, synthesis and compression in mammographic image analysis," *Medical Image Analysis*, vol. 7, no. 2, pp. 187–204, June 2003.

[68] A. Sarkar, M. K. Biswas, and K. M. S. Sharma, "A simple unsupervised GMRF model based image segmentation approach," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 801–811, May 2000.

[69] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2156, August 1994.

[70] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaption in anomaly detection," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1855–1871, August 2000.

[71] ——, "Efficient detection in hyperspectral imagery," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 584–597, April 2001.

[72] A. Speis and G. Healey, "Feature extraction for texture discrimination via random field models with random spatial interaction," *IEEE Transactions on Image Processing*, vol. 5, no. 4, pp. 635–645, April 1996.

[73] 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*, pp. 58–69, January 2002.

[74] L. Tarassenko, "Novelty detection for the identification of masses in mammograms," in *Proceedings of the 4th IEE International Conference on Artificial Neural Networks*, vol. 4. Cambridge, UK: IEE, 1995, pp. 442–447.

[75] L. Tarassenko, A. Nairac, N. Townsend, and P. Cowley, "Novelty detection in jet engines," in *IEE Colloquium on Condition Monitoring, Imagery, External Structures and Health*, Birmingham, UK, 1999, pp. 41–45.

[76] D. M. J. Tax and R. P. W. Duin, "Outlier detection using classifier instability," in *Advances in Pattern Recognition - The joint IAPR International Workshops*. Sydney, Australia: IAPR, 1998, pp. 593–601.

[77] L. van Kempen and H. Sahli, "Signal processing techniques for clutter parameters estimation and clutter removal in GPR data for landmine detection," in *Proc. 11th IEEE Workshop on Statistical Signal Processing*, Singapure, August 2001, pp. 158–161.

[78] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. John Wiley and Sons, July 2002.

[79] R. A. Weisenseel, W. C. K. D. A. Castanon, and C. A. Dimarzio, "Statistical approach to multi-channel spatial modeling for the detection of mine-like targets," in *Proc. SPIE, Detection and Remediation Technologies for Mines and Mine-Like Targets*, vol. 3392, April 1998, pp. 1044–1053.

[80] A. S. Willsky, "Multiresolution markov models for signal and image processing," in *Proceedings of the IEEE*, vol. 90, no. 8, August 2002, pp. 1396–1458.

[81] E. Wong, "Two-dimensional random fields and representation of images," *SIAM J. Appl. Math.*, vol. 16, pp. 756–770, 1968.

[82] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Transactions on Information Theory*, vol. IT-18, no. 2, pp. 101–109, March 1972.

[83] C. Wunsch and D. Stammer, "The global frequency-wavenumber spectrum of oceanic variability estimated from topex/poseidon altimetric measurements," *J. Geophys. Res.*, vol. 100, pp. 24,89524,910, 1995.

[84] Y. Yang and X. Liu, "A re-examination of text categorization methods," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM Press, 1999, pp. 42–49.

[85] D. Y. Yeung and C. Chow, "Parzen window network intrusion detectors," in *Proceedings of the International Conference on Pattern Recognition*, Quebec, Canada, 2002.

[86] B. Yue, "Sar sea ice recognition using texture methods," Master's thesis, University of Waterloo, Ontario, Canada, 2002.

[87] P. Zhao and J. Litva, "Consistency of modified LS estimation method for identifying 2-D noncausal SAR model parameters," *IEEE Transactions on Automatic Control*, vol. 40, no. 2, pp. 316–320, February 1995.