TWO-DIMENSIONAL GARCH MODEL WITH APPLICATION TO ANOMALY DETECTION

Amir Noiboar and Israel Cohen

Department of Electrical Engineering, Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel
noiboar@tx.technion.ac.il; icohen@ee.technion.ac.il

ABSTRACT

In this paper, we introduce a two-dimensional Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model for clutter modeling and anomaly detection. The one-dimensional GARCH model is widely used for modeling financial time series. Extending the one-dimensional GARCH model into two dimensions yields a novel clutter model which is capable of taking into account important characteristics of natural clutter, namely heavy tailed distribution and innovations clustering. We show that the two-dimensional GARCH model generalizes the casual Gauss Markov random field (GMRF) model, and develop a matched subspace detector (MSD) for detecting anomalies in GARCH clutter. Experimental results demonstrate that a reduced false alarm rate can be achieved without compromising the detection rate by using an MSD under GARCH clutter modeling, rather than GMRF clutter modeling.

1. INTRODUCTION

Anomaly detection is the process of detecting a portion of the data, which differs in some statistical sense from the background clutter. Anomaly detection is a well-studied problem with many practical applications including detection of targets in images [1–4], detection of defects in silicon wafers [5], detection of faults in seismic data [6], etc. The greatest factor in anomaly detection is the choice of an adequate statistical model for the clutter, which would enable to discriminate between anomalies and the clutter components. Unfortunately, clutter modeling is often obtained by utilizing a Gauss Markov random field (GMRF) model [6, 7], which is insufficient in its capability to model natural clutter. To overcome this limitation, and in order to decrease the false alarm rate while retaining the desired detection rate, various methods have been proposed. Among these methods we find multiscale representations for detecting anomalies in different scales [5, 8], performing segmentation as a preprocessing stage to anomaly detection [2], utilizing *a priori* information about the shape, size, and other characteristics of the anomaly [9], applying sliding windows to the image through which the model estimation and anomaly detection can be carried out locally [10], etc.

In this paper, we introduce a two-dimensional Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model for clutter modeling and anomaly detection. The one-dimensional GARCH model [11] is widely used for modeling financial time series. Extending the one-dimensional GARCH model into two dimensions yields a novel clutter model which is capable of taking into account important characteristics of natural clutter, namely heavy tailed distribution and innovations clustering. We show that the two-dimensional GARCH model generalizes the casual GMRF model, and develop a matched subspace detector (MSD) for detecting anomalies in GARCH clutter. The performance of the proposed anomaly detection method is evaluated on synthetic and real data. Experimental results demonstrate that reduced false alarm rate can be achieved without compromising the detection rate by using an MSD under GARCH clutter modeling, rather than GMRF clutter modeling.

The paper is organized as follows: In Sec. 2 we introduce the two-dimensional GARCH model. In Sec. 3 we address the model estimation problem. In Sec. 4 we develop an MSD for detecting anomalies in GARCH clutter. Finally, in Sec. 5 we demonstrate the advantage of the proposed anomaly detection approach over using the GMRF clutter modeling.

2. TWO DIMENSIONAL GARCH MODEL

Let $q_1, q_2, p_1, p_2 \geq 0$ denote the order of the GARCH model, and let Γ_1 and Γ_2 denote two neighborhood sets which are defined by

$$\begin{aligned} \Gamma_1 &= \{k\ell \mid 0 \leq k \leq q_1,\, 0 \leq \ell \leq q_2,\, (k\ell) \neq (0,0)\} \\ \Gamma_2 &= \{k\ell \mid 0 \leq k \leq p_1,\, 0 \leq \ell \leq p_2,\, (k\ell) \neq (0,0)\}\,. \end{aligned}$$

Let ϵ_{ij} represent a 2D stochastic process, and let h_{ij} denote its variance conditioned upon the information set $\psi_{ij} = \{\{\epsilon_{i-k,j-\ell}\}_{k\ell\in\Gamma_1}, \{h_{i-k,j-\ell}\}_{k\ell\in\Gamma_2}\}$. Define the 2D neighborhood of location ij as: $\Gamma = \{k\ell \mid k \leq i,\, \ell \leq j\}$ and let $\eta_{ij} \overset{iid}{\sim} N(0,1)$ be a stochastic 2D process independent of $h_{k\ell},\ \forall k\ell \in \Gamma$. The 2D $GARCH(p_1, p_2, q_1, q_2)$ process is defined as:

$$\epsilon_{ij} = \sqrt{h_{ij}}\,\eta_{ij} \tag{1}$$

$$h_{ij} = \alpha_0 + \sum_{k\ell\in\Gamma_1} \alpha_{k\ell}\epsilon^2_{i-k,j-\ell} + \sum_{k\ell\in\Gamma_2} \beta_{k\ell} h_{i-k,j-\ell}\,, \tag{2}$$

and is therefore conditionally distributed as:

$$\epsilon_{ij} \mid \psi_{ij} \sim N(0, h_{ij})\,. \tag{3}$$

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

$$\begin{aligned} \alpha_0 &> 0 \\ \alpha_{kl} &\geq 0, \quad k\ell \in \Gamma_1 \\ \beta_{kl} &\geq 0, \quad k\ell \in \Gamma_2\,. \end{aligned} \tag{4}$$

From (2) we see that at every spatial location, both the neighboring sample variances and the neighboring conditional variances play a role in the current conditional variance. This yields clustering of variations, which is an important characteristic of the GARCH process. Note that if $q_1 = q_2 = p_1 = p_2 = 0$ then ϵ_{ij} is simply white Gaussian noise (WGN).

It is shown in [11] that a sufficient condition for wide sense stationarity of the 1D GARCH process is that the sum of all parameters is smaller than one. A similar result is obtained in the 2D case as we prove in the following theorem.

Theorem 1: The $GARCH(p_1, p_2, q_1, q_2)$ process as defined in (1) and (2) is wide-sense stationary with:

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

if and only if $\sum_{k\ell\in\Gamma_1} \alpha_{k\ell} + \sum_{k\ell\in\Gamma_2} \beta_{k\ell} < 1$.

Proof: Substituting (1) into (2) yields:

$$h_{ij} = \alpha_0 \sum_{g=0}^{\infty} M(i, j, g)\,, \tag{5}$$

where $M(i, j, g)$ involves all terms of the form:

$$\prod_{k\ell\in\Gamma_1} \alpha_{k\ell}^{a_{k\ell}} \prod_{k\ell\in\Gamma_1} \beta_{k\ell}^{b_{k\ell}} \prod_{r=1}^{n} \eta^2_{(ij)-s_r}\,,$$

for

$$\sum_{k\ell\in\Gamma_1} a_{k\ell} + \sum_{k\ell\in\Gamma_2} b_{k\ell} = g\,;\ \sum_{k\ell\in\Gamma_1} a_{k\ell} = n$$

and

$$\begin{aligned} 0 &\leq |s_1| \leq |s_2| \leq \cdots \leq |s_n| \\ s_r &\equiv (s_{r_i}, s_{r_j}) \\ s_{r_i} &\leq \max\{kq_1, (g-1)q_1 + p_1\} \\ s_{r_j} &\leq \max\{kq_2, (g-1)q_2 + p_2\}\,. \end{aligned}$$

In general

$$\begin{aligned} M(i, j, g) &= \sum_{k\ell\in\Gamma_1} \alpha_{k\ell}\eta^2_{i-k,j-\ell} M(i-k, j-\ell, g) + \\ &\quad + \sum_{k\ell\in\Gamma_2} \beta_{k\ell} M(i-k, j-\ell, g)\,. \end{aligned} \tag{6}$$

Since η_{ij} is i.i.d., the moments of $M(i, j, g)$ are independent of (ij), and in particular

$$E\{M(i, j, g)\} = E\{M(k, \ell, g)\}\ \forall ijk\ell g\,. \tag{7}$$

From (6) and (7) we have:

$$E\{M(i, j, g+1)\} = \left[\sum_{k\ell\in\Gamma_1} \alpha_{k\ell} + \sum_{k\ell\in\Gamma_2} \beta_{k\ell}\right]^{g+1}. \tag{8}$$

Finally by (5) and (8):

$$E\{\epsilon^2_{ij}\} = \alpha_0 \left[1 - \sum_{k\ell\in\Gamma_1} \alpha_{k\ell} - \sum_{k\ell\in\Gamma_2} \beta_{k\ell}\right]^{-1},$$

if and only if

$$\sum_{k\ell\in\Gamma_1} \alpha_{k\ell} + \sum_{k\ell\in\Gamma_2} \beta_{k\ell} < 1\,.$$

$E(\epsilon_{ij}) = 0$ and $\mathrm{cov}(\epsilon_{\mathrm{ij}}, \epsilon_{\mathrm{k}\ell}) = 0,\ \forall (\mathrm{ij}) \neq (\mathrm{k}\ell)$ follows immediately.

3. MODEL ESTIMATION

In this section we find a maximum likelihood estimate for the GARCH model. We let ϵ_{ij} be innovations of a 2D linear regression, where y_{ij} is the dependent variable, $\boldsymbol{x}_{ij}$ a vector of explanatory variables and $\boldsymbol{b}$ a vector of unknown parameters:

$$\epsilon_{ij} = y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{b}\,, \tag{9}$$

If ϵ_{ij} in (9) is WGN the regression model is a casual GMRF. This is a special case of the GARCH process. Using (9) we can write (2) as:

$$\begin{aligned} h_{ij} &= \alpha_0 + \sum_{k,\ell\in\Gamma_1} \alpha_{k\ell}(y_{i-k,j-\ell} - \boldsymbol{x}_{i-k,j-\ell}^T \boldsymbol{b})^2 + \\ &\quad + \sum_{k\ell\in\Gamma_2} \beta_{k\ell} h_{i-k,j-\ell}\,. \end{aligned} \tag{10}$$

The conditional distribution of y_{ij} is Gaussian with mean $\boldsymbol{x}_{ij}^T\boldsymbol{b}$ and variance h_{ij}:

$$f(y_{ij} \mid \boldsymbol{x}_{ij}, \psi_{ij}) = \frac{1}{\sqrt{2\pi h_{ij}}} \exp\left(\frac{-(y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{b})^2}{2h_{ij}}\right). \tag{11}$$

Let

$$\begin{aligned} \boldsymbol{z}_{ij}^T &= [1, \epsilon^2_{i-1,j}, \ldots, \epsilon^2_{i-q_1,j-q_2}, h_{i-1,j}, \ldots, h_{i-p_1,j-p_2}] = \\ &= [1, (y_{i-1,j} - x_{i-1,j}^T b)^2, \ldots, (y_{i-q_1,j-q_2} - x_{i-q_1,j-q_2}^T b)^2, \\ &\quad h_{i-1,j}, \ldots, h_{i-p_1,j-p_2}] \end{aligned}$$

and let

$$\boldsymbol{\delta}^T = [\alpha_0, \alpha_{0,1}, \ldots, \alpha_{q_1,q_2}, \beta_{0,1}, \ldots, \beta_{p_1,p_2}]\,,$$

then (10) can be written as:

$$h_{ij} = [\boldsymbol{z}_{ij}(\boldsymbol{b})]^T \boldsymbol{\delta}\,. \tag{12}$$

The unknown parameters are collected into a column vector $\boldsymbol{\theta} = [\boldsymbol{b}^T, \boldsymbol{\delta}^T]^T$. Define the sample space Ω as a two dimensional lattice of size $N \times M$: $\Omega = \{ij \mid 1 \leq i \leq N,\ 1 \leq j \leq M\}$. The conditional sample log likelihood is:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \sum_{ij\in\Omega} \log f(y_{ij} \mid \boldsymbol{x}_{ij}, \psi_{ij}) = \\ &= -\frac{1}{2}[(N+M)log(2\pi) + \sum_{ij\in\Omega} \log([\boldsymbol{z}_{ij}(\boldsymbol{b})]^T\boldsymbol{\delta}) + \\ &\quad + \sum_{ij\in\Omega} (y_{ij} - \boldsymbol{x}_{ij}^T\boldsymbol{b})^2 / ([\boldsymbol{z}_{ij}(\boldsymbol{b})]^T\boldsymbol{\delta})]\,. \end{aligned} \tag{13}$$

The parameter vector $\boldsymbol{\theta}$ is found by numerically solving a constrained maximization problem on the log likelihood function with respect to the unknown parameters (see for example [12]). The constraints used are those presented in (4) and in Theorem 1. To solve the maximization problem knowledge of values of ϵ_{ij} and

h_{ij} where $i, j \leq 0$ is required. As in [11] we set these boundary conditions as:

$$\epsilon_{ij} = h_{ij} = \frac{1}{NM} \sum_{k\ell \in \Omega} (y_{ij} - \boldsymbol{x}_{ij}^T \boldsymbol{b})^2; \; \forall ij \leq 0 \,. \quad (14)$$

4. ANOMALY DETECTION

In this section we develop our anomaly detection approach, which is based on modeling the image as a 2D causal autoregressive model with GARCH innovations. We assume that the anomalies are sparse within the image and therefore their influence on the model estimation and on the estimated conditional variance field is negligible. Model estimation is performed as described in Section 3. The conditional variance field h_{ij} is calculated based on the estimated model parameters and is later used in our detection process. The anomalies are 2-dimensional with a spatial size $K \times L$ which is much smaller than $N \times M$. The anomalies are assumed to lie within a known subspace spanned by G image chips $w_g, g = 1, 2, \cdots, G$, each of size $K \times L$. A matrix H, whose columns span the anomaly subspace is created by row stacking every image chip into a column vector and setting these vectors as columns in H.

We model the interference subspace in a similar manner using T image chips $s_t, t = 1, 2, \cdots, T$, each of size $K \times L$. A matrix spanning the interference subspace S is created.

In [9] an MSD is developed for the detection of subspace signals in subspace interference and WGN. Here we develop a MSD for the detection of subspace signals in GARCH clutter and subspace interference.

Let $y(s)$ represent a pixel at spatial location s. For each pixel $y(s)$ we create a column vector $\boldsymbol{y}(s)$ by row stacking an image chip of size $K \times L$ centered around s. Let $\boldsymbol{\epsilon}(s)$ be a result of row stacking a chip of a GARCH field of size $K \times L$ centered around s. Similarly let $\boldsymbol{x}$(s) be a vector representing the explanatory variable field ($\boldsymbol{x}_{ij}\boldsymbol{b}$) in the $K \times L$ neighborhood of s. Let $\boldsymbol{\phi}(s)$, $\boldsymbol{\psi}(s)$ be the weight vectors for the interference and anomaly subspaces respectively. We define two hypotheses, H_0 and H_1, which represent absence and respectively presence of an anomaly:

$$\begin{aligned} H_0 : \boldsymbol{y}(s) &= S\boldsymbol{\phi}(s) + \boldsymbol{x}(s) + \boldsymbol{\epsilon}(s) \\ H_1 : \boldsymbol{y}(s) &= H\boldsymbol{\psi}(s) + S\boldsymbol{\phi}(s) + \boldsymbol{x}(s) + \boldsymbol{\epsilon}(s) \,. \end{aligned} \quad (15)$$

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

$$\begin{aligned} H_0 : \boldsymbol{y}(s) &\sim N(S\boldsymbol{\phi}(s) + \boldsymbol{x}(s), \Sigma(s)) \\ H_1 : \boldsymbol{y}(s) &\sim N(H\boldsymbol{\psi}(s) + S\boldsymbol{\phi}(s) + \boldsymbol{x}(s), \Sigma(s)) \,. \end{aligned}$$

Define P_S as the projection into the subspace spanned by the columns of S, and define P_{HS} as the projection into the subspace spanned by the columns of the concatenated matrix $[HS]$, that is:

$$\begin{aligned} P_S &= S(S^T S)^{-1} S^T \\ P_{HS} &= [HS] \left([HS]^T [HS] \right)^{-1} [HS]^T \,. \end{aligned} \quad (16)$$

From (15) and (16) we find the GARCH innovations field under the two hypothesis:

$$\begin{aligned} H_0 : \boldsymbol{\epsilon}_0(s) &= \boldsymbol{y}(s) - \boldsymbol{x}(s) - S\boldsymbol{\phi}(s) = \\ &= (I - P_S)[\boldsymbol{y}(s) - \boldsymbol{x}(s)] \quad (17) \\ H_1 : \boldsymbol{\epsilon}_1(s) &= \boldsymbol{y}(s) - \boldsymbol{x}(s) - S\boldsymbol{\phi}(s) - H\boldsymbol{\psi}(s) = \\ &= (I - P_{HS})[\boldsymbol{y}(s) - \boldsymbol{x}(s)] \,. \quad (18) \end{aligned}$$

The conditional likelihood function of $\boldsymbol{\epsilon}$ under the two hypothesis is:

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

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

The generalized likelihood ratio (GLR) is:

$$L(s) = 2 \log \left(\frac{\ell_1(s)}{\ell_0(s)} \right) \,, \quad (19)$$

which yields:

$$\begin{aligned} L(s) &= \boldsymbol{\epsilon}_0(s)^T \Sigma(s)^{-1} \boldsymbol{\epsilon}_0(s) - \boldsymbol{\epsilon}_1(s)^T \Sigma(s)^{-1} \boldsymbol{\epsilon}_1(s) = \\ &= [\boldsymbol{\epsilon}_0(s) - \boldsymbol{\epsilon}_1(s)]^T \Sigma(s)^{-1} [\boldsymbol{\epsilon}_0(s) - \boldsymbol{\epsilon}_1(s)] = \\ &= [(P_{HS} - P_S)(\boldsymbol{y}(s) - \boldsymbol{x}(s))]^T \, \Sigma(s)^{-1} \\ &\quad \times [(P_{HS} - P_S)(\boldsymbol{y}(s) - \boldsymbol{x}(s))] \,. \end{aligned} \quad (20)$$

Define the ratio between the energy of the signal which does not lie in the interference subspace and the innovations' conditional variance as the signal to noise ratio (SNR):

$$SNR = [(H\boldsymbol{\psi}(s))(I - P_S)]^T \, \Sigma(s)^{-1} \, [(H\boldsymbol{\psi}(s))(I - P_S)] \,. \quad (21)$$

The GLR is a sum of squared conditionally independent normally distributed variables and therefore is conditionally chi-square distributed with $u = rank(H)$ degrees of freedom:

$$\begin{aligned} H_0 &: L(s) \sim \chi_u^2(0) \\ H_1 &: L(s) \sim \chi_u^2(SNR) \,. \end{aligned}$$

Under hypothesis H_1, the non-centrality parameter of the chi-square distribution of $L(s)$ is equal to the SNR [9]. Detection is performed by applying a threshold η to the GLR: $L(s) \underset{H_0}{\overset{H_1}{\gtrless}} \eta$. The threshold is determined by the tradeoff between the desired conditional detection and false alarm rates, where these can be calculated by:

$$\begin{aligned} P_{FA} &= 1 - P \left[\chi_u^2(0) \leq \eta \right] \quad (22) \\ P_D &= 1 - P \left[\chi_u^2(SNR) \leq \eta \right] \,. \quad (23) \end{aligned}$$

The following describes the steps taken in order to detect anomalies within an region of interest (ROI):

1. Select an ROI in the image and a threshold η using (22).
2. Estimate the unknown parameters using the model estimation method described in Section 3.

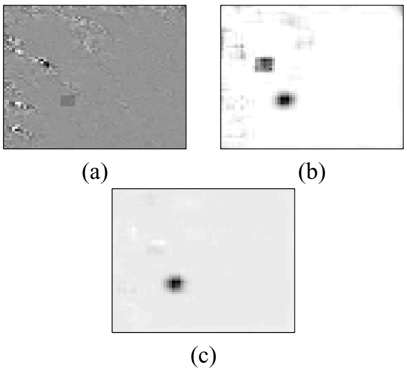

Fig. 1. Anomaly detection in synthetic GARCH clutter: (a) Original image with embedded rectangular anomaly; (b) GLR based on GMRF clutter modeling; (c) GLR based on GARCH clutter modeling;

3. Calculate the conditional variance field $\boldsymbol{h}(s)$ using (12).
4. Calculate the GARCH innovation fields $\boldsymbol{\epsilon}_k(s)$, $k \in \{0, 1\}$. for every spatial location s, using (17),(18)
5. Find the GLR for every spatial location s using (20).
6. Compare the GLR to the chosen threshold to achieve the anomaly detection image.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we demonstrate the performance of our anomaly detection approach on synthetic and real data, and show the advantage of using GARCH clutter modeling compared to using GMRF modeling.

5.1. Synthetic Data

We use synthetic image data, which was generated using a GARCH(1,1,1,1) model with the following parameters: $\alpha_0 = 0.002, \alpha_{0,1} = 0.3, \alpha_{1,0} = 0.25, \alpha_{1,1} = 0.1$, $\beta_{0,1} = 0.03, \beta_{1,0} = 0.02, \beta_{1,1} = 0.04$, $\boldsymbol{b} = [0.04, 0.03, 0.05]^T$ and $\boldsymbol{x}_{ij} = [y_{i,j-1}, y_{i-1,j}, y_{i-1,j-1}]^T$. A 5×5 anomaly is planted in the synthetic image. Figure 1(a) shows the synthetic image with the inserted anomaly. The anomaly does not stand out as much as the clustered variations to its upper left. We set the anomaly size to $K = L = 7$ and create an anomaly subspace using a single image chip. No interference subspace is assumed. We perform parameter estimation as described in Section 3 and anomaly detection as detailed in Section 4. Figure 1(b) shows the GLR when performing an MSD based anomaly detection in a strongly casual GMRF clutter. Figure 1(c) shows the GLR obtained by (20) when modeling the image as a GARCH process. Typically, as demonstrated in Figure 1, the MSD under GARCH modeling yields a lower false alarm rate than under GMRF modeling.

5.2. Real Data

We now demonstrate the performance of our anomaly detection approach on a real silicon wafer image. Figure 2(a) shows an image of a silicon wafer containing a manufacturing defect. Figures 2(b) and 2(c) show the detection results using a GMRF clutter model and a GARCH clutter model, respectively. We used an anomaly subspace constructed from a single image chip of size 5×5. Note that we did not use training images of typical defects to create the anomaly subspace. The advantage of GARCH modeling over GMRF modeling is clearly evident.

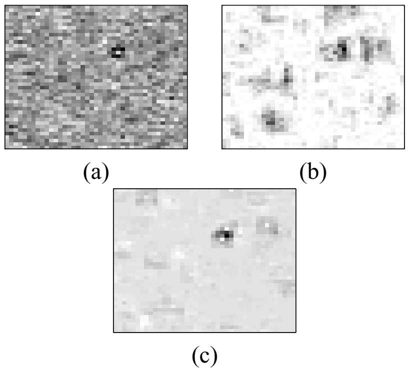

Fig. 2. Silicon wafer defect detection: (a) Original image; (b) GLR based on GMRF clutter modeling; (c) GLR based on GARCH clutter modeling;

6. REFERENCES

[1] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1760–1770, October 1990.

[2] E. A. Ashton, "Detection of subpixel anomalies in multispectral infrared imagery using an adaptive Bayesian classifier," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, no. 2, pp. 506–517, March 1998.

[3] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 58–69, January 2002.

[4] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Mag.*, vol. 19, no. 1, pp. 29–43, January 2002.

[5] A. Goldman and I. Cohen, "Anomaly subspace detection based on a multi-scale Markov random field model," *to appear in Signal Processing*.

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

[7] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inform. Theory*, vol. 46, no. 5, pp. 1855–1871, August 2000.

[8] R. N. Strickland and H. I. Hahn, "Wavelet transform methods for object detection and recovery," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 724–735, May 1997.

[9] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 2146–2157, August 1994.

[10] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Processing*, vol. 10, no. 4, pp. 584–597, April 2001.

[11] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

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