

Breast Cancer Detection Using Communications Technology

Hacer Varol¹, Nidhal Bouaynaya² and William O'Neill³

¹ Dept. of Applied Science, University of Arkansas at Little Rock

² Dept. of Systems Engineering, University of Arkansas at Little Rock

³ Department of Bioengineering, University of Illinois at Chicago

Abstract. In this paper, we formulate the breast cancer diagnosis as a signal detection problem, where the signal of interest is the tumor and the noise refers to the healthy tissue in the breast. In communication engineering, it is known that the optimal detector of a known signal in noise is the North filter. However, the tumor characteristics vary between patients and stages of the disease. We show that, for an unknown signal of interest, the optimal filter is given by the Wiener filter correlator. We apply and compare both filters to cancerous and healthy ultrasound breast images. We show that the North filter provides 83% accuracy, and the estimator correlator achieves 91% accuracy.

Keywords: Breast Cancer, Estimator-Correlator, Matched Filter, North Filter

1. Introduction

In this paper, we use a communication engineering technology, first developed for radar signal identification, to detect tumors in ultrasound breast images. In radar, we are interested in determining the presence or absence of an approaching aircraft. This decision making problem is tackled as a hypothesis testing problem where the signal detected by the radar is classified as either an aircraft reflection signal or noise due to ambient radiation and the receiver electronics. Similarly, we formulate the tumor detection problem in breast images as a hypothesis testing problem, where the signal of interest is the tumor. The proposed detection system will help the radiologist differentiate between benign and cancerous images. Probabilities of detection, error, false alarm, sensitivity, specificity and accuracy are the evaluation criteria of the system.

There are a number of different techniques available to detect breast tumors, which were classified in [1] as geometrical based methods, learning-based methods, pyramidal-based or multi-resolution techniques, and model-based techniques. Except for the model-based techniques, all other methods assume that the imaging breast data is deterministic. In other words, the techniques' success depends on the quality of the healthy and cancerous images. On the other hand, statistical methods weight the noise factor while taking a decision about the existence or absence of a tumor in the image. It is therefore surprising that relatively few researchers have attempted to investigate the efficiency of statistical-based methods as computer-aided diagnostic (CAD) paradigms for breast cancer detection. In [1], the authors proposed a CAD system for tumor detection and classification (cancerous vs. benign) in ultrasound breast images based on a two-dimensional Auto-Regressive-Moving-Average (ARMA) model of the breast image. The estimated 2D ARMA parameters are used as the basis for statistical inference and biophysical interpretation of the breast image. In [2], the detection of microcalcifications in digital mammograms is formulated as a statistical change detection problem in the local properties of the image. A generalized likelihood ratio test is used to detect these changes and estimate their location in the image.

2. Methodology

2.1 The known template

The Matched Filter: In telecommunication engineering, a received waveform is often examined for the presence or absence of a signal of interest. In radar, for instance, we are interested in the presence or absence of an aircraft. It is the function of the signal processor to decide whether the received waveform consists of noise only (no aircraft) or an echo in noise (aircraft present). When an echo is present, the received signal is somewhat different from the transmitted waveform, although possibly not by much. This is because the received echo is attenuated due to propagation loss and possibly distorted due to the interaction of multiple reflections. The optimal detector for the radar problem is the Neyman- Pearson (NP) detector, which maximizes the probability of detection for a given probability of false alarm [3]. When the underlying signal of interest (to be detected) is assumed to be known, the Neyman-Pearson theorem leads to the matched filter solution.

Similarly, we formulate the breast cancer diagnosis as a signal detection problem, where the signal of interest is the tumor and the noise refers to the healthy tissue of the breast. That is we build a CAD system based on this communication engineering technology for signal detection. The similarities between the radar signal detection problem and the tumor detection in breast images are striking and are summarized in Table I.

Let us adopt the standard parlance of hypothesis testing theory, where H_0 denotes the hypothesis that the observed signal is due solely to noise (no tumor), and H_1 the alternative hypothesis that the observed signal, $x[n]$, is due to both the underlying signal of interest (the malignant lesion), $s[n]$, and noise, $w[n]$.

$$\begin{aligned} H_0 &: x[n] = w[n], n = 0, 1, \dots, N-1 \quad (1) \\ H_1 &: x[n] = s[n] + w[n], n = 0, 1, \dots, N-1. \end{aligned}$$

Table I: The Analogy Between The Radar Signal Detection Problem and Malignant Region Detection in the Breast

The Radar Problem	The Breast Cancer Problem
The received noisy signal	The ultrasound
The signal of interest	A malignant lesion in the breast
The ambient noise	The healthy tissue in the breast
The matched filter to the transmitted signal	The matched filter to the lesion

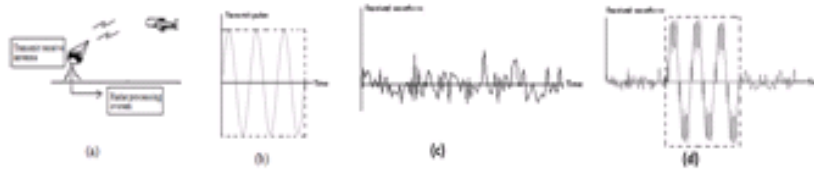


Fig. 1: Depiction of a radar detection problem: (a) the radar system; (b) the transmitted signal; (c) the received signal when no airplane is present; (d) The received signal when an airplane is present.

The signal of interest $s[n]$ is assumed known and $w[n]$ is white Gaussian noise (WGN) with variance σ^2 .

In the decision process, there are two types of concerned error: a Type I error, of deciding in favor of H_1 when H_0 is true, also known as a false alarm, and a Type II error, of deciding in favor of H_0 when H_1 is true, also known as the miss. The probabilities of making such errors are denoted by α and β , respectively. Maximizing the probability of detection in breast cancer (minimizing the Type II error β for a given, predetermined value of α (false alarm probability) is given by the North filter [4]. The NP detector decides H_1 if the likelihood ratio exceeds a threshold or

$$L(x) = \frac{p(x; H_1)}{p(x; H_0)} > \gamma \quad (2)$$

where $x = [x[0]x[1] \dots x[N-1]]^T$ is the observed signal. Given the WGN assumption, it can be easily shown that the NP detector decides H_1 if:

$$\frac{1}{\sigma^2} \sum_{n=0}^{N-1} x[n]s[n] - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} s^2[n] > \ln \gamma. \quad (3)$$

Since $s[n]$ is assumed to be known (and thus not a function of the data), we may incorporate the energy term into the threshold to yield

$$T(x) = \sum_n^{N-1} x[n]s[n] > \sigma^2 \ln \gamma + \frac{1}{2} \sum_{n=0}^{N-1} s^2[n]. \quad (4)$$

Calling the right-hand-side of inequality a new threshold γ' , we decide H_1 if :

$$T(x) = \sum_{n=0}^{N-1} x[n]s[n] > \gamma'. \quad (5)$$

Equation (5) defines the NP detector for a known underlying signal or a known tumor template given by $s[n]$. The detector is shown in Figure 2.

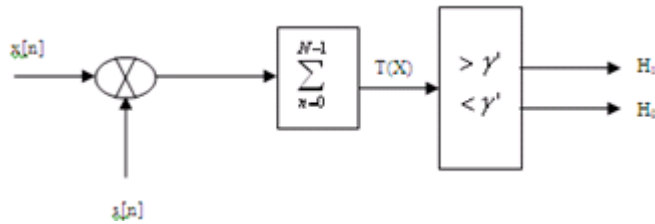


Fig. 2: The NP Detector

The North Filter: The matched filter detector assumes that the noise (healthy breast tissue) is white or uncorrelated. This assumption may not be true given the local homogeneity of the breast tissue. The North filter is a generalization of the matched filter in the case of a colored noise. It can be shown that if $N(f)$ is the Fourier transform of the noise autocovariance function, then the North filter is given by:

$$G(f) = \frac{S^*(f)}{N(f)} \quad (6)$$

where $*$ denotes complex conjugate, and $S(f)$ is the Fourier transform of the known signal of interest $s[n]$.

2.2 The random template model

The matched filter approach assumes that we know *a priori* the signal of interest (the tumor). The matched filter approach therefore relies on a pre-defined template of the tumor signal. In this section, we relax the assumption of known deterministic tumor signal, and consider instead the tumor to be an unknown random signal. This is a valid presumption as the tumor changes its characteristics (shape, size, location, etc) from one patient to another and even from one stage of the disease to another for the same patient. Therefore, unless we have a catalog of all possible tumors in breast images, it is unrealistic to assume that the tumor is perfectly known for all types of breast cancer and all patients.

We adopt a more realistic approach where we assume that the tumor is a random process with a known covariance structure. We examine the optimal detectors that result from the random signal model. To this end we assume $s[n]$ is a zero mean, Gaussian random process with covariance matrix C_s . As before, $w[n]$ is WGN with variance σ^2 . The detection problem is formulated by the same hypothesis testing in Eq. (1), where $s[n]$ is assumed to be normally distributed with covariance matrix C_s and is independent of the noise $w[n]$. In particular, we have the following distribution of the data:

$$x \sim \begin{cases} N(0, \sigma^2 I) & \text{under } H_0 \\ N(0, C_s + \sigma^2 I) & \text{under } H_1 \end{cases} \quad (7)$$

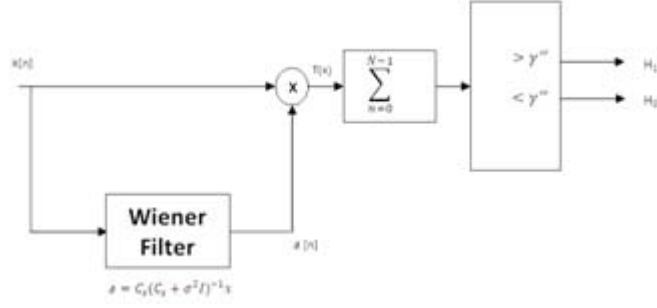


Fig. 3: The estimator-correlator

Computing the logarithm of the likelihood ratio and retaining only the data-dependent terms yields, the NP detector decides H_1 if

$$-\frac{1}{2}x^T \left[(C_s + \sigma^2 I)^{-1} - \frac{1}{\sigma^2} I \right] x > \gamma \quad (8)$$

Using the matrix inversion lemma

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1} \quad (9)$$

We have upon letting $A = \sigma^2 I, B = D = I, C = C_s$

$$(\sigma^2 I + C_s)^{-1} = \frac{1}{\sigma^2} I - \frac{1}{\sigma^4} \left(\frac{1}{\sigma^2} I + C_s^{-1} \right)^{-1} \quad (10)$$

Hence, we decide H_1 if

$$T(x) = x^T s = \sum_{n=0}^{N-1} x[n]s[n] > \gamma'' \quad (11)$$

where $s = C_s(C_s + \sigma^2 I)^{-1}x$. The NP detector correlates the received data with an estimate of the signal $s[n]$, i.e., $s[n]$. It is therefore termed an estimator-correlator (Fig. 3). Note that the test statistic in Eq. (11) is a quadratic form in the data and thus will not be Gaussian random variable. Moreover, it can be shown that $s[n]$ is the minimum mean square error estimator of $s[n]$, also called the Wiener filter [3].

3. Simulations

3.1. The known template model

In ultrasound images, lesions are almost invariably darker than the background, and thus the gray-scale image of the breast is inverted prior to filtering in order to avoid having zero values for tumor areas after correlation.

The matched filter: Figure 4-a shows an ultrasound image, which reflects a malignant type of tumor in the breast (in the upper left corner of the image). In particular, observe that the lesion shows an anisotropic diffusion behavior within the breast tissue. Any template used for the tumor should, therefore, reflect this anisotropic behavior. In Figure 4-b, we plotted the trace (the row) along the center of the lesion. Given the diffusive behavior of most lesions in ultrasound images, we propose to use a trapezoidal malignant lesion template with diffusive boundaries as shown in Fig. 4-c. The correlation curve of the trace with the template is shown in Fig. 4-d.

We process the ultrasound image in a scanline fashion (i.e., row by row). We correlate each trace (row) with the trapezoidal template, which provides an optimal (in the NP sense) automatic system to detect tumors in ultrasound breast images. The images that contain a tumor exhibit correlation curve peaks which are higher on average than healthy images. The threshold γ' was chosen experimentally based on a training set of ultrasound breast images. Figure 4-e shows the average (along all scanlines) correlation curve of a healthy ultrasound breast (in red) and cancerous breasts. Observe that the correlation curve of the healthy breast is below the correlation curves of all shown cancerous images.

The North filter: Unlike the matched filter, the North filter takes into account the correlation structure of the noise, which may not necessarily be white. It is actually intuitive that the healthy tissue in the breast possesses some type of correlation structure. In the North filter implementation, the tumor template is still

assumed to be known *a priori*, modeled by the diffusive template in Fig. 4-c. We found that the noise sequences (scans of the healthy tissues in the breast) can be accurately represented by a second order autoregressive (AR) model. Figure 6 shows a noise sequence and its second order AR model, their corresponding spectrum, the magnitude of the time-domain North filter, and the average (along the image scanlines) North filters of a cancerous and a healthy breast. Observe that the North filter of a cancerous breast shows higher amplitudes than a healthy breast. In practice, the average North filter of each ultrasound image is computed, and compared to the threshold. If there is no peak that exceeds the threshold, the CAD system concludes that the examined image reflects a healthy breast.

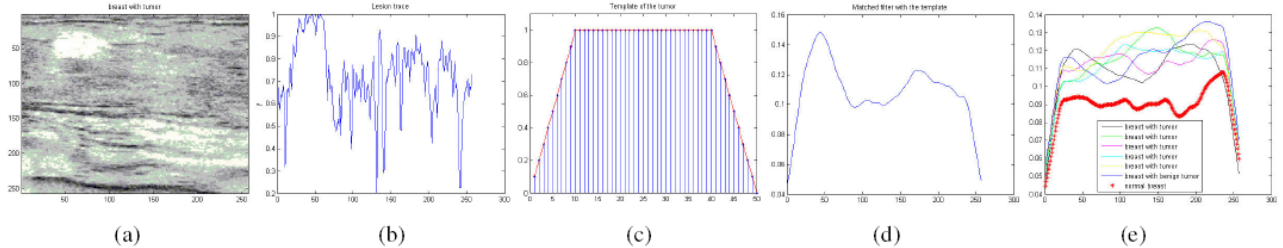


Fig. 4: The matched filter: (a) an ultrasound image of a malignant lesion in breast; (b) trace (row plot) along the center of the lesion; (c) The trapezoidal template of the tumor; (d) The correlation curve of the trace in (b) with the template in (c); (e) The average (along all traces of the image) correlation curve of a healthy breast in red, and cancerous breasts in other colors.

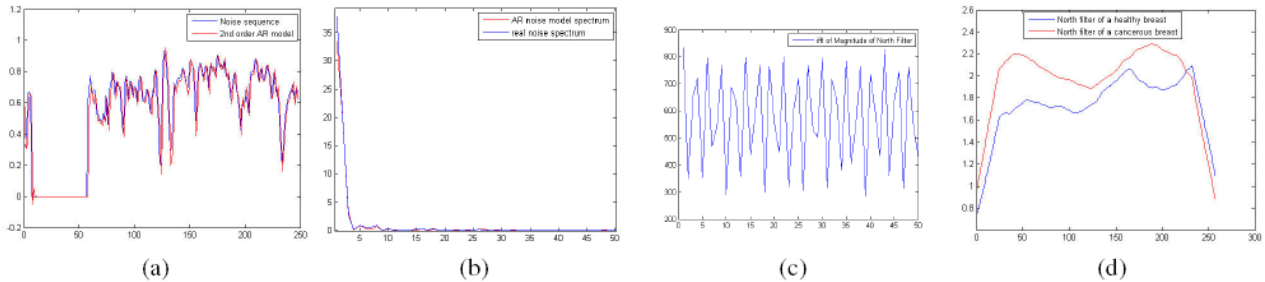


Fig. 5: The North filter: (a) A noise sequence and its second order AR model; (b) The spectrum of a noise sequence and its AR model; (c) The magnitude of the North filter $g[n]$; (d) The North filters of a healthy and cancerous breasts.

3.2. The random template model: *The estimator-correlator*

The covariance matrix of the lesion, C_s , is estimated as the average of the covariance matrices of a training set of cancerous ultrasound breast images. Each trace of the image is then correlated with the estimated Wiener filter of the lesion, and compared to the experimentally chosen threshold to detect the presence or absence of a lesion. The Wiener filter of a cancerous trace and a healthy trace are displayed in Figs. 6-b and 6-c, respectively. Table II shows the performance measures (accuracy, specificity, sensitivity) of the matched filter, North filter and estimator-correlator. As we argued earlier, the random lesion model leads to superior results to the fixed template model because it provides a more realistic model for breast tumors, which vary in size, shape, location, etc., according to the type of the disease, the stage of the cancer, and the profile of the patient.

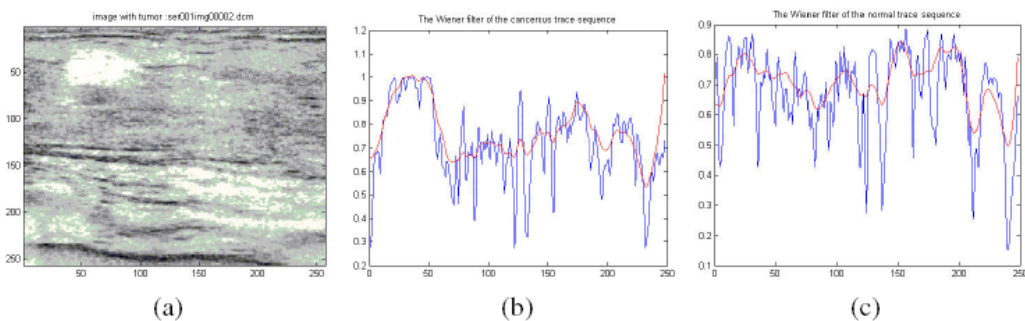


Fig. 6: The estimator-correlator: (a) A cancerous ultrasound image; (b) A trace along the tumor and its Wiener filter in red; (c) A trace along a healthy breast and its Wiener filter in red.

TABLE II: Performance Analysis of the Matched Filter, North Filter and Estimator-Correlator

	The Matched Filter	The North Filter	The Estimator-Correlator
Accuracy	83%	83%	91%
Specifity	67%	83%	83%
Sensitivity	100%	83%	100%

4. Conclusion

In this paper, we introduced a statistical approach to the problem of tumor detection in ultrasound images based on optimal detection theory used in communication engineering. We investigated two cases: a known fixed template of the malignant lesion, and a random model of the lesion. The known template model results in the matched filter detector (for white noise) and the North filter detector (for colored noise). The random template model results in the estimator-correlator or the Wiener filter correlator. We observed that the performance of the detection of the tumor strongly depends on the template used. An effective North filter can be implemented as a CAD system for breast cancer diagnosis provided a library of tumor templates is available to enhance the performance of the system. Breast cancer lesions vary in their morphological characteristics like size, shape, and location. The estimator-correlator detector provides better results than the matched filter and the North filter. Although the estimator-correlator provides a more realistic model of the tumor, we believe that the North filter will be as robust with a catalog of all possible tumors of the breast.

5. References

- [1] J. Zielinski N. Bouaynaya and D. Shonfeld, Two-dimensional ARMA modelling for breast cancer detection and classification, in *IEEE International Conference On Signal Processing And Communications*, July 2010.
- [2] J. Zielinski N. Bouaynaya and D. Shonfeld, Statistical sequential analysis for detection of microcalcifications in digital mammograms, in *IEEE International Conference On Signal Processing And Communications*, July 2010.
- [3] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.
- [4] D. O. North, Analysis of the factors which determine signal/noise discrimination in radar, in *Proceedings of the IRE*, July 1963, vol. 51, pp. 1016–1027.