This is to certify that the

thesis entitled

ADAPTIVE WAVELET-DOMAIN DIGITAL IMAGE WATERMARKING: A DETECTION-THEORETIC APPROACH

presented by

Keith J. Jones

has been accepted towards fulfillment
of the requirements for

MS degree in EE

Major professor

Date 5/6/99

ADAPTIVE WAVELET-DOMAIN DIGITAL IMAGE WATERMARKING: A DETECTION-THEORETIC APPROACH

BY

KEITH J. JONES

A THESIS

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

1999

ABSTRACT

Adaptive Wavelet-Domain Digital Image Watermarking: A Detection-Theoretic Approach

By

Keith J. Jones

The goal of digital watermarking is the ability to embed information into a file so any attempts to remove the information would render the file useless to the attacker. This technological tool is used for a variety of purposes including ownership declaration, authentication, content control, and covert communications. The basic engineering trade-offs involved when designing a digital watermark are with the following three competing factors: information embedding rate, distortion of the original file, and robustness due to intentional or unintentional attempts to remove the watermark information. A general framework which addresses these trade-offs is presented and an existing adaptive wavelet-domain scheme is further developed using hypothesis testing. A new adaptive wavelet domain watermarking method is also proposed which is superior to the existing method, in real world application.

For my beautiful wife, Andrea.

ACKNOWLEDGEMENTS

I would like to acknowledge all the help from Dr. Robert Nowak, from the initial ideas for the foundation of this thesis, to the proof reading of the final draft. Without him, this would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Digital media provides an extraordinary medium to share information such as audio, images, and video. Due to the ability to transmit digital media easily, content providers find it advantageous to embed additional information in their files. This introduction will provide a background about watermarking that will be useful for the analysis of the specific algorithms.

There are many uses for digital watermarks. Some of them include:

1. *Proof of ownership* of a file can allow the owner to distribute their content in a medium, such as the internet, where this file could be copied repeatedly. The watermark, which contains the ownership information, may be the only proof the owner will have to collect any necessary royalties.

2. *File authentication* would allow a person to receive a file and be reassured it was produced and not tampered with in route to the receiving entity.

3. Using a watermark to *control a file's usage* would allow the owner to place restrictions such as the *copy/no copy/copy once* scheme for DVD disks.

4. *Covert communications* would allow for two entities the ability to transmit information to each other secretly. The security in this application is high because third parties would not know to look within a transmitted file for the communication.

For further discussion on watermark usage, see [1].

For many of these applications to be successful, the embedding scheme must be robust. A watermarking scheme is considered to be robust when removing or altering ("attacking") the watermark causes the host file (in which the watermark

is embedded) to be of little or no use. For example, if the host is an image, then removing a "robust" watermark will drastically degrade the visual quality of the image. There are two types of attacks against which a watermarking algorithm must be robust: the unintentional attack and the intentional attack. Unintentional attacks consist of any transformation that may be applied without the intent to harm a watermark and still leave the visual quality of the image acceptable. Some examples of unintentional attacks on images are:

1. Filtering.
2. Cropping.
3. Resizing.
4. Rotation.
5. Digital to analog, and analog to digital conversions.
6. Format conversions.
7. Compression.
8. Non linear transformations
9. Color manipulations.
10. Multiple watermarks.
11. Noise.

Intentional attacks are performed with the specific aim to remove the watermark information. Intentional attacks on images include:

1. Collusion
2. Attack on the detector.

3. Attack on the encoder.

4. False watermark information representation.

5. Multiple watermarking.

6. Removal of sections of the image.

It is important to mention the existence of software packages to test the robustness of watermarking methods. One such package, *StirMark* [2],[3],[4], automates many of the attacks above. For further discussion on attack methods, see [1].

Watermarks can be used in nearly every file type. Some examples of file types that can be watermarked are the following:

1. *Text documents* - ASCII or document processing programs.

2. *Audio files* - WAV and MP3.

3. *Image files* - GIF, JPEG, TIFF, PGM, and RAW.

4. *Video files* - AVI and MPG.

5. *CAD files* - watermark the textures applied to the object.

For the rest of this thesis, the concentration will be on gray-scale images that are 256×256 pixels.

Currently, there are two general types of watermarks used: visible watermarks and invisible watermarks. A visible watermark provides a robust method but may also degrade the visual usefulness of the file. This type of watermark usually consists of a symbol, phrase, or trademark and appears in the image, visible to the naked eye. An invisible watermark would try to alleviate any visual degradation to the file, but its robustness is limited. This type of watermark is invisible to the naked eye and requires a special detection algorithm to decode the watermark. An

invisible watermark usually consists of an ASCII phrase or a binary number. A watermark causing little visible degradation and that is highly robust to a number of "attacks", or intentional attempts to remove it, is very desirable.

From an engineering perspective, the basic trade-offs involve the following competing factors when designing a watermarking algorithm:

1. The amount of information embedded.

2. The degradation to the original image.

3. The robustness of the watermark against attacks.

The visible degradation stems from having to hide the information within the digital image which degrades some of the visual quality. Since the human visual system is not perfect, a small perturbation from the original image to generate a watermarked image may not be detected by the human eye. Therefore there is a range of perturbations the original image, also called the host image, can endure during the watermarking process before the watermarked image is considered inferior to the host in visual quality.

To make a watermark robust, a scheme would typically watermark a large portion of the image and be highly redundant. Watermarking a large amount of visual information could erode the visual quality of a watermarked image. Embedding the watermark information redundantly will limit the amount of information embedded. Hence, the trade-offs involved are readily apparent.

The work in this thesis improves upon an existing watermark embedding method by formal hypothesis testing. In addition, a new digital watermark encoding method is proposed that is practically superior to the existing method. The new method increases the robustness while keeping the visual degradation of the encoding phase and the amount of information encoded, a constant. The analysis

here also sheds light on several important features of the adaptive watermarking methods.

CHAPTER 2

OVERVIEW OF EXISTING METHODS

2.1 Spread Spectrum Approaches

Many watermarks use an approach defined as *spread spectrum* embedding. The approach adds pseudo-noise information to the host image to construct the watermarked image. *Least significant bit (LSB)* method is one example of this approach. The LSB method adds the information to the least significant bit of each pixel value to represent the watermark. This approach is easily removed because the attacker can subtract their pseudo-noise information signal from the watermarked image and create a new watermark in the image. This process is known as the *IBM attack* [5]. Since this method is easily attacked and removed, there is a need to develop a more robust system. The following section proposes the next logical step in such an approach.

2.2 Quantization Approaches

In [6], a generalized model of the quantization technique that addresses the tradeoffs among visual degradation, robustness, and amount of data embedded is proposed. In this model, the host is represented by a vector $x \in \Re^N$, where $\Re$ is the real line. The host signal can be either the image itself or a transformation, such as the Discrete Wavelet Transform (DWT), of the image. The data will be embedded into x at a rate of R bits per host signal sample. Denote by m, the integer chosen from a set $\{0, 1, 2, ..., 2^{N \cdot R} - 1\}$, as the information to be embedded. $s(x, m)$ is the quantized (watermarked) data indexed by m, defined as a *quantization index modulation* function. Also define a noise vector, n, which can be stochastic or determinant, signal dependent or independent, that represents an

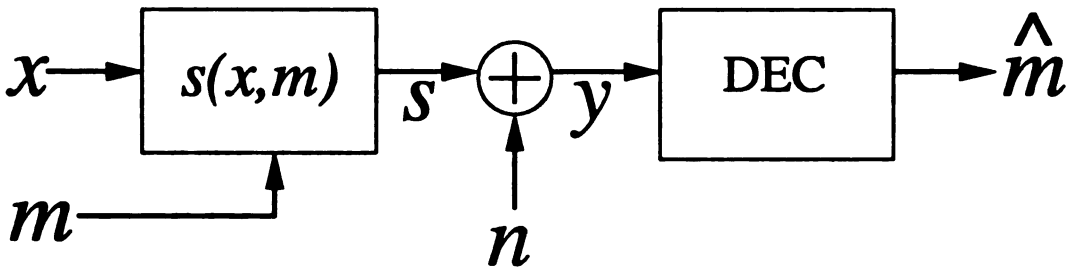


Figure 2.1: General watermark embedding model.

attack on the watermark. In the case of no attack, the noise vector n, is ideally zero. The decoder will calculate the estimate $\hat{m}$ of m from the attacked image y. This completely general model can be observed in Fig. 2.1.

The authors of [6] further the approach by a *coded dither modulation* scheme in which $q[\cdot]$ represents a given quantizer. This quantizer is used at shifted increments to represent a *dithered quantizer*, parameterized by a dither vector $d(m)$ for each possible value of m, so that:

$$s(x;m) = q[x + d(m)] - d(m) \tag{2.1}$$

The goal of this approach is to ensure that two dithered quantizers are the maximum possible distance from each other. As a simple example, $q[\cdot]$ could be defined as a uniform, scalar quantizer with step size Δ. See [6] for further explanation of this example. Although this general model is useful, a watermarking method that adapts to the host signal could provide more robustness.

2.3 Adaptive Quantization Approaches

Although the approach in the previous section provides a more robust scheme to watermark images, it is possible to exploit characteristics of the human visual system (HVS) to strengthen the robustness of the quantization method. The HVS

is unable to detect noise in areas of similar frequency patterns. This phenomenon is known as *frequency masking* [1]. The HVS is also known to be unable to detect the visibility of other features that are spatially close to them, and this is known as *spatial masking* [1]. Together, a new type of masking, known as *edge masking*, allows for a stronger signal to be hidden in the edge features at different frequency levels and spatial locations within the image. Where edges are absent, the watermark information will be less obtrusive. A watermarking scheme which exploits edge masking is characterized as an *adaptive method*.

In contrast to the dithered quantizer in the previous section, adaptive quantizers are signal dependent. Near edges of the host image, $s(x, m)$ will watermark the image harshly without noticeably degrading the visual quality of the image. Such an adaptive quantizer will increase the robustness of the watermarking algorithm. The methods proposed in the following chapters are all adaptive.

CHAPTER 3

FORMULATION OF ADAPTIVE WAVELET-BASED WATERMARKING METHODS

3.1 The Wavelet Transform

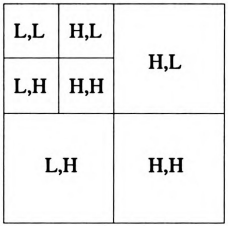
The discrete wavelet transform (DWT) provides localization in frequency and space. Because of this fact, the wavelet transform provides a much more suitable domain to adaptively watermark information. Furthermore, the wavelet transform is known to be sparse, and the edges of the image tend to be evident in the highest absolute values of the wavelet coefficients [7]. This edge phenomenon can be seen as the white pixels in Fig. 3.1(c). This feature is effective for exploiting the edge masking property when embedding a watermark. For further explanation of the DWT, see [7].

The two-dimensional DWT (2DWT) can be efficiently computed using a filter bank structure. The sub-image in the lower right corner of Fig. 3.1(b) is obtained by applying a high-pass filter (H) to the rows and columns of the original image and decimating in each direction. The sub-images H,H, (diagonal details) L,H, (horizontal details) and H,L (vertical details) are composed of the wavelet coefficients at the first scale. The sub-image in the upper-left corner represents the scaling coefficients at the next scale and is obtained by reiterating the filter bank on the previous scale's L,L sub-image. Fig. 3.1 illustrates the process of performing the two-dimensional wavelet transform on a test image.

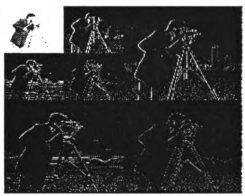
3.2 The Overview of an Existing Method

One promising method using wavelet-domain adaptive watermark encoding was proposed in [8]. The embedding rate, defined in Section 2.2, is $R = \frac{1}{3}$. A value

(a) The Original Image.



(b) The 2-D wavelet transform diagram.



(c) The 2-D wavelet transform example.

Figure 3.1: The 2-D wavelet transform example.

of Q is chosen to represent a trade-off between robustness and visual degradation of this algorithm. Q is chosen on a trial-and-error basis. It is desirable to choose the lowest integer Q value that does not produce visual degradation to the host image. Typically, values of $6 \leq Q \leq 10$ do not produce noticeable distortions. The framework for the system proposed can be generalized into the following algorithm for embedding a watermark:

1. Compute the DWT of the host image. Denote the wavelet coefficient image as w.

2. For each scale, l, and each position (i, j) within l, perform the following steps:

   (a) There are three distinct sets of wavelet coefficients, corresponding to the three orientations; vertical, horizontal, and diagonal. Take the triple of coefficients at position (i, j) and scale l, and order them so that:

   $$w_{l,d_1}(i,j) < w_{l,d_2}(i,j) < w_{l,d_3}(i,j) \tag{3.1}$$

   where $d_k \in$ {horizontal,vertical,diagonal}, each distinct. If

   $$w_{l,d_k}(i,j) = w_{l,d_m}(i,j)$$

   when $k \neq m$, then perturb the value of $w_{l,d_k}(i,j)$ slightly so that the algorithm may continue.

   (b) Given Q, calculate the quantization step size with the following equa-

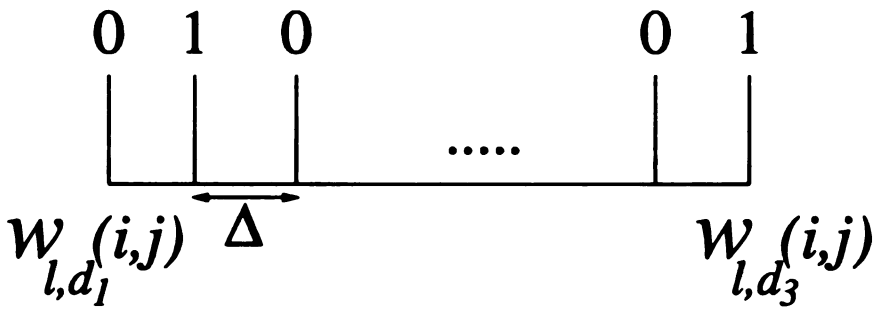


Figure 3.2: The quantization model.

tion:

$$\Delta = \frac{w_{l,d_3}(i,j) - w_{l,d_1}(i,j)}{2Q-1} \tag{3.2}$$

(c) To embed a watermark bit of value one, quantize $w_{l,d_2}(i,j)$ to the nearest quantization level with 1 written over it in Fig. 3.2. To embed a watermark bit of value zero, quantize $w_{l,d_2}(i,j)$ to the nearest quantization level with a 0 written over it in Fig. 3.2.

The decoding process is also formulated into an algorithm as follows:

1. Compute the DWT of the host image. Denote the wavelet coefficient image as w.

2. For each scale, l, and each position (i,j) within l, perform the following steps:

   (a) There are three distinct sets of wavelet coefficients, corresponding to the three orientations; vertical, horizontal, and diagonal. Take the triple of coefficients at position (i,j) and scale l, and order them so that:

$$w_{l,d_1}(i,j) < w_{l,d_2}(i,j) < w_{l,d_3}(i,j) \tag{3.3}$$

where $d_k \in$ {horizontal,vertical,diagonal}, each distinct. If

$$w_{l,d_k}(i,j) = w_{l,d_m}(i,j)$$

when $k \neq m$, then perturb the value of $w_{l,d_k}(i,j)$ slightly so that the algorithm may continue.

(b) Given Q, calculate the quantization step size with the following equation:

$$\Delta = \frac{w_{l,d_3}(i,j) - w_{l,d_1}(i,j)}{2Q-1} \tag{3.4}$$

(c) Find the closest quantization level in Fig. 3.2 to the w_{l,d_2} value. If there is a 0 over the line, a zero is decoded. If there is a 1 over the line, a one is decoded.

Decoding the wavelet coefficient to the nearest quantization level in Fig. 3.2 is equivalent to finding the maximum value of the following expression, with respect to n, where the integer $n \in [0, 2Q-1]$:

$$\Gamma(W,n) = -\left(w_{l,d_2}(i,j) - w_{l,d_1}(i,j) - n\frac{w_{l,d_3}(i,j) - w_{l,d_1}(i,j)}{2Q-1}\right)^2 \tag{3.5}$$

If the value of n that maximizes Eq. (3.5) is even, a zero is decoded. If the value of n that maximizes Eq. (3.5) is odd, a one is decoded. Eq. (3.5) is approximately the log likelihood expression for the observed data after an additive attack, modeled as an additive Gaussian White Noise (GWN). This decoder provides invariance for affine transformations to a given wavelet triple representing the three directions for the coordinates (i,j) in level l.

The value of Q used is also very important. $2Q-1$ denotes the number of quantization levels in Fig. 3.2. For higher values of Q, the range is quantized with higher resolution, while lower values of Q quantize coarsely. The value of Q represents part of the engineering trade-off for the proposed system because a lower value of Q will produce a watermark more robust and a higher level of visual degradation. A higher value of Q will produce a system less robust, but the watermarked image will bare a closer resemblance to the host image.

This watermarking scheme is adaptive because at areas with edge features, the wavelet coefficients will be large. When there are large distances between w_{l,d_1} and w_{l,d_3}, the watermark perturbs the host image more aggressively compared to the case when w_{l,d_1} and w_{l,d_3} are near each other. When w_{l,d_1} and w_{l,d_3} are near each other, the edge feature is missing and the perturbation to the host image is small.

3.2.1 A Decision - Theoretic Approach

The proposed method in Section 3.2 may not produce the n which minimizes the probability of an error. Given a model of the attack process, the goal of this chapter is to minimize the probability of error for the existing method by finding the most likely encoded bit using the hypothesis test:

$$H_0 \ : \ n \text{ is even, 0 encoded,}$$

$$H_1 \ : \ n \text{ is odd, 1 encoded.}$$

First, the preliminaries must be established. Define a vector Y as the observations of the wavelet coefficient triple in a given image suspected of containing a watermark:

$$Y = \begin{bmatrix} y_{l,d_1}(i,j) \\ y_{l,d_2}(i,j) \\ y_{l,d_3}(i,j) \end{bmatrix} \tag{3.6}$$

Define the vector W as the original watermarked wavelet coefficients before the attack:

$$W = \begin{bmatrix} w_{l,d_1}(i,j) \\ w_{l,d_2}(i,j) \\ w_{l,d_3}(i,j) \end{bmatrix} \tag{3.7}$$

This suspected image is assumed to have an attack, modeled as an additive Gaussian noise, defined as the vector ϵ, with the covariance matrix Σ_Y, given below, and a zero mean vector.

$$\Sigma_Y = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix} \cdot \sigma^2 \tag{3.8}$$

Notice this is only one such choice for the covariance matrix, other more general covariance structures will be proposed in Chapter 5. This particular covariance matrix is chosen because it adds the next logical step of complexity beyond an independent error model (diagonal covariances), allowing for possible correlation, ρ, between the wavelet coefficient errors. This level of generality, as opposed to an independent error model, is important since many attacks, e.g., JPEG, may introduce correlated errors. The attack is modeled by:

$$\begin{bmatrix} y_{l,d_1}(i,j) \\ y_{l,d_2}(i,j) \\ y_{l,d_3}(i,j) \end{bmatrix} = \begin{bmatrix} w_{l,d_1}(i,j) + \epsilon_1 \\ w_{l,d_2}(i,j) + \epsilon_2 \\ w_{l,d_3}(i,j) + \epsilon_3 \end{bmatrix} \tag{3.9}$$

Furthermore, the log likelihood function is:

$$\tau_Y = -(Y - E[Y])' \Sigma_Y^{-1} (Y - E[Y]) \tag{3.10}$$

A transformation from the three variable system in Eq. (3.9) can be made to the two variable system in Eq. (3.11) without loss, since the wavelet coefficient differences alone contain the necessary decoding information.

$$B = \begin{bmatrix} b_{l,1}(i,j) \\ b_{l,2}(i,j) \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{l,d_1}(i,j) \\ y_{l,d_2}(i,j) \\ y_{l,d_3}(i,j) \end{bmatrix} \tag{3.11}$$

Many statistical values can be evaluated:

$$t = w_{l,d_3}(i,j) - w_{l,d_1}(i,j) \tag{3.12}$$

$$w_{l,d_2} - w_{l,d_1} = \frac{nt}{2Q-1} \tag{3.13}$$

$$E[b_{l,1}(i,j)] = t \tag{3.14}$$

$$E[b_{l,2}(i,j)] = \frac{nt}{2Q-1} \tag{3.15}$$

The derivation for the new transformed covariance matrix is given in Eq. (3.16)

through (3.27):

$$\begin{aligned}
\text{var}(b_1) &= E[(\epsilon_3 - \epsilon_1)^2] && (3.16)\\
&= E[\epsilon_3^2 - 2\epsilon_3\epsilon_1 + \epsilon_1^2] && (3.17)\\
&= (1 - 2\rho + 1)\sigma^2 && (3.18)\\
&= 2(1-\rho)\sigma^2 && (3.19)\\
\text{var}(b_2) &= E[(\epsilon_2 - \epsilon_1)^2] && (3.20)\\
&= E[\epsilon_2^2 - 2\epsilon_2\epsilon_1 + \epsilon_1^2] && (3.21)\\
&= (1 - 2\rho + 1)\sigma^2 && (3.22)\\
&= 2(1-\rho)\sigma^2 && (3.23)\\
\text{cov}(b_1, b_2) &= E[(\epsilon_3 - \epsilon_1)(\epsilon_2 - \epsilon_1)] && (3.24)\\
&= E[\epsilon_3\epsilon_2 - \epsilon_1\epsilon_2 - \epsilon_3\epsilon_1 + \epsilon_1^2] && (3.25)\\
&= (\rho - \rho - \rho + 1)\sigma^2 && (3.26)\\
&= (1-\rho)\sigma^2 && (3.27)
\end{aligned}$$

These calculations lead to the form of the covariance matrix below:

$$\Sigma_B = \text{Cov}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}(1-\rho)\sigma^2 \qquad (3.28)$$

The new covariance matrix has a structure which is invariant to the value of ρ. Therefore, the overall probability of error is a function of ρ, but the detector structure (decision rule) will be indifferent. The performance analysis with different values of ρ will be discussed in section 3.3.1. The position notation (i, j) and the scale index l will be omitted where unnecessary to keep the equations simple

throughout the rest of this analysis. The new log likelihood equation is:

$$\tau_B = -(B - E[B])'\Sigma_B^{-1}(B - E[B]) \tag{3.29}$$

The hypothesis test involves the choice of whether z, the bit initially encoded, is zero or one. The minimum probability of error decision [9] is determined by the *likelihood ratio test (LRT)* defined as:

$$\Lambda(B) = \frac{p(B|n,t,z=1)}{p(B|n,t,z=0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1 \tag{3.30}$$

Several nuisance parameters prevent a direct calculation of this test. The nuisance parameters are:

- $n \in \{0,2,4,6,8,...,2Q-2\}|z=0$
- $n \in \{1,3,5,7,8,...,2Q-1\}|z=1$
- $t \in \Re^+$

The nuisance parameter n, providing the exact position of the bit encoded, and t, the quantization distance, are not known. One approach to solve this problem is the *generalized likelihood ratio test (GLRT)* where the nuisance parameters are replaced by the maximum likelihood estimates (MLE) of their values. The GLRT is defined as follows:

$$\tilde{\Lambda}(B) = \frac{\max\limits_{n \text{ odd}, t \in \Re^+} p(B|n,t,z=1)}{\max\limits_{n \text{ even}, t \in \Re^+} p(B|n,t,z=0)} \underset{H_0}{\overset{H_1}{\gtrless}} 1 \tag{3.31}$$

Computing the MLE for t and substituting it into the likelihood ratio test, Eq. (3.30), nearly completes the GLRT and provides Eq. (3.32). The completion of the GLRT involves searching for the n that maximizes Eq. (3.32) within the range of $[0, 2Q-1]$. Finally, if the maximizing n is even, then it is assumed a zero was encoded; if it is odd then a one was encoded. Hence, this completes the hypothesis test.

$$\max_{t} \log p(B|n,t) = \frac{(b_2 + b_1 n - 2b_2 Q)^2}{2(\rho-1)(n+n^2+(1-2Q)^2-2nQ)} \tag{3.32}$$

3.3 The New Adaptive Watermarking Method

Here a new method is proposed which is a variant/extension of the previous method. The new method defines an information bit with two integers (m,n), representing a zero if m and n are even, and a one if m and n are odd. Notice the embedding rate, defined in Section 2.2, is also $R = \frac{1}{3}$ for this method. A Q value will be used which serves the same purpose as in the previous method. The integer values of m and n are defined by:

$$(m,n) \in [-(2Q-1), 2Q-1] \times [-(2Q-1), 2Q-1] \tag{3.33}$$

The encoding process can be summarized in the following algorithm:

1. Compute the DWT of the host image. Denote the wavelet coefficient image as w.

2. For each scale, l, and each position (i,j) within l, perform the following steps:

(a) There are three orientations in the DWT image. Take the same coordinates, (i, j), in l and order the wavelet coefficients so that:

$$|w_{l,d_1}(i,j)| < |w_{l,d_2}(i,j)| < |w_{l,d_3}(i,j)| \tag{3.34}$$

where $d_k \in$ {horizontal,vertical,diagonal}, each distinct. If

$$|w_{l,d_k}(i,j)| = |w_{l,d_m}(i,j)|$$

when $k \neq m$, then it is possible to perturb the value of $w_{l,d_k}(i,j)$ slightly so that the algorithm may continue.

(b) The embedding process makes the following changes to the wavelet coefficient triple:

$$w'_{l,d_1}(i,j) = \frac{m\, w_{l,d_3}(i,j)}{2Q-1} \tag{3.35}$$

$$w'_{l,d_2}(i,j) = \frac{n\, w_{l,d_3}(i,j)}{2Q-1} \tag{3.36}$$

To embed a zero, (m, n) are both chosen to be even such that the distance between the host, w, and new, w', wavelet coefficient values are minimal. If a one is to be embedded, (m, n) are both chosen to be odd such that the new, y, values are minimal distance from the host, w', corresponding wavelet coefficients.

The new method also has a similar decoding sequence to the previous method. It is summarized as follows:

1. Compute the DWT of the image in question. Denote the wavelet coefficient image as w.

2. For each scale, l, and each position (i,j) within l, perform the following steps:

   (a) There are three orientations in the DWT image. Take the same coordinates, (i,j), in l and order the wavelet coefficients so that:

$$|w_{l,d_1}(i,j)| < |w_{l,d_2}(i,j)| < |w_{l,d_3}(i,j)| \tag{3.37}$$

   where $d_k \in$ {horizontal,vertical,diagonal}, each distinct. If

$$|w_{l,d_k}(i,j)| = |w_{l,d_m}(i,j)|$$

   when $k \neq m$, then it is possible to perturb the value of $w_{l,d_k}(i,j)$ slightly so that the algorithm may continue.

   (b) Find the values of (m,n) which are both even or both odd that maximize the following equation:

$$\begin{aligned}\Gamma(W,m,n) = \; & -\left(w_{l,d_1}(i,j) - \frac{m\, w_{l,d_3}(i,j)}{2Q-1}\right)^2 - \\ & \left(w_{l,d_2}(i,j) - \frac{n\, w_{l,d_3}(i,j)}{2Q-1}\right)^2\end{aligned} \tag{3.38}$$

   This is equivalent to finding the m and n values which are both even or odd and minimize the distance when compared to the $w_{l,d_3}(i,j)$ coefficient. Notice this is also in the form of a log likelihood expression.

We can derive a similar decision rule by following the more formal hypothesis test as before. The hypotheses are defined as:

$$H_0 \ : \ n, m \text{ are even, 0 encoded,}$$

$$H_1 \ : \ n, m \text{ are odd, 1 encoded.}$$

To set up the decision rule, one must define more notation. Define a vector Y as the observations of a suspect image according to Eq. (3.6) and (3.39).

$$|y_{l,d_1}(i,j)| < |y_{l,d_2}(i,j)| < |y_{l,d_3}(i,j)| \tag{3.39}$$

Also define a vector W similar to Eq. (3.7) and denote this vector as the true value of the wavelet coefficient triples before an attack. Again, an attack is modeled using Eq. (3.9). The statistical properties of ϵ are the same as that assumed in the previous method using the covariance matrix in Eq. (3.8) and a zero mean vector. Again, this choice of covariance matrix is only one of many, Chapter 5 will recommend more for future research. The expected values of the observation Y are given below.

$$E[y_{l,1}(i,j)] = \frac{m\, w_{l,d_3}(i,j)}{2Q-1} \tag{3.40}$$

$$E[y_{l,2}(i,j)] = \frac{n\, w_{l,d_3}(i,j)}{2Q-1} \tag{3.41}$$

$$E[y_{l,3}(i,j)] = w_{l,d_3}(i,j) \tag{3.42}$$

Using a similar argument as the previous formal decision rule, in this case the GLRT is given by:

$$\tilde{\Lambda}(W) = \frac{\max\limits_{\text{m,n odd, } w_{l,d_3} \in \Re} p(W|m,n,w_{l,d_3},z=1)}{\max\limits_{\text{m,n even, } w_{l,d_3} \in \Re} p(W|m,n,w_{l,d_3},z=0)} \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} 1 \tag{3.43}$$

By finding the maximum likelihood estimate for $w_{l,d_3}(i,j)$ and substituting it into Eq. (3.10), the result nearly completes the GLRT and gives Eq. (3.55).

$$y_1 = y_{l,d_1}(i,j) \tag{3.44}$$

$$y_2 = y_{l,d_2}(i,j) \tag{3.45}$$

$$y_3 = y_{l,d_3}(i,j) \tag{3.46}$$

$$a = y_3^2(m^2+n^2-2mn\rho) \tag{3.47}$$

$$b = 2y_2y_3(n-m\rho-m^2\rho+mn\rho-2nQ+2m\rho Q) \tag{3.48}$$

$$c = y_2^2(1+m^2+2m\rho-4Q-4m\rho Q+4Q^2) \tag{3.49}$$

$$d = y_1^2(n^2+(1-2Q)^2+n(2\rho-4\rho Q)) \tag{3.50}$$

$$e = -2y_1(\rho(1+n-2Q)(y_2+ny_3-2y_2Q)) \tag{3.51}$$

$$f = -2y_1(m(ny_2-y_3+y_2\rho-ny_3\rho+2y_3Q-2y_2\rho Q)) \tag{3.52}$$

$$g = (-1+\rho)(m^2(1+\rho)+n^2(1+\rho)+(1+\rho)(1-2Q)^2) \tag{3.53}$$

$$h = (-1+\rho)(-2m\rho(-1+n+2Q)+n(2\rho-4\rho Q)) \tag{3.54}$$

$$\max_{w_{l,d_3}} \log p(W|m,n,w_{l,d_3}) = \frac{a+b+c+d+e+f}{g+h} \tag{3.55}$$

In a similar manner as before, we search over integer values $(m,n) \in [-(2Q-1),2Q-1] \times [-(2Q-1),2Q-1)]$, both even or both odd, and find the pair that

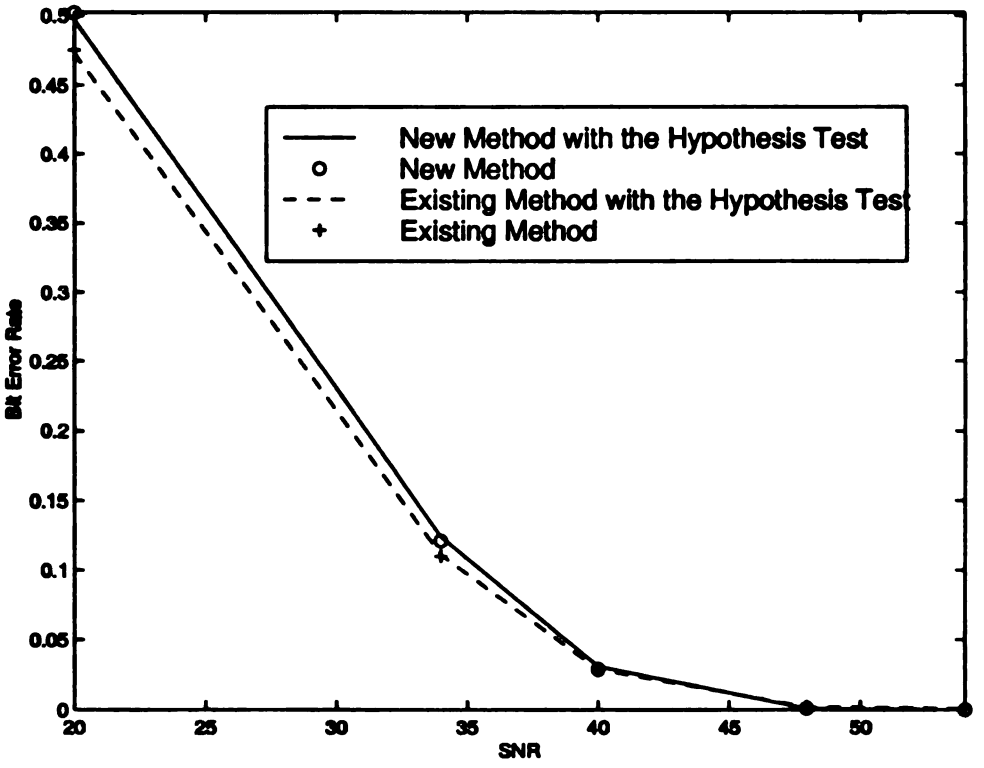


Figure 3.3: The bit-error rate for $\rho = 0.4$.

produces the maximum value of Eq. (3.55). If the maximizing (m, n) are both even, a zero is assumed to be encoded. If they are odd, a one is assumed to be encoded. Hence, this is the completion of the hypothesis test.

An important note is the value for ρ in the covariance matrix can be different for each level of wavelet resolution, l, being decoded. This provides for greater flexibility and better detection.

3.3.1 Comparison/Analysis

The performance simulation is structured such that the host wavelet triples are independently Gaussian distributed. Each watermarking method will be applied to these triples, separately, and attacked with the addition of Gaussian noise. Each method will use the same host wavelet triple values to encode the watermark, the same noises, and the same encoding bit values.

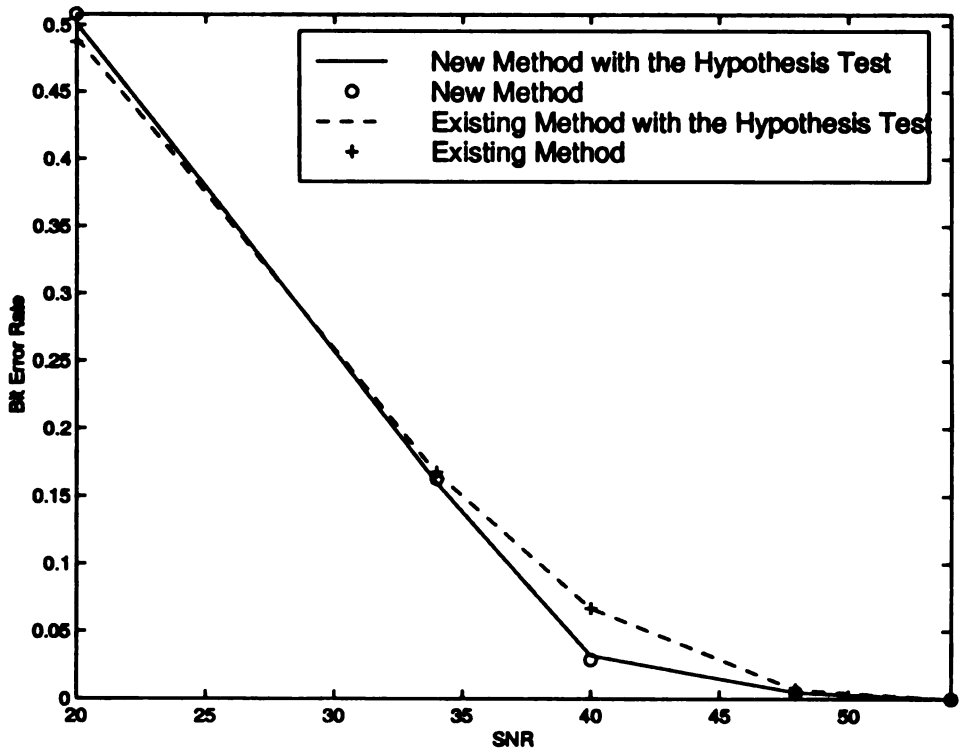


Figure 3.4: The bit-error rate for $\rho = 0$

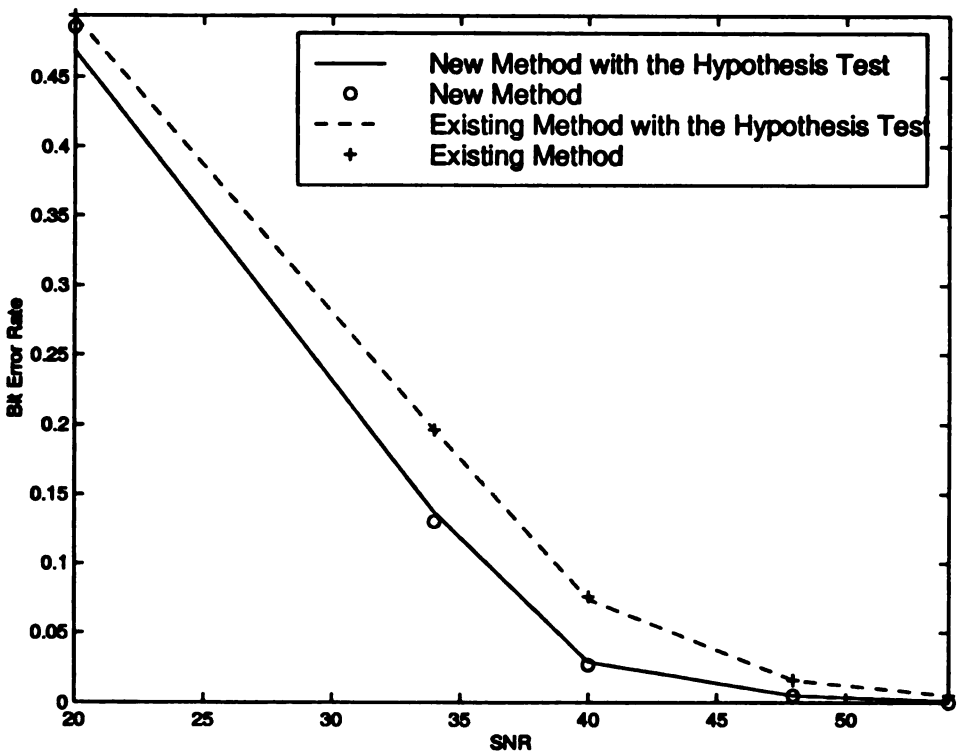


Figure 3.5: The bit-error rate for $\rho = -0.4$

The calculated results from the simulation for the bit-error rate represents a situation when a watermarked image may be attacked by the addition of Gaussian noise. This is an ideal scenario which fits the hypothesis test model perfectly. Because the data and the noise are generated in this manner, a value of power can be associated and will provide a measure of signal-to-noise ratio (SNR) for each iteration. The purpose of this type of simulation is to plot several performance curves for different values of the SNR.

The result of probability of bit-error rate versus signal-to-noise ratio for each of the four decoders are depicted in Fig. 3.3, 3.4, and 3.5. The simulation used a Q value of 10, performed 1000 repetitions for each SNR value, and the stated ρ value. The Q value was chosen as 10 because this is the lowest value observed that does not produce visual degradation.

Fig. 3.3 shows that the proposed method under-performing the existing method. Fig. 3.4 and 3.5 shows the proposed method has superior performance to the existing method at higher SNR values. Therefore, one conclusion can be drawn concerning the value of ρ: if the noise added is uncorrelated or slightly negatively correlated, then the new method is superior.

Another conclusion is that the hypothesis test, Eq. (3.32), for the existing method does not perform much better, or worse, than the *ad hoc* scheme, Eq. (3.5), when different values of ρ are used to generate the correlated noise. This can be explained with the following argument. If the range between $w_{l,d_1}(i,j)$ and $w_{l,d_3}(i,j)$ is quantized with a large Q, the most likely values of $y_{l,d_1}(i,j)$ and $y_{l,d_3}(i,j)$ are the values $w_{l,d_1}(i,j)$ and $w_{l,d_3}(i,j)$, respectively. That is, if a very fine quantization step is used, then the maximization is essentially unconstrained and produces the same results as the *ad hoc* method. Therefore, this argument suggests Eq. (3.5) is actually the same maximization as Eq. (3.32) and the curves are as expected.

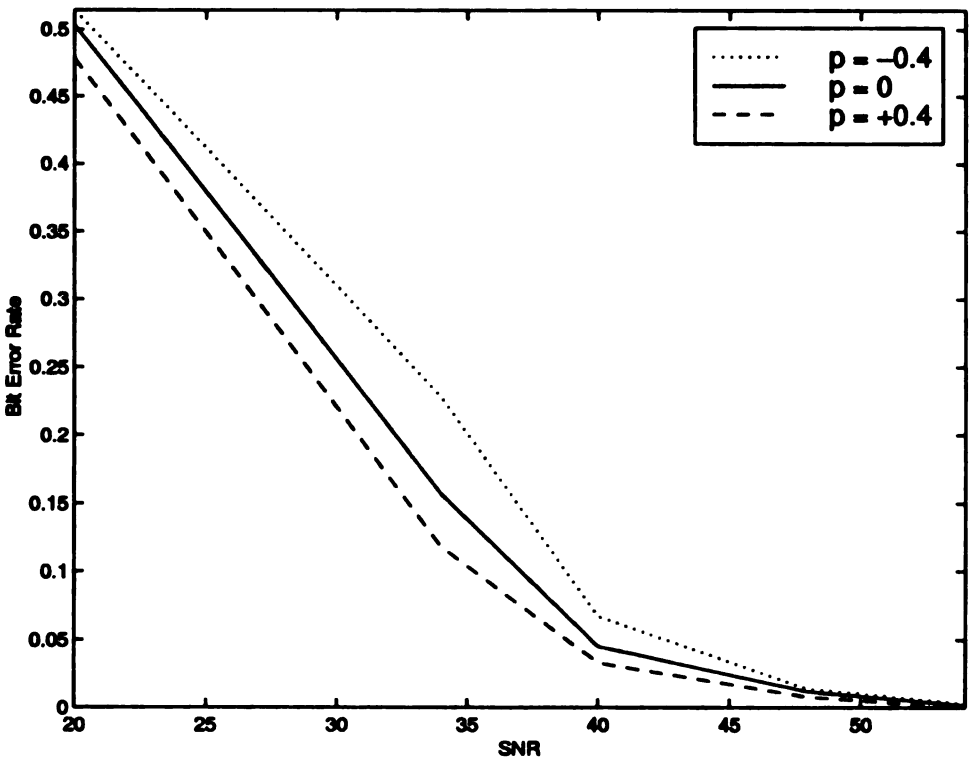


Figure 3.6: The bit-error rate for varying ρ values using the existing method with the hypothesis test.

In contrast, the covariance structure presented in Eq. (3.28) makes the decision rule for the existing method with the hypothesis test invariant to the choice of ρ, but the performance is not. Fig. 3.6 compares the probability of detection for different choices of ρ for the existing method. Notice the detector performance is better with positive values of ρ as opposed to negative values. As the value of ρ tends toward one, the covariance matrix becomes zero. Therefore, the overall covariability becomes zero and the performance of the detector is maximum.

Conversely, the new method is dependent on ρ when using the hypothesis test. When ρ is chosen as a good estimate of the added correlated noise, the proposed detector based on a hypothesis test performs slightly better than the *ad hoc* version. The increase in performance using the hypothesis test for the new method is very slight, however.

CHAPTER 4

APPLICATION

It is desirable to test each watermark embedding method on a variety of real images. Fig. 4.1 represents a typical assortment of test images because there are solid objects (Fig. 4.1(a)), granular (Fig. 4.1(j)), and combinations of straight edges (Fig. 4.1(c) and 4.1(d)). The other images represent different mixtures of each. Furthermore, this subset is extensively used for testing and evaluations purposes in the research community.

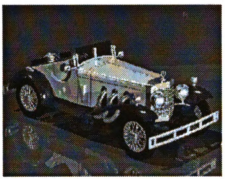
Fig. 4.2 displays an example image selected from the subset watermarked with the existing method and the new method. An example of two attacks can be viewed in Fig. 4.3 and 4.4 for both watermarking schemes. Each figure represents an image with nearly zero visual degradation using JPEG compression of 60% and low pass filtering using a third-order two-dimensional Butterworth filter [10] with cutoff frequency $W_n = 0.9$. Various bit-error rates can be viewed in Tables 4.1 through 4.5. The value of ρ chosen for the tabular results varied, but was usually slightly negative and constant across each wavelet resolution level, l.

Tables 4.1 through 4.5 provide numerous values for the bit-error rates using different strengths of the two attacks. The conclusion drawn from the tables is obvious: the newly proposed method performs superior compared to the existing method. In some instances, the new method performs nearly 30% better than the existing method.

Due to the actual correlation value, ρ, due to these attacks tended to be slightly negative, Fig. 3.3 through 3.5 strengthens the conclusion that the new method performs better than the existing method. This correlation is the key factor in determining the difference in the performance of the two methods.

The hypothesis test for the previous (existing) method performed the same

as the *ad hoc* detector in Eq. (3.5), therefore the argument regarding the rough equivalence of the two decoders given in the last chapter is strengthened. The hypothesis test for the new method seems to do better or worse depending on the value of ρ chosen for the simulation. Since the purpose of this thesis is not to find the optimum value of ρ, it is important to note that an increase in performance could be achieved with enough trial and error experiments to find the best estimate of ρ for each instance of attack and level of wavelet resolution, l.



(a) Benz

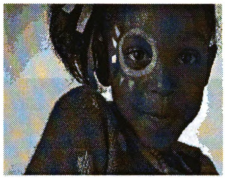
(b) Bridge

(c) Building

(d) Camera

(e) Fruit

Figure 4.1: The host images.

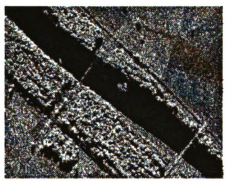


(f) Girl

(g) Glasses

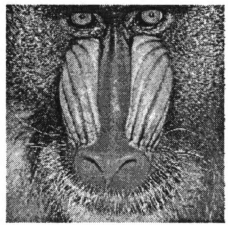
(h) Lenna

(i) Mandrill

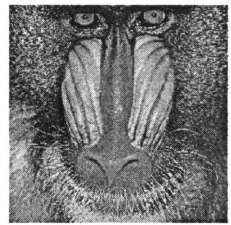


(j) River

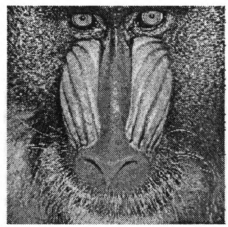
Figure 4.1: The host images (con't).



(a) Original Image

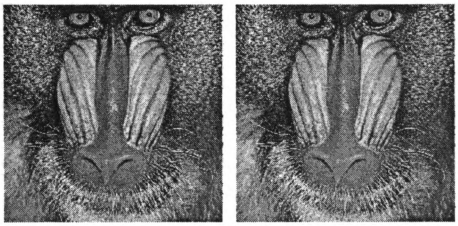


(b) Existing Method



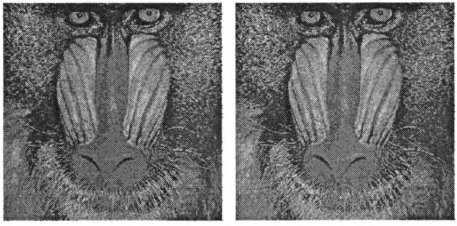
(c) New Method

Figure 4.2: An example of watermarking methods.



(a) Fig. 4.2(b) Attacked (b) Fig. 4.2(c) Attacked

Figure 4.3: An example of JPEG attack, 60% compression.



(a) Fig. 4.2(b) Attacked (b) Fig. 4.2(c) Attacked

Figure 4.4: An example of low-pass filter attack, $W_n = 0.9$.

Image	P_e Method 1 Eq. (3.5)	P_e Method 1 Eq. (3.32)	P_e Method 2 Eq. (3.38)	P_e Method 2 Eq. (3.55)
benz	0.2360	0.2360	0.2015	0.1993
bridge	0.1162	0.1132	0.0868	0.0838
building	0.1971	0.1985	0.1787	0.1772
camera	0.2221	0.2213	0.2081	0.2103
fruit	0.1647	0.1662	0.1412	0.1426
girl	0.2235	0.2235	0.1941	0.2029
glasses	0.1191	0.1184	0.1007	0.1022
lenna	0.1522	0.1529	0.1176	0.1213
mandrill	0.1044	0.1044	0.0669	0.0691
river	0.0596	0.0559	0.0353	0.0368

Table 4.1: The bit-error rates using JPEG compression of 90%.

Image	P_e Method 1 Eq. (3.5)	P_e Method 1 Eq. (3.32)	P_e Method 2 Eq. (3.38)	P_e Method 2 Eq. (3.55)
benz	0.3243	0.3250	0.3096	0.3140
bridge	0.2757	0.2757	0.2507	0.2449
building	0.2875	0.2868	0.2647	0.2632
camera	0.3360	0.3375	0.2831	0.2772
fruit	0.3029	0.3022	0.2860	0.2779
girl	0.3441	0.3449	0.3059	0.3081
glasses	0.2537	0.2522	0.2235	0.2324
lenna	0.2654	0.2647	0.2368	0.2412
mandrill	0.2471	0.2456	0.2353	0.2390
river	0.1904	0.1919	0.1941	0.1816

Table 4.2: The bit-error rates using JPEG compression of 75%.

Image	P_e Method 1 Eq. (3.5)	P_e Method 1 Eq. (3.32)	P_e Method 2 Eq. (3.38)	P_e Method 2 Eq. (3.55)
benz	0.3794	0.3794	0.3632	0.3691
bridge	0.3493	0.3485	0.3176	0.3243
building	0.3493	0.3500	0.2919	0.2956
camera	0.3654	0.3654	0.3434	0.3434
fruit	0.3750	0.3743	0.3221	0.3272
girl	0.4096	0.4110	0.3485	0.3515
glasses	0.3544	0.3566	0.3287	0.3279
lenna	0.3471	0.3463	0.3044	0.3022
mandrill	0.3596	0.3588	0.3375	0.3390
river	0.3162	0.3184	0.3081	0.3132

Table 4.3: The bit-error rates using JPEG compression of 60%.

Image	P_e Method 1 Eq. (3.5)	P_e Method 1 Eq. (3.32)	P_e Method 2 Eq. (3.38)	P_e Method 2 Eq. (3.55)
benz	0.1956	0.1949	0.1772	0.1654
bridge	0.0846	0.0824	0.0478	0.0449
building	0.1794	0.1794	0.1634	0.1331
camera	0.1794	0.1787	0.1515	0.1449
fruit	0.0934	0.0956	0.0721	0.0750
girl	0.1699	0.1691	0.1287	0.1346
glasses	0.0838	0.0831	0.0640	0.0640
lenna	0.0980	0.0985	0.0654	0.0654
mandrill	0.0669	0.0669	0.0537	0.0500
river	0.0801	0.0801	0.0471	0.0434

Table 4.4: The bit-error rates using a low-pass filter with $W_n = 0.99$.

Image	P_e Method 1 Eq. (3.5)	P_e Method 1 Eq. (3.32)	P_e Method 2 Eq. (3.38)	P_e Method 2 Eq. (3.55)
benz	0.4353	0.4360	0.3890	0.3860
bridge	0.4581	0.4588	0.4199	0.4250
building	0.4103	0.4096	0.3434	0.3485
camera	0.4382	0.4375	0.3860	0.3801
fruit	0.4221	0.4235	0.3463	0.3390
girl	0.4162	0.4154	0.3368	0.3309
glasses	0.4375	0.4353	0.3640	0.3500
lenna	0.4206	0.4176	0.3787	0.3765
mandrill	0.4426	0.4404	0.4360	0.4368
river	0.4382	0.4404	0.4059	0.3949

Table 4.5: The bit-error rates using a low-pass filter with $W_n = 0.9$.

CHAPTER 5

CONCLUDING REMARKS AND RECOMMENDATIONS

There are some areas of this analysis where future research would be very beneficial. One recommended topic for future research is to classify attacks and characterize each with an appropriate value of ρ for use with the statistical versions of the new detectors.

Another recommended topic for future research is to derive a statistical model for the methods using a different covariance matrix that encompasses more variables:

$$\Sigma = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \cdot \sigma^2 \tag{5.1}$$

In general, the decision rule will not be invariant to the choice of the covariance parameters like it is in the existing method. More parameters used in the covariance matrix may provide extra flexibility and better performance after a complete classification of attacks is developed. Furthermore, the methods discussed in this thesis use a detector which treats different sets of wavelet coefficient triples independently. Many attacks, in general, may not effect these triples independently and a joint maximization of a likelihood ratio for all sets of wavelet triples could be researched.

The last recommended topic for future research is the examination of higher embedding rates for the given watermarking methods. If the quantized values in Fig. 3.2 were not modulus two, but rather modulus 2^r, then there is a possibility

of increasing the embedding rate to $R = \frac{r}{3}$. Note, due to the basic engineering trade-offs, this will decrease the other two competing factors.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] M. D. Swanson, M. Kobayashi, A. H. Tewfik, "Multimedia Data-Embedding and Watermarking Technologies," Proceedings of the IEEE, Vol. 86, No. 6, pp. 1064-1087, June 1998

[2] Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn, "Attacks on copyright marking systems", in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98, Portland, Oregon, USA, April 15-17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239

[3] Fabien A. P. Petitcolas and Ross J. Anderson, "Evaluation of copyright marking systems" To be presented at IEEE Multimedia Systems (ICMCS'99), 7–11 June 1999, Florence, Italy

[4] Fabien A. P. Petitcolas. "Image Watermarking - StirMark". [Online] Available http://www.cl.cam.ac.uk/users/fapp2/steganography/image_watermarking/stirmark/

[5] R. B. Wolfgang, E. J. Delp, "A Watermarking Technique for Digital Imagery: Further Studies," Proc. of the Int. Conf. on Imaging Science, Systems, and Technology, pp. 279-287, June 30 - July 3, 1997

[6] B. Chen, G. W. Wornell, "An Information-Theoretic Approach To The Design of Robust Digital Watermarking Systems," Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 4, pp. 2061-2064, March 1999

[7] C. S. Burrus, R. A. Gopinath, H. Guo. "Introduction to Wavelets and Wavelet Transforms A Primer," Prentice-Hall, Inc., New Jersey, 1998

[8] D. Kundur, D. Hatzinakos, "Digital Watermarking Using Multiresolution Wavelet Decomposition," Proc. of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol.5, pp. 2969-2972, May 1998

[9] M.D. Srinath, P.K. Rajasekaran, R. Viswanathan, "Introduction to Statistical Signal Processing with Applications," Prentice Hall, Upper Saddle River, New Jersey, 1996

[10] A. Ambardar, "Analog and Digital Signal Processing," PWS Publishing Company, Boston, MA, 1995