12-15-2007

Watermarking With Wavelet Transforms

Kristen Michelle Parker

WATERMARKING WITH WAVELET TRANSFORMS

By

Kristen M. Parker

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfllment of the Requirements
for the Degree of Master of Science
in Engineering
in the Department of Electrical & Computer Engineering

Mississippi State, Mississippi

December 2007

WATERMARKING WITH WAVELET TRANSFORMS

By

Kristen M. Parker

Approved:

James E. Fowler Professor of Electrical & Computer Engineering (Director of Thesis)	Lori M. Bruce Professor of Electrical & Computer Engineering (Committee Member)
Nicholas H. Younan Professor of Electrical & Computer Engineering and Graduate Coordinator (Committee Member)	Roger L. King Associate Dean of Research and Graduate Studies James Worth Bagley College of Engineering

Name: Kristen M. Parker

Date of Degree: December 14, 2007

Institution: Mississippi State University

Major Field: Engineering (Electrical Engineering)

Major Professor: Dr. James E. Fowler

Title of Study: WATERMARKING WITH WAVELET TRANSFORMS

Pages in Study: 40

Candidate for Degree of Master of Science

Digital watermarking algorithms based on wavelet transforms provide increased performance and perceptual quality. This thesis proposes two wavelet-based schemes: one robust and one fragile. Robust watermarks should withstand attacks, such as compression, while maintaining the data integrity. The fr st approach presented is an algorithm which implements image watermarking in the domain of an overcomplete, or redundant, wavelet transform. Alternately, fragile watermarks are intended for use in applications wherein any loss of image quality is not acceptable. In the second approach presented, data embedding in the domain of an integer wavelet transform is considered. An algorithm is proposed that uses a bilevel image coder to compress a chosen bitplane, thereby providing space in which to store a payload while guaranteeing perfect image recovery.

ACKNOWLEDGMENTS

I am immensely grateful to my advisor, Dr. James E. Fowler, whose wisdom and knowledge guided me through my graduate-school achievements. I also want to thank the Electrical and Computer Engineering faculty and staff, particularly Dr. Lori Bruce and Dr. Nick Younan, for their willingness to always lend an ear and a helping hand. Finally, I am grateful to my parents for always being proud of me; my friends for believing in me; and my husband Brock, for his undying support and encouragement along the way.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Digital media provides a vital source of non-degradable, easily manipulated information. However, the ease at which digital images can be modifed makes the verifcation of image integrity of paramount importance. As a consequence, watermarking of images is becoming increasingly of interest in tasks such as copyright control, image identifcation, authentication, verifcation, and data hiding.

Spread-spectrum watermarking [1], one of the most popular methods for image watermarking, embeds a white-noise watermark into transform coeffcients of an image and verifes the presence of the watermark by measuring the correlation between the watermarked coeffcients and the watermark sequence. Spread-spectrum watermarking is intended to be *robust* such that the watermark is designed to survive attempts—intentional or unintentional—to remove it. Such robust watermarking is of use, for example, in copyright control in order to track the origin of illicit copies made of proprietary imagery. It has been shown that the discrete wavelet transform (DWT) is an effective venue for the spread-spectrum method due to natural similarities between the space-frequency tiling of the DWT and the operating characteristics of the human visual system (HVS) [2].

Although robust watermarking has been widely explored, other forms of watermarking exist. For example, invertible data embedding is a process of hiding a payload of information into a digital image such that, after the data payload is removed,

the image is fully and losslessly restored to its original state. Although payloads of a general and arbitrary nature can be carried in such an invertible data-embedding method, when the payload consists of a cryptographic hash of the image itself, the data embedding can be considered to be a form of *fragile watermarking*, because, if the image is manipulated or modifed in any way, such modifcation can be detected via a mismatch between calculated and embedded hashes [3]. Unlike robust watermarking which aims to protect against watermark removal, fragile watermarking is often intended for use in applications where no amount of distortion is acceptable, and perfect recovery of the image is required. The fragile watermark thus detects unauthorized image manipulation or modifcation in these applications. Examples would include medical imaging or military surveillance where normal HVS rules do not necessarily apply and even small distortions will be scrutinized by analysts and examiners. Thus, a guarantee of perfect image recovery is an important characteristic for data embedding in such applications.

The main contributions of this thesis are two algorithms—one for robust watermarking and one for fragile watermarking. The fr st algorithm presented implements robust watermarking in the domain of an overcomplete, or redundant, wavelet transform. This algorithm expands on a previous method, pixel-wise masking (PWM) [2], which employs a traditional, critically sampled wavelet transform coupled with perceptually-based watermark casting and optimal Neyman-Pearson detection. As an alternative to the DWT, the redundant discrete wavelet transform (RDWT) [4–6] has also been considered for watermarking [7, 8]. From a mathematical perspective, the RDWT is a frame expansion, and frame expansions have long been known to be robust to added noise. Since, in spread-spectrum watermarking, the watermark signal takes the form of added noise, the RDWT is particularly attractive as a venue for watermarking.

This thesis describes an RDWT-domain version of the PWM technique of [2]; the resulting PWM-RDWT algorithm—originally developed in [9]—is shown to provide greater robustness than the original PWM of [2].

The second main contribution of this thesis takes the form of an approach to fragile watermarking, or, more generally, invertible data embedding. The second method uses a bilevel image coder to compress a chosen bitplane, thereby providing space in which to store a payload while guaranteeing perfect image recovery. While similar prior approaches (e.g., [10]) embed data in a simple 1D Haar transform, the method proposed in this thesis uses a 2D integer-valued transform with longer flter s, signifcantly reducing the distortion incurred by data embedding. Unlike the 1D Haar, though, embedding in the 2D transform cannot be applied to all images due to overfo w and underfo w that can occur as a result of the embedded data. However, on images in which embedding is successful, there is considerable gain in image quality and payload size. The proposed approach can be used in a system in which the conventional 1D Haar embedding is used as a fall-back in the event that embedding fails in the proposed scheme. In this fashion, the proposed algorithm is "nearly invertible"—i.e., it is invertible in cases in which the embedding is successful, which we hope to be true for most images. As a consequence, we denote our proposed approach to fragile watermarking as nearly invertible data embedding (NIDE).

An overview of the remainder of this thesis is as follows. The next chapter presents a review of watermarking methods, providing an in-depth look at the robust-watermarking technique PWM-DWT of [2] as well as the invertible-watermarking approach using the 1D Haar of [10]. Subsequently, Chap. III presents the fr st major contribution of this thesis, the PWM-RDWT technique; this algorithm is described in detail and

experimental results are presented. Next, in Chap. IV, the second major contribution, NIDE watermarking, is proposed and discussed along with experimental results. Finally, Chap. V offers conclusions drawn from the results of the experiments of the preceding two chapters.

CHAPTER II

BACKGROUND

The use of watermarking in digital images can be divided into two groups based on application. The fr st group—robust watermarking—strives to preserve the data embedded into the image, while the second—fragile watermarking—strives to preserve the image itself. In this chapter, we review two approaches to watermarking, one from each group. First, we discuss the robust pixel-wise masking (PWM) of [2] followed by an overview of fragile watermarking using the invertible 1D Haar transform [10].

2.1 Robust Watermarking

In spread-spectrum watermarking [1], a low-energy noise signal is added to transform coeffcients, such that, once the inverse transform is taken, the addition of the watermark is visually imperceptible. Then, to test for the presence of a known watermark in a given image, the watermark is correlated with the transform coeffcients of the image in question. A large magnitude output from the correlator indicates the presence of the desired watermark; a near-zero output indicates the absence of the watermark. A threshold can be set to determine the decision between these two using classical Neyman-Pearson detection theory [2]. A spread-spectrum watermark is designed to withstand attempts to remove it; such attacks may be intentional, as in the case of someone explicitly attempting to circumvent copy controls, or unintentional, as in the case of incidental processing, like compression, being applied to the image. In general, there is a tradeoff between the robustness of the watermark and its visibility—

the stronger the noise power, the more likely the watermark is to survive an attack, but the more likely it is that the watermark will be visible.

In spread-spectrum watermarking, the goal is to embed as much watermark noise into an image as possible so as to maximize the correlation-detector performance, thereby maximizing robustness, while simultaneously leaving the perceptual quality of the image unchanged. As a consequence, the guiding principle of perceptually-based spread-spectrum watermarking is that the watermark energy should be placed in locations that are the least perceptible to the human visual system (HVS). Locating those least-perceptible areas accurately is key to placing large amounts of watermark information into the image.

The PWM method as proposed in [2] is deployed in the domain of a discrete wavelet transform (DWT); henceforth, we refer to it as PWM-DWT. In PWM-DWT, the perceptibility of each DWT coeffcient is determined from the model of the HVS originating in [11]. This model consists of three components—orientation and level of detail, local brightness, and local texture—which are combined in a product expression that is then used as a weighting factor for the watermark information during watermark casting. The product expression, in fact, computes a weighting factor $w_l^\theta(i,j)$ for the coeffcient at location (i,j) in the subband at orientation $\theta \in \{0,1,2\}$ in the decomposition level l. The weighting factor itself is

$$w_l^\theta(i,j) = \frac{1}{2}\Theta(l,\theta)\Lambda(l,i,j)\Xi(l,i,j)^{0.2}, \tag{2.1}$$

consisting of the product of three distinct factors corresponding to three HVS effects.

The fr st factor in (2.1) accounts for HVS perceptibility of a DWT subband based on its level and orientation. This factor is defned as

$$\Theta(l,\theta) = \quad(\theta)\ \ (l), \tag{2.2}$$

where

$$(\theta) = \begin{cases} \sqrt{2}, & \theta = 1, \\ 1, & \text{else}, \end{cases} \tag{2.3}$$

and

$$(l) = \begin{cases} 1.00, & l = 0, \\ 0.32, & l = 1, \\ 0.16, & l = 2, \\ 0.10, & l = 3. \end{cases} \tag{2.4}$$

These (θ) and (l) values were determined via perceptual experiments in [11].

The second factor in (2.1) accounts for perceptibility based on local brightness. This factor is

$$\Lambda(l,i,j) = 1 + L'(l,i,j), \tag{2.5}$$

where

$$L'(l,i,j) = \begin{cases} 1 - L(l,i,j), & L(l,i,j) < 0.5, \\ L(l,i,j), & \text{else}, \end{cases} \tag{2.6}$$

and

$$L(l,i,j) = \frac{1}{256} I_3^3 \left(1 + \left\lfloor \frac{i}{2^{3-l}} \right\rfloor, 1 + \left\lfloor \frac{j}{2^{3-l}} \right\rfloor \right). \tag{2.7}$$

Here, I_l^θ is the DWT subband at orientation θ and scale l, so I_3^3 corresponds to the baseband subband in a 4-level DWT decomposition.

Finally, the third factor in (2.1) estimates image texture for a coeffcient at scale l located spatially at (i, j) by examining the variance in co-located 2×2 blocks in each subband in the DWT,

$$\Xi(l,i,j) = \left[\sum_{k=0}^{3-l} \frac{1}{16^k} \sum_{\theta=0}^{2} \sum_{x=0}^{1} \sum_{y=0}^{1} \left[I_{k+l}^{\theta} \left(y + \frac{i}{2^k}, x + \frac{j}{2^k} \right) \right]^2 \right] \cdot \mathrm{Var}\left\{ I_3^3 \left(y + \frac{i}{2^{3-l}}, x + \frac{j}{2^{3-l}} \right) \right\}_{x,y=0,1}, \quad (2.8)$$

where I_l^{θ} is the DWT subband at orientation θ and scale l. We note that, in [2], watermarking is applied only to the highest-resolution subbands (i.e., for $l = 0$) in order to minimize overall perceptibility of the watermark.

As used in (2.8), fx ed-size blocks in a DWT subband correspond to increasingly larger spatial areas in the original image as the resolution of the subband decreases (l increases), resulting in the texture measure being less local for the lower-resolution subbands. Fig. 2.1 illustrates this effect in a two-scale DWT, wherein it can be seen that the fx ed-sized blocks cover an increasingly larger spatial area as l increases. Additionally, the size of the blocks that can be used in the texture estimation of (2.8) is limited in practice since the blocks should not become larger than the size of the lowest-resolution subbands when the number of scales of decomposition is large.

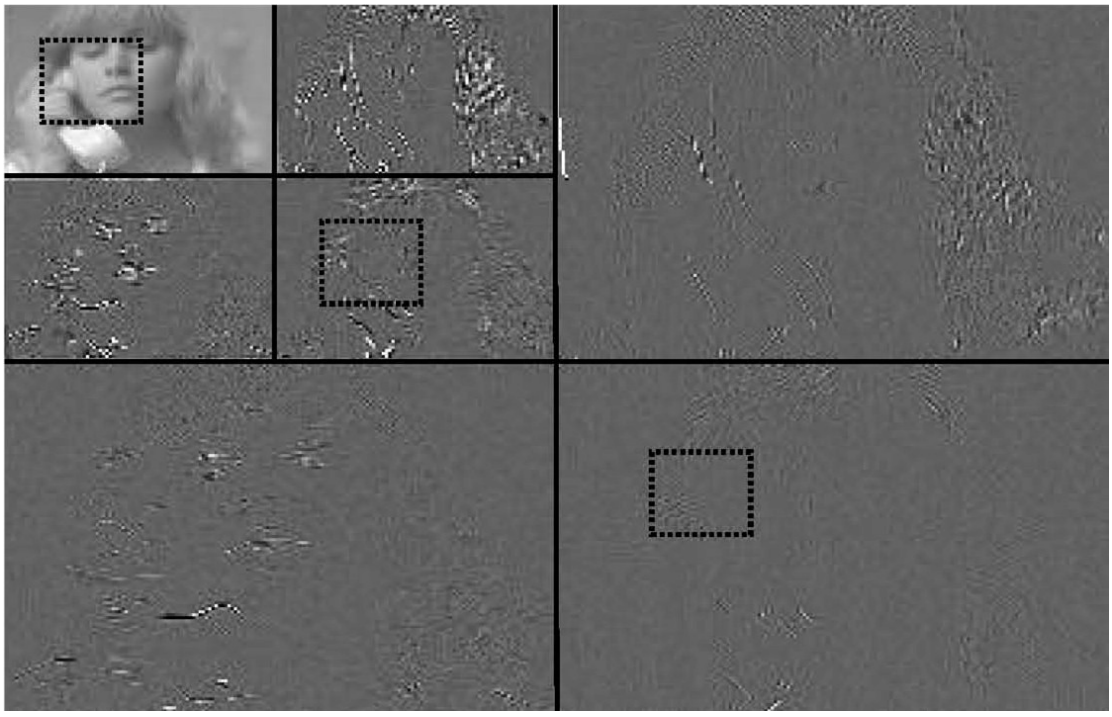


Figure 2.1: Spatial area of fx ed-size blocks in a two-scale DWT.

The detector used in the PWM-DWT technique of [2] follows the correlation approach to watermark detection typical of spread-spectrum watermarking—correlation is calculated between the transform coeffcients and the desired watermark, with a large magnitude correlation indicating presence of the watermark. That is, correlation ρ is

$$\rho = \frac{1}{3MN} \sum_{\theta=0}^{2} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \widetilde{I}_0^{\theta}(i,j) x^{\theta}(i,j), \tag{2.9}$$

where $\widetilde{I}$ are the DWT coeffcients of the watermarked image, and x is the watermark that is to be detected. In this manner, PWM-DWT uses blind watermark detection (i.e., the detector does not have access to the original image), and a Neyman-Pearson approach is employed to minimize the probability of missed detection of the watermark given a fx ed false-detection probability. In [2], the Neyman-Pearson detection threshold was determined to be

$$T_\rho = 3.97\sqrt{2\sigma_\rho^2} \tag{2.10}$$

for a false-detection probability of 10^{-8} under the assumption that the correlation-detector output, ρ, is normally distributed, with σ_ρ^2 being the variance of ρ when the image is watermarked with some watermark other than the target watermark.

2.2 Fragile Watermarking and Invertible Data Embedding

Although robust watermarking via the spread-spectrum approach as described above accounts for much work in the watermarking area, there has also been increasing interest in developing fragile watermarks. The most prominent work in the area of fragile watermarking for images is the invertible data-embedding scheme implemented through spatial-domain least-signifcant-bit (LSB) compression as proposed in [3]. Fragile watermarking via spatial-domain LSB involves a hash calculated over the image, compression of an LSB bitplane using a lossless bilevel coder such as JBIG [12], and

embedding of the resulting bitstream (hash plus JBIG-compressed bitplane) in place of the LSB bitplane. This process is repeated on each bitplane, starting with the least-signifcant bitplane, until a bitplane is found that compresses enough to allow for the image hash to be accommodated. Although originally proposed in the context of image authentication in [3], this general embedding procedure could be applied by replacing the hash with any arbitrary payload; the data to be embedded then consists of the payload plus a compressed bitplane that allows recovery of the bits modifed by the embedding process.

Invertible data embedding was originally conducted in the spatial domain in [3]; a number of subsequent approaches considered embedding in the domain of a wavelet transform. Most of these wavelet-based approaches are built upon Tian's difference expansion (DE) scheme [10] which, in essence, consists of a 1D Haar transform applied horizontally to image rows. The transform itself is a lifting-based integer-to-integer implementation (e.g., [13]) guaranteeing transform reversibility. Careful embedding into the least signifcant bitplane of the wavelet coeffcients ensures that overfo w and underfo w beyond the valid $[0, 255]$ interval for 8-bit pixels is avoided.

The DE watermark proposed by Tian [10] consists of a simple integer-valued, lifting-based Haar applied in 1D (the transform is applied horizontally to rows of image pixels). The reversible integer-valued Haar wavelet transform, assuming an 8-bit grayscale pair of horizontally adjacent pixels x and y, is

$$l = \left\lfloor \frac{x+y}{2} \right\rfloor, \qquad h = x - y. \tag{2.11}$$

The inverse transform is

$$x = l + \left\lfloor \frac{h+1}{2} \right\rfloor, \qquad y = l - \left\lfloor \frac{h}{2} \right\rfloor. \tag{2.12}$$

Due to the simplicity of the Haar scheme, it is possible to modify only those coeffcients that will not cause overfo w or underfo w upon the inverse transform. By restricting x and y in the range of $[0, 255]$, overfo w and underfo w is prevented; conditions for this to hold are

$$0 \leq l + \left\lfloor \frac{h+1}{2} \right\rfloor \leq 255, \tag{2.13}$$

and

$$0 \leq l - \left\lfloor \frac{h}{2} \right\rfloor \leq 255. \tag{2.14}$$

Invertibility for the data embedding is assured for any image, since the decoder can identify which pixels the encoder determined were susceptible to overfo w/underfo w to correctly extract the embedded payload.

While DE-based embedding ensures perfect recovery while successfully avoiding overfo w/underfo w issues, the 1D Haar transform used is not particularly effcient as an image transform. Ideally, one would prefer an integer-to-integer version of a 2D transform in popular use; the integer-valued 5/3 transform from the JPEG2000 standard [14] would be a reasonable candidate. One would expect compression of bitplanes in such a 2D wavelet transform to be more effcient, allowing embedding into a lower bitplane than is possible in the original spatial domain or in the 1D Haar. In this case, a lower distortion can be achieved from the embedding process; equivalently, a larger payload can be embedded for a given distortion. However, preventing overfo w/underfo w in transforms more sophisticated than Tian's 1D Haar is non-trivial [15].

Zou *et al*. [16] propose one approach to invertible data embedding using a 2D integer-value 5/3 transform instead of the 1D Haar of [10]. In essence, blocks of wavelet coeffcients are modifed by shifting their block means in order to embed bits from the payload. In order to prevent overfo w/underfo w, blocks that could produce these effects

are detected and not modifed. This results in an occasional "error" being made in the embedding process; i.e., avoiding overfo w/underfo w causes the "wrong" bit to be embedded for these blocks. However, an error-correcting code is added to the embedded bitstream to detect and correct these bit errors. It is expected that, for most images, few errors of this sort will need corrected, as the overfo w/underfo w issue does not arise for most blocks [16]. However, it is possible that, if overfo w/underfo w does occur, several blocks may be affected. Such a "bursty" error pattern may surpass the abilities of the error-correcting code, resulting ultimately in failure to losslessly recover the image.

In this chapter, we have overviewed techniques from the two main classes of strategies for image watermarking—robust watermarking and fragile watermarking. In the subsequent two chapters, we propose two watermarking algorithms, one from each class. We start in the next chapter with a proposed modifcation to the robust PWM technique of [2] in which we replace the critically sampled DWT used originally with a redundant transform.

CHAPTER III

PIXEL-WISE MASKING USING THE RDWT

Spread-spectrum watermarking [1], one of the most popular methods for image watermarking, embeds a white-noise watermark into transform coeffcients of an image and verifes the presence of the watermark by measuring the correlation between the watermarked coeffcients and the watermark sequence. It has been shown that the discrete wavelet transform (DWT) is an effective venue for the spread-spectrum method due to natural similarities between the space-frequency tiling of the DWT and the operating characteristics of the human visual system (HVS) [2].

As an alternative to the DWT, the redundant discrete wavelet transform (RDWT) [4–6] has also been considered for watermarking [7, 8]. In essence, the RDWT—often implemented as the *algorithme à trous* [4, 5]—removes the downsampling operation from the DWT to produce an overcomplete and shift-invariant transform. From a mathematical perspective, the RDWT is a frame expansion, and frame expansions have long been known to be robust to added noise. Specifcally, white noise added in the transform domain results in signifcantly reduced noise power in the original signal domain due to the fact that the inverse frame operator is a pseudo-inverse that involves a projection onto the range space of the forward transform [17].

Intuitively, one would expect that the robustness to noise provided by frame expansions such as the RDWT would be ideally suited to the spread-spectrum watermarking procedure. Indeed, more watermarking energy can be accommodated

in the RDWT domain for the same distortion incurred in the original signal domain as compared to traditional DWT-based watermarking [8]. However, it has been shown that the same pseudo-inverse projection that decreases the noise power also results in a corresponding decrease in correlation-detector performance, such that overcomplete and complete transforms offer the same watermarking performance from a theoretical perspective [8].

Still, the redundancy provided by the RDWT can be exploited in ways other than for noise robustness. Since the redundancy in the transform facilitates the location of edges and other salient features in an image [18], it has been argued that the RDWT domain is well-suited for perceptually guiding the casting of watermarks [7]. In this chapter, we demonstrate this advantage using the well-known, perceptually-based watermarking method, the pixel-wise masking (PWM) technique originating in [2] and described in this thesis in Sec. 2.1. PWM was originally formulated with the critically sampled DWT (i.e., PWM-DWT as we call it here); in this section, we adapt PWM-DWT to the overcomplete RDWT, producing what we call PWM-RDWT. We note that the discussion and results of this chapter originated in [9].

3.1 PWM-RDWT Watermark Casting

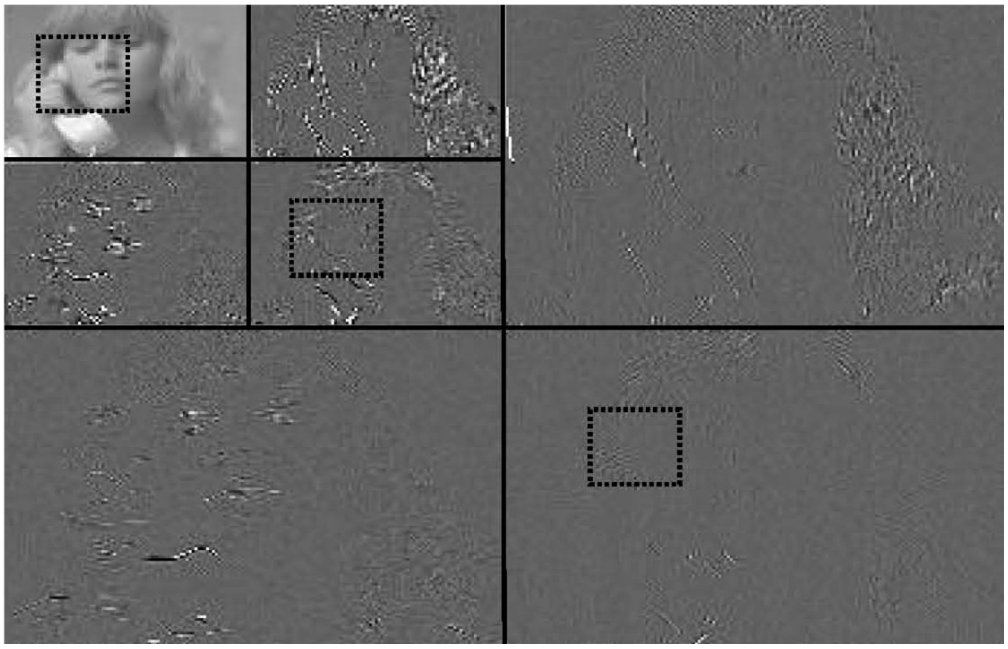
Fundamental to PWM-DWT is the determination of the perceptibility of each DWT coeffcient through use of an HVS model involving orientation and level of detail, local brightness, and local texture. In PWM-DWT, the local texture perceptibility is calculated in a blockwise manner using co-located 2×2 blocks in each subband of the DWT as described in (2.8). However, in the DWT, fx ed-size blocks correspond to varying spatial areas in the original image, depending on the resolution level of the subband in which the block in question resides.

Because the RDWT is not downsampled, the subbands have the same size as the original image for each level of decomposition; therefore, decreasing subband resolution does not increase the spatial area associated with a fx ed-size block. Additionally, we can employ a larger block size with the current coeffcient itself as the center (the 2×2 blocks of PWM-DWT are offset relative to the current coeffcient). Consequently, more accurate estimation of local texture activity surrounding the current coeffcient can be achieved in the RDWT domain. In our proposed PWM-RDWT method, we replace (2.8) with

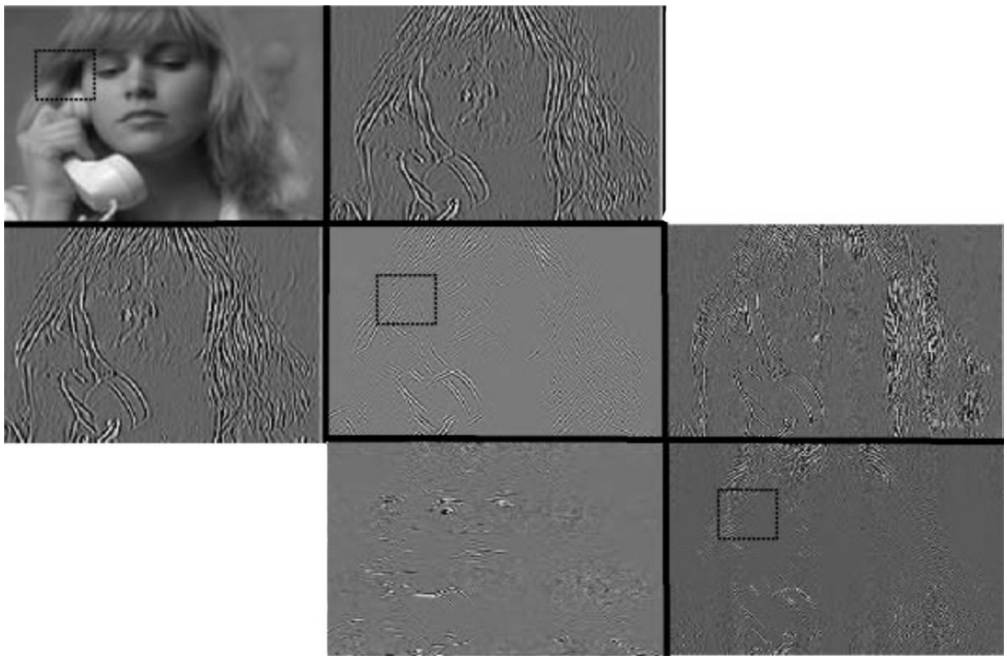
$$\Xi(l,i,j) = \left[\sum_{k=0}^{3-l} \frac{1}{16^k} \sum_{\theta=0}^{2} \sum_{x=-n}^{n} \sum_{y=-n}^{n} \left[I_{k+l}^{\theta}\left(y+i, x+j\right)\right]^2\right] \cdot \operatorname{Var}\left\{ I_3^3\left(y+i, x+j\right)\right\}_{x,y=-n,\ldots,n}, \quad (3.1)$$

wherein we have assumed $n \times n$ blocks centered about the current coeffcient. Fig. 3.1 illustrates the difference in character of fx ed-size blocks between the DWT and RDWT domains. Fig. 3.1(b) shows that fx ed-sized blocks in an RDWT correspond to the same spatial area in each subband, in contrast to the varying spatial area of the fx ed-sized blocks in the DWT of Fig. 3.1(a).

We note that the remaining components of the PWM procedure, i.e., (2.1)–(2.7), are unaffected by the use of the RDWT and so remain unchanged from those of [2]. We note also that, as in [2], we watermark only the transform subbands with the highest resolution as a compromise between robustness and perceptual invisibility.



(a)



(b)

Figure 3.1: Spatial area of fx ed-size blocks, (a) two-scale DWT, (b) two-scale RDWT.

3.2 PWM-RDWT Watermark Detection

In [8], the case of watermarking with a tight-frame expansion was considered, and it was determined that the Neyman-Pearson threshold for the tight-frame case, T'_ρ, is related to the Neyman-Pearson threshold for the critically-sampled case, T_ρ, as

$$T'_\rho = \sqrt{A}\, T_\rho, \tag{3.2}$$

where A is the frame bound for the tight-frame expansion, and it is assumed that the false-detection probability is the same in both cases.

For our PWM-RDWT technique, we observe that the RDWT is a tight frame only when one level of decomposition is used [19, 20]. However, if the watermark is cast into only the highest-resolution subbands of the transform (as was done in the PWM-DWT approach of [2] and in our implementation of PWM-RDWT), then the Neyman-Pearson threshold for PWM-RDWT will be given approximately by (3.2) with $A = 4$ for a 2D transform and T_ρ being the threshold used for PWM-DWT as given by (2.10). In the following results, we verify experimentally the validity of this approximation to the optimal threshold.

3.3 Experimental Results

We compare our proposed PWM-RDWT to the PWM-DWT technique of [2]. For both techniques, we adjust the watermark strength to the level of just-noticeable distortion (JND), and evaluate watermark detection using correlation-based detection with a Neyman-Pearson threshold as described above. All transforms are implemented using the popular biorthogonal 9/7 wavelet [21] with symmetric extension. We initially fx the block size for PWM-RDWT to $n = 3$.

For PWM-RDWT, the detector response, correlation ρ, is calculated for 1000 different watermarks—only one being the correct embedded watermark—and the resulting detector responses shown in Fig. 3.2. The magnitude of the correct watermark is comparatively much larger than any of the "incorrect" watermarks.

Fig. 3.3 compares the detector response to the correct watermark for PWM-RDWT and PWM-DWT under attack with SPIHT [22] compression. Also shown are the Neyman-Pearson thresholds given by (3.2) and (2.10), respectively. As is evident in Fig. 3.3, the PWM-DWT detector response falls below its optimal threshold at a compression ratio of 61, while PWM-RDWT does not cross its threshold until a compression ratio of 130. The amount of watermark information embedded using the PWM-RDWT method was considerably larger due to the more accurate locating of pixels masked by the HVS.

Fig. 3.4 demonstrates the validity of (3.2) as an approximation to the ideal threshold. We see that the second-highest detector response is consistently below the approximate threshold (3.2) as the compression ratio varies.

Finally, we consider a block size of $n = 5$ in Fig. 3.5. As can be seen, increasing the block size results in slightly greater robustness (a compression ratio of roughly 140 can be withstood using a block size of $n = 5$). We have also tested larger block sizes, but did not observe further improvement beyond $n = 5$.

In this chapter, we have considered an algorithm for robust watermarking taking place in the domain of the overcomplete RDWT. In contrast, in the next chapter, we will consider watermarking in the domain of the critically sampled DWT; however, our focus will be on fragile watermarking instead.

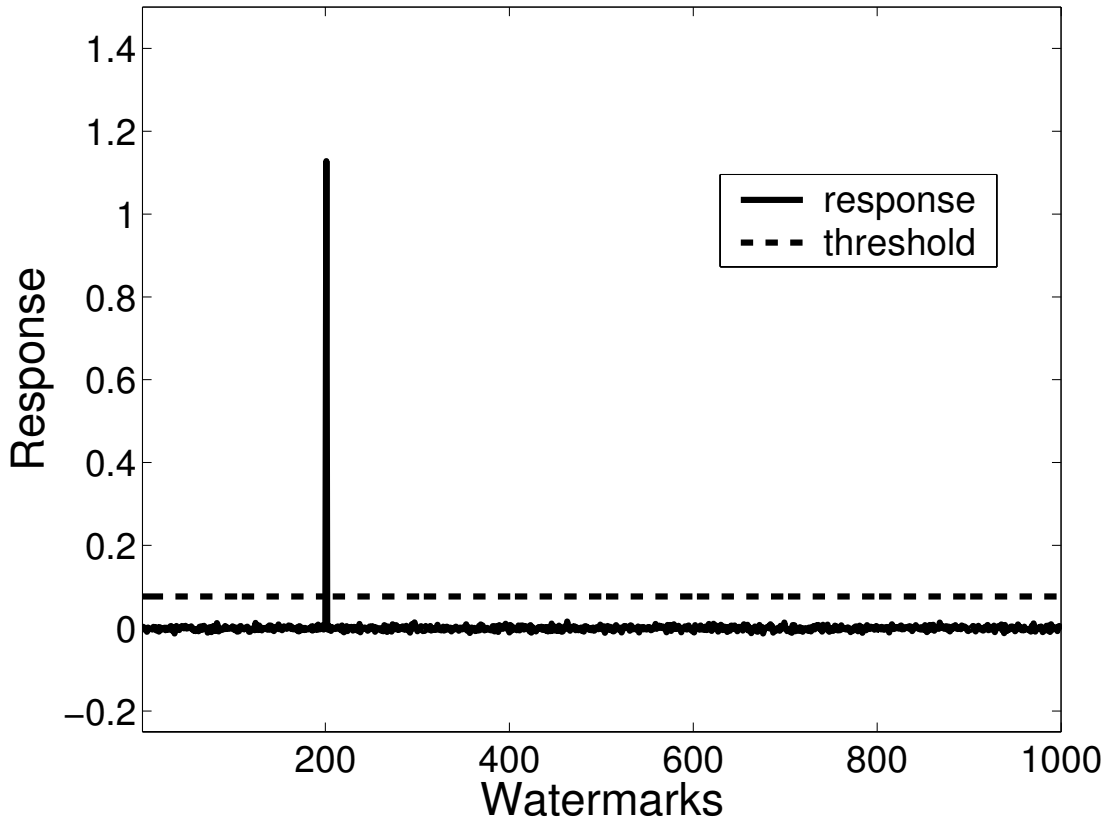


Figure 3.2: Detector response for 1000 different watermarks for the "Lenna" image with PWM-RDWT; $n = 3$.

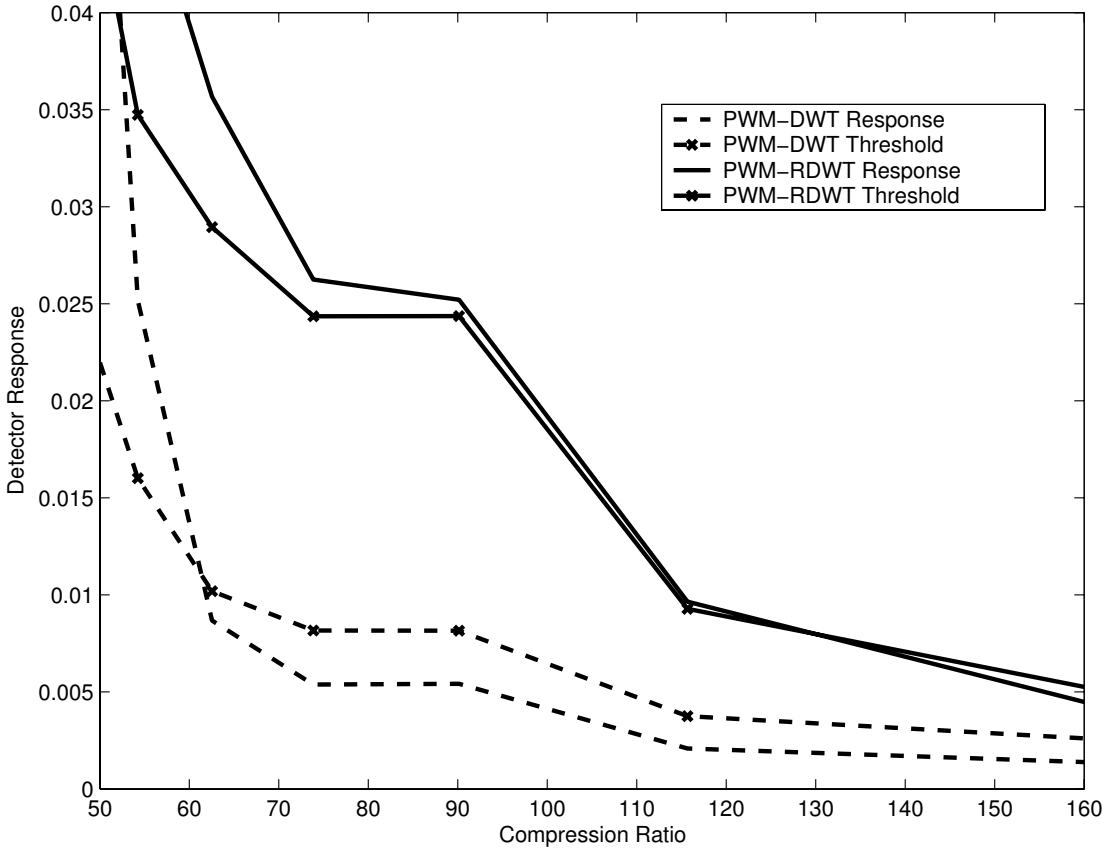


Figure 3.3: Detector performance for PWM-RDWT and PWM-DWT [2] for the "Lenna" image under compression with SPIHT [22]; $n = 3$.

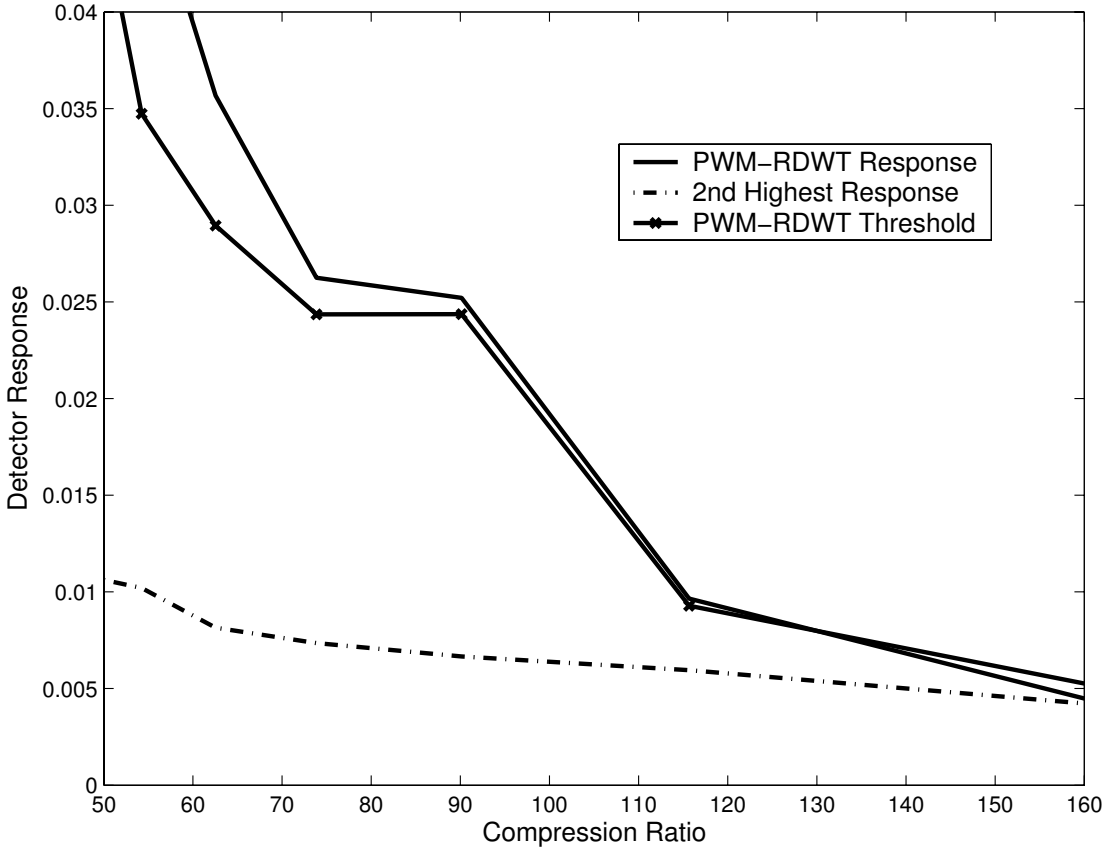


Figure 3.4: Detector response and second-highest response for the "Lenna" image under compression with SPIHT [22]; $n = 3$.

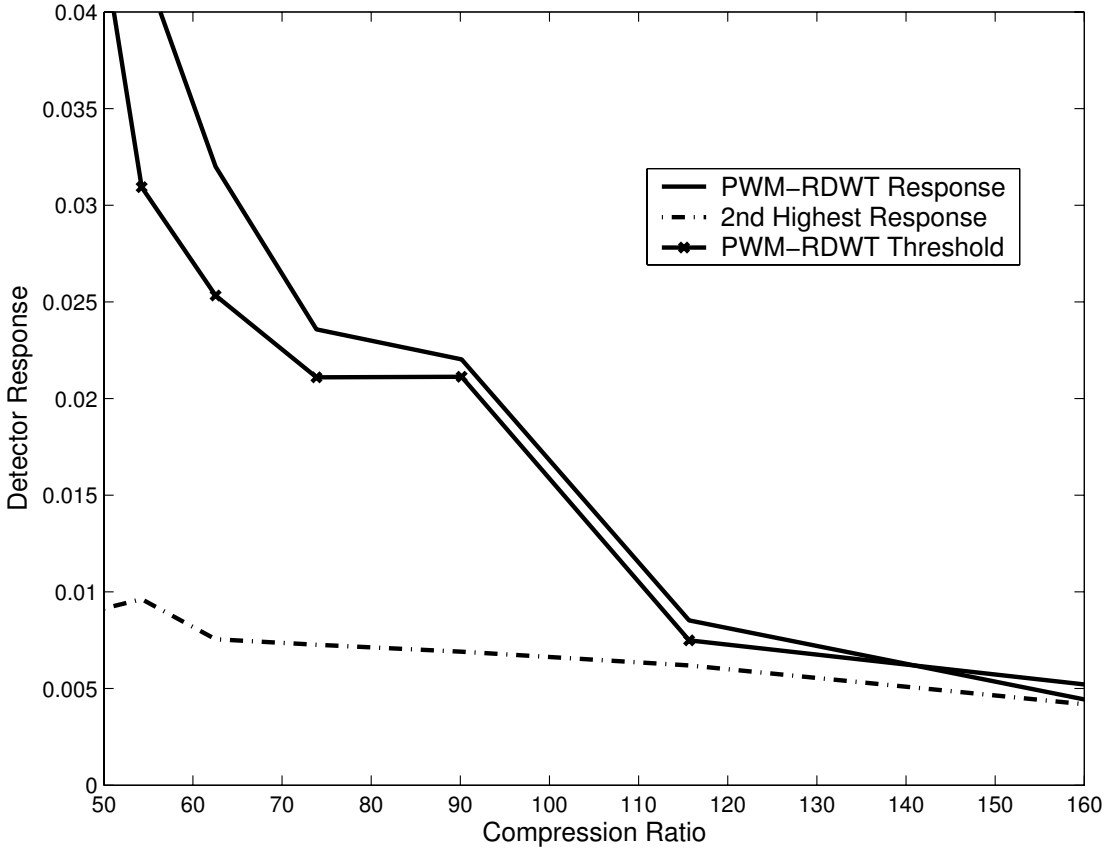


Figure 3.5: Detector response and second-highest response for the "Lenna" image under compression with SPIHT [22]; $n = 5$.

CHAPTER IV

NEARLY INVERTIBLE DATA EMBEDDING

While robust watermarking, such as the PWM-RDWT technique proposed in the preceding chapter, accounts for much activity in the watermarking feld, there is increasing interest in fragile watermarking and invertible data embedding. Invertible data embedding can be used in a variety of applications to embed various data payloads into an image while permitting perfect image recovery without loss after the payload is extracted. When such a payload contains a cryptographic hash of the image, such data embedding is commonly referred to as fragile watermarking; such fragile watermarking permits image authentication and verifcation. Although originally conducted on spatial-domain pixels, fragile watermarking, like its robust counterpart, has also been applied in the domain of a wavelet transform. Unlike robust watermarking, though, it is a challenge to apply fragile watermarking in arbitrary transforms and retain the ability to recover the original image.

Most wavelet-domain approaches (e.g, [23]) to fragile watermarking are based on the difference-expansion (DE) scheme of [10], embedding data in a simple 1D Haar transform taken row-wise in an image. In this chapter, however, we focus on the use of a 2D integer-valued transform with longer flter s which signifcantly reduces the distortion incurred by data embedding from that of the 1D Haar.

Specifcally , in this chapter, we adopt the 2D integer-valued 5/3 transform which is popular for invertible applications such as lossless compression within the JPEG2000

standard [14]. Like other fragile-watermarking techniques, we perform embedding similar to that of [3] in the wavelet-coeffcient magnitudes. However, unlike [10, 16, 23], we make no special accommodations to avoid overfo w or underfo w of pixel values. That is, we expect that, as observed in [16], overfo w/underfo w should be rarely encountered. Of course, in the event overfo w or underfo w occurs for some image, our proposed embedding scheme fails to embedded the payload. For this reason, we refer to our proposed approach as "nearly invertible" data embedding (NIDE)—the scheme is guaranteed to be invertible and yield lossless image recovery, provided that embedding was successful in the fr st place, which we hope to be true for many images.

4.1 The NIDE Algorithm

The goal in invertible data embedding is to maximize payload size as well as to minimize the distortion between the data-embedded and original images while ensuring that the original image is exactly recoverable from the data-embedded image. Improving upon performance of previous methods of invertible data embedding based on DE, we employ a more sophisticated 2D transform, the integer-valued 5/3 DWT. Bitplane compression and payload embedding take place in a bitplane from the magnitudes of the DWT coeffcients.

Specifcally , let B_i be the bilevel bitplane image created by extracting bitplane i from an image of size $N \times M$ pixels. Let $\mathrm{JBIG}(B_i)$ be the bitstream generated by applying JBIG compression [24] to B_i and defne the redundancy of bitplane i as

$$R_i = NM - |\mathrm{JBIG}(B_i)|\,, \tag{4.1}$$

where $|\cdot|$ indicates length in bits. R_i measures how many bits of payload can be accommodated in bitplane B_i.

The NIDE algorithm using the 2D DWT can be described as follows, starting with $i = 0$:

1. Employ the forward integer 5/3 DWT to the original image
2. Extract bitplane B_i from the magnitudes of the DWT coeffcients
3. Apply JBIG to compress the selected bitplane
4. Calculate R_i via (4.1)
5. If $R_i \geq |P|$, where P is the payload bitstream, replace B_i with $\mathrm{JBIG}(B_i) \circ P$; otherwise, $i \leftarrow i + 1$ and goto 2 ("$\circ$" denotes bitstream concatenation)
6. Perform inverse 5/3 DWT

Therefore, the fnal bitplane will consist of the compressed bitplane followed by a payload; random bits can be concatenated, as many as is needed to fll out to the end of the bitplane, in the case that $R_i > |P|$. Clearly, the distortion decreases as the bitplane for embedding decreases—a higher bitplane will allow for more payload, at the cost of increased distortion.

Although the integer DWT is perfectly lossless, the drawback to using this transform occurs when coeffcients are modifed. Specifcally, after the inverse transform, it is possible that some of the spatial-domain pixel values will be greater than 255 or less than 0, even though the original image was confned to the range of $[0, 255]$. Simply "clipping" the pixels to $[0, 255]$ before outputting the fnal image is not an option, as such an action would likely render the embedded bitstream undetectable. The simple nature of the 1D Haar as used in the DE scheme of [10] permits one to determine exactly which coeffcients are at risk for producing overfo w or underfo w; one can then avoid embedding data into those coeffcients. However, the 2D 5/3 transform

is much more complex, and such overfo w/underfo w prediction is not possible to the best of our knowledge. It is possible, however, to detect the occurrence of overfo w or underfo w after the fact; that is, the watermark-embedding process can determine whether watermark insertion was successful or not since the overfo w and underfo w conditions can be detected as they arise when conducting the inverse 5/3 transform.

We anticipate that the nearly-invertible paradigm will be useful in a number of situations. In applications in which image capture or generation can be controlled, the dynamic range of the image pixels can be reduced so as to provide a "margin" about 0 and 255 to guard against overfo w and underfo w. Alternatively, DE could be employed as a "fall-back" method when our embedding fails. That is, should overfo w/underfo w occur, it will be known by the data embedder, in which case the image can be re-embedded using DE. During detection of the payload and image recovery, if a valid payload is not found using our approach, one can assume the fall-back DE method was used and search for a DE-embedded payload. In this manner, we gain the benefts of the 2D 5/3 transform, in the absence of overfo w/underfo w, but can embed in some way into all images.

4.2 Experimental Results

We compare the 2D 5/3 transform of our NIDE method to the 1D Haar transform as used in DE. We focus on merely the effcienc y of the transform used, employing essentially identical embedding schemes within each transform domain in order to measure the increase in effcienc y arising from the 2D 5/3 transform. For both techniques, we embedded data into the lowest bitplanes that have positive redundancy and could, therefore, support a payload. For our experimental results, we use the "Lenna," "Goldhill," and "Barbara" images, shown in Figs. 4.1–4.3, which happen to have pixel ranges of around $[15, 245]$ instead of the full dynamic range of $[0, 255]$; this

reduced range will help NIDE avoid overfo w and underfo w. Throughout, we measure distortion as the peak signal-to-noise ratio (PSNR) in dB.

For “Lenna,” we were able to embed into bitplanes 2 and 3 for NIDE and only bitplane 3 for the 1D Haar. Fig. 4.4 illustrates PSNR versus the number of bits embedded into the image (i.e., payload size). The payload embedded for the 2D transform is approximately 2.5 times the length of payload that can be accommodated by the 1D Haar for an equivalent PSNR. Alternatively, comparing approximately the same payload size of 60 kbits, the PSNR is over 5 dB higher for the 2D transform.

Fig. 4.5 compares bitplane 3 for the 1D Haar to bitplane 2 for the 2D 5/3 for approximately equal payload sizes (about 27 kbits). The 2D 5/3 method shows signifcantly less distortion, at close to 38 dB, compared to the 1D Haar method, with 32 dB. We also successfully tested the method on “Barbara” (Fig. 4.6) with similar results, embedding a payload of 46 kbits for a PSNR of 37.8 dB.

In this chapter, we have presented an approach to invertible data embedding and fragile watermarking that is “nearly invertible”—in the likely case that embedding is successful, the integer 2D 5/3 DWT used is signifcantly more effcient that the 1D Haar transform that underlies the DE approach of [10] that forms the basis of many wavelet-based fragile-watermarking schemes. In the next chapter, we make some fnal concluding remarks concerning our NIDE approach as well as the robust PWM-RDWT technique of the preceding chapter.

Figure 4.1: Test image used for NIDE 2D DWT verifcation, “Lenna.”

Figure 4.2: Test image used for NIDE 2D DWT verifcation, “Barbara.”

Figure 4.3: Test image used for NIDE 2D DWT verifcation, “Goldhill.”

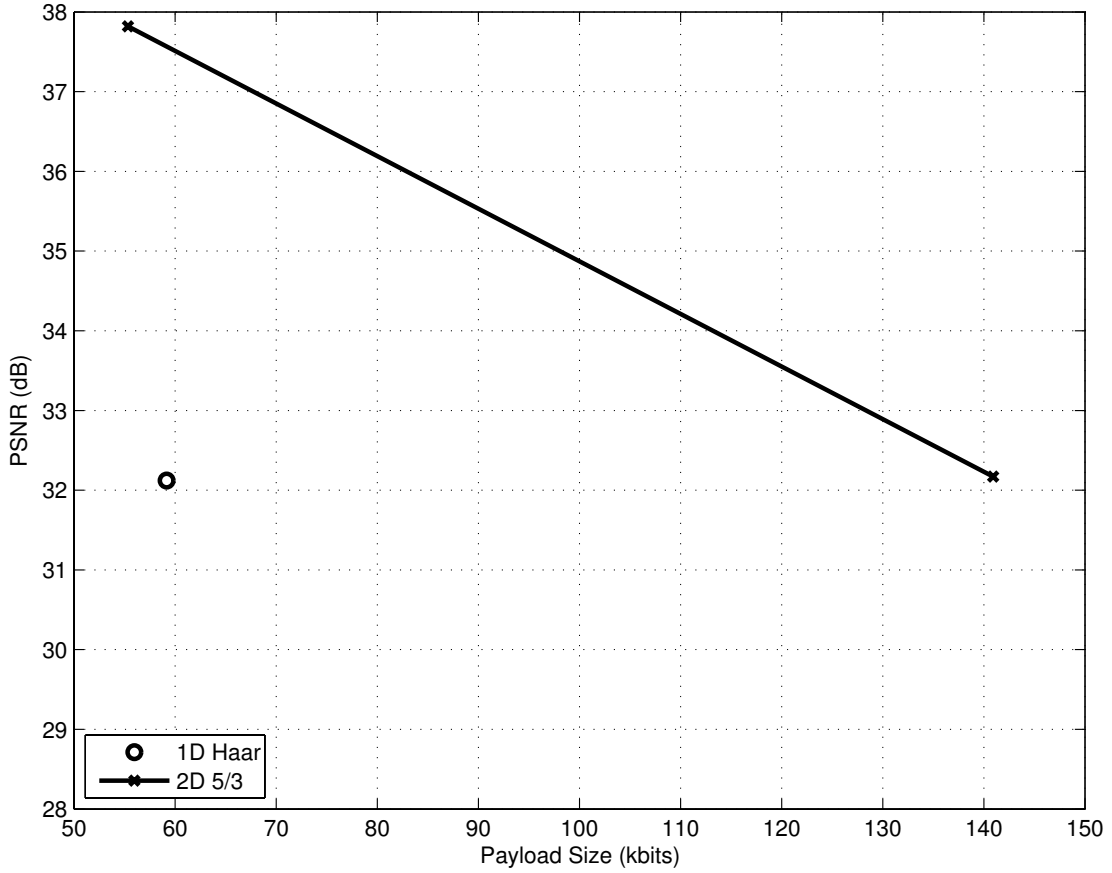


Figure 4.4: PSNR versus amount of data embedded (payload size) for “Lenna” using 1D Haar and 2D 5/3 DWT.

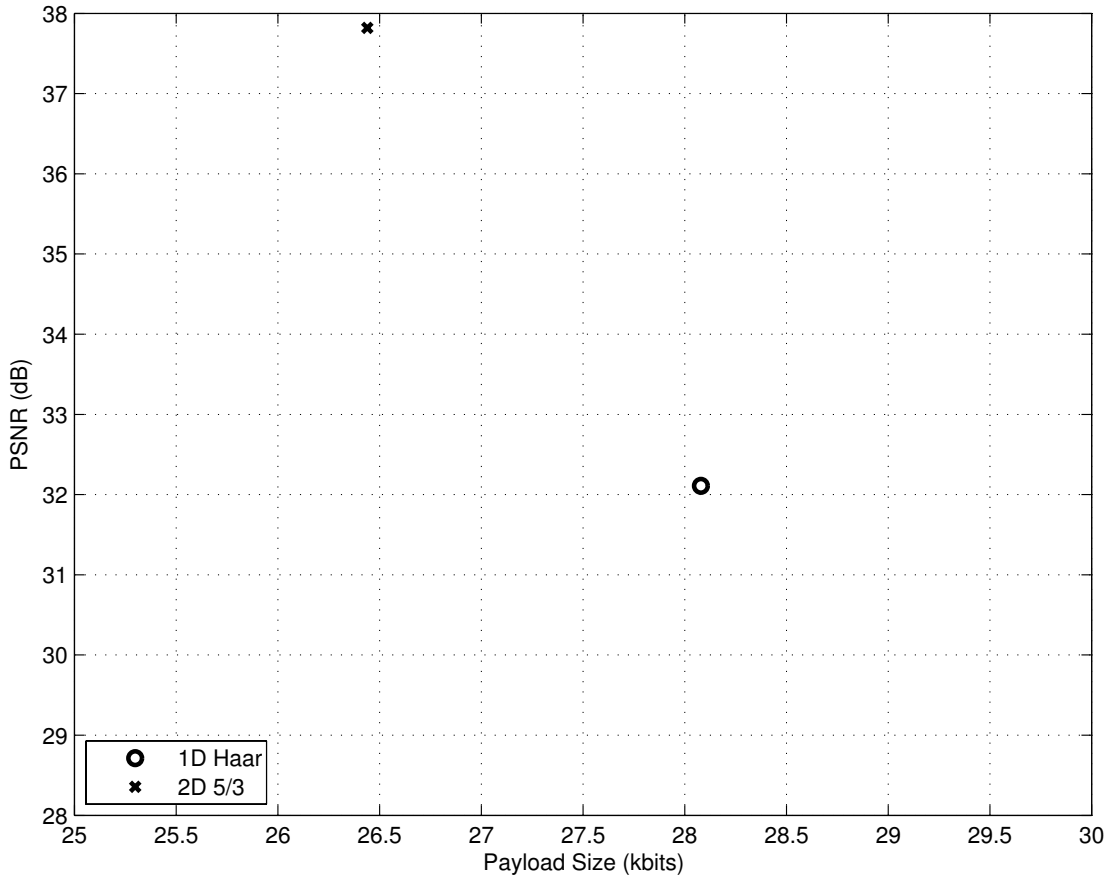


Figure 4.5: PSNR versus amount of data embedded (payload size) for “Goldhill” using 1D Haar and 2D 5/3 DWT.

Figure 4.6: “Barbara” after embedding with 2D 5/3 DWT yielded 37.82 dB PSNR with 46 kbits embedded.

CHAPTER V

CONCLUSIONS

Although watermarking as a concept has existed for hundreds, even thousands, of years, the introduction of the spread-spectrum technique [1] for robust watermarking resulted in a subsequent decade of signifcant interest in the development of watermarking and data-embedding schemes for digital media. This thesis has focused on the development of two algorithms, one from each of the two major classes of watermarking algorithms—robust watermarking and fragile watermarking.

As the fr st contribution of this thesis, we have adapted the robust pixel-wise masking (PWM) of [2] to the context of a redundant transform by modifying the approach to texture estimation which guides watermark casting and by accommodating the overcomplete nature of the transform in a Neyman-Pearson detection threshold. The proposed RDWT-domain texture measure more accurately estimates local texture activity since the equivalent DWT-based technique must consider increasingly larger spatial regions as resolution decreases due to the changing temporal sampling of the DWT. The resulting PWM-RDWT technique is shown to produce increased watermark robustness as compared to the original PWM-DWT approach in the face of a compression attack.

As the second contribution of this thesis, we have demonstrated improved performance for nearly invertible data embedding (NIDE) using a 2D, integer-valued 5/3 transform, known to be a more effcient image transform than the horizontal 1D Haar

transform that underlies a majority of prior approaches (e.g., [10, 23]) for invertible data embedding. In experimental results, the use of this proposed 2D transform provides a notable reduction in distortion as compared to the 1D transform for similar embedding rates. However, the NIDE scheme we propose cannot provide the simplicity of a 1D Haar transform with respect to invertibility, and therefore cannot predict or prevent overfo w/underfo w. Coupling our NIDE method with the 1D Haar method as a fall-back scheme for images with overfo w/underfo w provides better embedding when possible with a guaranteed invertibility on all images.

As digital media gains in prevalence, so too will the use of watermarking become increasingly more widespread. The techniques presented in this thesis are but two approaches in an ever-increasing diversity of watermarking strategies which are being adopted into an ever-increasing range of practical applications. The techniques put forth in this thesis fall into the two categories of robust watermarking and fragile watermarking, both of which have seen recent application in practice. For example, the recent Digital Cinema Standard [25] issues specifcations for robust watermarking to counter attempts to generate illicit copies of theatrical-release motion pictures. Such robust watermarking is intended to combat the most common form of piracy plaguing the motion-picture industry: the illicit taping of movies with a camcorder. Watermarks in this setting must be suffciently robust to survive the analog-to-digital conversion inherent to the taping process as well as any subsequent compression occurring in DVD mastering or online posting. On the other hand, fragile watermarking plays a key role in the authentication of digital data. As an example, an increasing number of states in the U.S. employ fragile watermarks in driver photographs, among other security features, to deter the counterfeiting of driver's licenses [26]. In this setting, the watermark must be fragile so that any modifcation of the image will destroy the watermark, resulting in

detection of any illicit tampering. These are just two examples of practical applications into which the techniques developed in this thesis may fnd eventual use.

REFERENCES

[1] I. J. Cox, J. Killian, F. T. Leighton, and T. Sharmoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, December 1997.

[2] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Transactions on Image Processing*, vol. 10, no. 5, pp. 783–791, May 2001.

[3] J. Fridrich, M. Goljan, and R. Du, "Invertible authentication," in *Security and Watermarking of Multimedia Contents III*, P. W. Wong and E. J. Delp III, Eds. San Jose, CA: Proc. SPIE 4314, January 2001, pp. 197–208.

[4] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 286–297, Proceedings of the International Conference, Marseille, France, December 14–18, 1987.

[5] P. Dutilleux, "An implementation of the "algorithme à trous" to compute the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.-M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 298–304, Proceedings of the International Conference, Marseille, France, December 14–18, 1987.

[6] M. J. Shensa, "The discrete wavelet transform: Wedding the à trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464–2482, October 1992.

[7] J.-G. Cao, J. E. Fowler, and N. H. Younan, "An image-adaptive watermark based on a redundant wavelet transform," in *Proceedings of the International Conference on Image Processing*, vol. 2, Thessaloniki, Greece, October 2001, pp. 277–280.

[8] L. Hua and J. E. Fowler, "A performance analysis of spread-spectrum watermarking based on redundant transforms," in *Proceedings of the IEEE*

International Conference on Multimedia and Expo, vol. 2, Lausanne, Switzerland, August 2002, pp. 553–556.

[9] K. M. Parker and J. E. Fowler, "Redundant-wavelet watermarking with pixel-wise masking," in *Proceedings of the International Conference on Image Processing*, vol. 1, Genoa, Italy, September 2005, pp. 685–688.

[10] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890–896, August 2003.

[11] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 244–250, April 1992.

[12] *Information Technology—-Coded Representation of Picture and Audio Information—Progressive Bi-Level Image Compression*, ISO/IEC 11544, 1993, JBIG Bi-Level Image Coding Standard.

[13] A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Lossless image compression using integer to integer wavelet transforms," in *Proceedings of the International Conference on Image Processing*, vol. 1, Santa Barbara, CA, October 1997, pp. 596–599.

[14] *Information Technology—JPEG 2000 Image Coding System—Part 1: Core Coding System*, ISO/IEC 15444-1, 2000.

[15] L. Kamstra and H. J. A. M. Heijmans, "Reversible data embedding into images using wavelet techniques and sorting," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2082–2090, December 2005.

[16] D. Zou, Y. Q. Shi, Z. Ni, and W. Su, "A semi-fragile lossless digital watermarking scheme based on integer wavelet transform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 10, pp. 1294–1300, October 2006.

[17] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1992.

[18] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 7, pp. 710–732, July 1992.

[19] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 629–632, September 2005.

[20] ——, "The redundant discrete wavelet transform and additive noise," Mississippi State ERC, Mississippi State University, Tech. Rep. MSSU-COE-ERC-04-04, March 2004.

[21] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, April 1992.

[22] A. Said and W. A. Pearlman, "A new, fast, and effcient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[23] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1147–1156, August 2004.

[24] *Information Technology—Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/s*, ISO/IEC 11172-2, 1993, MPEG-1 Video Coding Standard.

[25] *Digital Cinema System Specification*, Digital Cinema Initiatives, LLC, April 2007, version 1.1.

[26] J. Tuohey, "States begin security checks of driver's license photos," *PC World*, December 2004.