A DWT-BASED DIGITAL IMAGE WATERMARKING

Mir Delwar Hossain*, Jannatul Naim Tanima, Mohammad Za Uddin, Mir Feroz Uddin, Mohammed Shoaiibur Rahman

Abstract

Digital Watermarking is a technique which embeds a watermark signal into the host image in order to authenticate it. The operation of embedding and extraction of watermark are done in high frequency domain of Discrete Wavelet Transform (DWT) since small modifications in this domain are not perceived by human eyes. This watermarking scheme deals with the extraction of the watermark information in the absence of original image, hence the blind scheme was obtained. Peak Signal to Noise Ratio (PSNR) is computed to measure image quality. The experimental evaluation of the proposed method showed very good results in terms of robustness and transparency to various attacks such as median filtering, Gaussian noise, and JPEG compression.

Introduction

A “watermark” is a signal or secret information that is securely, imperceptibly, and robustly embedded into original content such as an image, video, or audio signal, producing a watermarked signal. Digital Watermarking is the process that embeds data called a watermark into a multimedia object such that watermark can be detected or extracted later to make an assertion about the object [Dugad et al, 1998]. Information hiding can be mainly divided into three processes - cryptography, steganography and watermarks. Cryptography is the process of converting information to an unintelligible form so that only the authorized person with the key can decipher it. As many advances were made in the field of communication it became rather simple to decrypt a ciphertext. Hence more sophisticated methods were designed to offer better security than what cryptography could offer. This led to the discovery of stenography and watermarking. Stenography is the process of hiding information over a cover object such that the hidden information cannot be perceived by the user. Thus, even the existence of secret information is not known to the attacker. Watermarking is closely related to stenography, but in watermarking the hidden information is usually related to the cover object. Hence it is mainly used for copyright protection and owner authentication.

A watermarking system is usually divided into three distinct steps, embedding, attack and detection. In embedding, an algorithm accepts the host and the data to be embedded and produces a watermarked signal. The watermarked signal is then transmitted or stored, usually transmitted to another person. If this person makes a modification, this is called an attack. There are many possible attacks. Detection is an algorithm which is applied to the attacked signal to attempt to extract the watermark from it. If the signal was not modified during transmission, then the watermark is still present and it can be extracted. If the signal is copied, then the information is also carried in the copy. The

*Corresponding Authors’ Email: za_maruf@yahoo.com
embedding takes place by manipulating the content of the digital data, which means the information is not embedded in the frame around the data, it is carried with the signal itself.

Fig. 1 Watermarking block diagram

The original image and the desired watermark are embedded using one of the various schemes that are currently available. The obtained watermarked image is passed through a decoder in which usually a reverse process to that employed during the embedding stage is applied to retrieve the watermark. The different techniques differ in the way in which it embeds the watermark on to the cover object. A secret key is used during the embedding and the extraction process in order to prevent illegal access to the watermark [Abrar].

Requirements

A number of watermarking system requirements as well as the tradeoffs among them has been described in this section.

- Security:
- Imperceptibility:
- Capacity:
- Robustness:

Applications of Watermarking

Basically watermarking algorithms provide a data transmission channel that can be used in existing distribution channels. This data transmission is compatible in the sense that every existing channel that is able to carry image also is able to carry watermarked image. Hence watermarking can be utilized in a wide field of applications [Wikipedia][Christian etal].

Applications include

- Authentication and verification
- Fingerprinting
- Ownership Assertion
- Content labeling
- Usage Control

Techniques of Watermarking

The basic components of any watermarking technique consist of a marking algorithm that inserts information, the watermark, into an image. The watermark is inserted into the image in the spatial domain or spatial frequency domain. As part of the watermarking
Conflict over extraction method

technique, a testing algorithm must be defined that tests an image to see if a particular watermark is contained in the image.

- Spatial Domain Techniques
- Frequency Domain Techniques

Attacking on Watermarked Image

General processes involved in a watermarking system. The watermark is encoded into the cover data in the embedding phase. An optional encryption mechanism may also be used to add another layer of security. In the transmission phase, the watermarked image can be subject to attack from third party. This in turns provides some challenges to the decoder to retrieve as accurate as possible the hidden data from the received watermarked image. Digital watermarking is not as secure as date encryption. Therefore, digital watermarking is not immune to hacker attacks. An image can be attacked by the following way [Chunlin et al].

![Fig. 2 Attacks on Watermarked Image](image-url)

Watermarking attacks are broadly divided into the following categories:

- Removal Attacks
- Geometrical Attacks
- Cryptographic Attack
- Protocol Attacks

![Fig. 3 Types of watermarking attacks.](image-url)

Discrete Wavelet Transform (DWT) Based Digital Image Watermarking

In this project using DWT, adding information in a digital image is decoded but without losing its original structure. The processes are described below.

Overview of Wavelet Transformation

When the time localization of the spectral components is needed, a transform giving the time-frequency representation of the signal is needed. The ultimate solution is the wavelet transform that provides the time-frequency representation. The foundations of the DWT go back to 1976 when Croisier, Esteban, and Galand devised a technique to decompose discrete time signals. Crochier, Weber, and Flanagan made a similar work on coding of
speech signals in the same year. They named their analysis scheme as sub band coding. In 1983, Burt defined a technique very similar to sub band coding and called it pyramidal coding which is also known as multi resolution analysis. Later in 1989, Vetterli and Le Gall did some improvements to the sub band coding scheme, removing the existing redundancy in the pyramidal coding scheme.

In the Discrete Wavelet Transform, filters of different cutoff frequencies are used to analyze the signal at different scales. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies.

The resolution of the signal, which is a measure of the amount of detail information in the signal, is changed by the filtering operations, and the scale is changed by up sampling and down sampling (sub sampling) operations. Sub sampling a signal corresponds to reducing the sampling rate, or removing some of the samples of the signal. For example, sub sampling by two refers to dropping every other sample of the signal. Sub sampling by a factor n reduces the number of samples in the signal n times. Up sampling a signal corresponds to increasing the sampling rate of a signal by adding new samples to the signal. For example, up sampling by two refers to adding a new sample, usually a zero or an interpolated value, between every two samples of the signal. Up sampling a signal by a factor of n increases the number of samples in the signal by a factor of n.

The basic idea in the DWT for a one dimensional signal is the following. A signal is split into two parts, usually high frequencies and low frequencies. The edge components of the signal are largely confined to the high frequency part. The low frequency part is split again into two parts of high and low frequencies. This process is continued an arbitrary number of times, which is usually determined by the application at hand. Furthermore, from these DWT coefficients, the original signal can be reconstructed. This reconstruction process is called the inverse DWT (IDWT) [Dugad et al, 1998].

The DWT and IDWT can be mathematically stated as follows. Let

\[ H(w) = \sum_k h_k e^{-jkw} \quad \text{and} \quad G(w) = \sum_k g_k e^{-jkw} \] ..........................(1)

be a lowpass and a highpass filter, respectively, which satisfy a certain condition for reconstruction to be stated later. A signal, \( F[n] \) can be decomposed recursively as

\[ f_{J+1}^{\text{low}}(k) = \sum_n h_{n-2k} f_j(n) \] ..........................(2)

\[ f_{J+1}^{\text{high}}(k) = \sum_n g_{n-2k} f_j(n) \] ..........................(3)

for \( j = J+1, J, \ldots, J_0 \) where \( f_{J+1}(k) = F(f), k \in Z \).

\( J + 1 \) is the highest resolution level index and \( J_0 \) is the lowest resolution level index.

The coefficients

\( f_{J_0}^{\text{low}}(k), f_{J_0}^{\text{high}}(k), f_{J_1}^{\text{high}}(k), \ldots, f_J^{\text{high}}(k) \)

are called the DWT of the signal \( F[n] \), where \( f_{J_0}^{\text{low}}(k) \) is the lowest resolution part of \( F[n] \) (the approximation) and the \( f_j^{\text{high}}(k) \) are the details of \( F[n] \) at various bands of frequencies. Furthermore, the signal \( F[n] \) can be reconstructed from its DWT coefficients recursively,

\[ f_j^{\text{low}}(n) = \sum_k h_{n-2k} \cdot f_{j-1}^{\text{low}}(k) + \sum_k g_{n-2k} \cdot f_{j-1}^{\text{high}}(k) \] ..........................(4)
To ensure the above IDWT and DWT relationship, the following orthogonality condition on the filters $H(w)$ and $G(w)$ is needed:

$$|H(w)|^2 + |G(w)|^2 = 1$$  \hspace{2cm} (5)

An example of such $H(w)$ and $G(w)$ is given by,

$$H(w) = \frac{1}{2} + \frac{1}{2} e^{-jw}$$  \hspace{2cm} (6)

and

$$G(w) = \frac{1}{2} - \frac{1}{2} e^{-jw}$$  \hspace{2cm} (7)

Which is known as the Haar wavelet Filter. Other common filters used in image processing are the family of Daubechies orthogonal (D-4, D-6, D-8, D-10, etc.), Symlets, Coiflets, Morlet, and also biorthogonal filters. The DWT and IDWT for a two dimensional image $F(m,n)$ can be similarly defined by implementing the one dimensional DWT and IDWT for each dimension $m$ and $n$ separately, resulting in the pyramidal representation of an image [Dugad et al, 1998].

The 2D-DWT transform divides the image into 4 sub-bands:
- **LL** (approx.) - Lower resolution version of image
- **LH** - Horizontal edge data
- **HL** - Vertical edge data
- **HH** - Diagonal edge data

Most DWT watermarking algorithms embed only in the HL, LH and HH sub-bands.

**Implementation of Watermarked Image**

A DWT-based semi-blind image watermarking scheme is implemented by following a similar approach of Dugad et al. Excluded low pass bands and embedded the watermark into high pass bands. Since high pass bands typically contain edge related information of the image, each coefficient in the high frequency bands affects only a spatially limited portion of the image. Therefore, adding the watermark to significant coefficients in the high frequency bands is equivalent to embedding the watermark to only the edge areas of the image. In Dugad’s work, threshold values of $T_1$ was 40.0 and $T_2$ was 50.0 and were obtained after three levels decomposition. We have also measured...
the quality of the watermarked image by calculating PSNR (Peak Signal-to-Noise Ratio) value which is larger and maintain better image quality [Dugad et al, 1998].

The equation used for watermark embedding is similar to those used in:

\[ V'_i = V_i + \alpha |V'_i| x_i \] ...........................................(8)

Where \( i \) run over all DFT coefficients \( > T_1 \)

\( V'_i \) refer to the corresponding DWT coefficient of the original image and \( V'_i \) refer to the DWT coefficient of the watermarked image. \( x_i \) is the watermark value at the position of \( V'_i. x_i \) is generated from a normal distribution of zero mean and unit variance. \( \alpha \) is taken as 0.2.

The correlation \( z \) between the DWT coefficients \( \hat{V} \) of the corrupted watermarked image and a possibly different watermark \( Y \) is computed as

\[ z = \frac{1}{M} \sum_i \hat{V}_i y_i \] ...........................................(9)

Where \( i \) run over all DFT coefficients \( > T_2 > T_1 \) and \( M \) is the number of such coefficients.

The threshold \( S \) is defined as

\[ S = \frac{\alpha}{2M} \sum_i \hat{V}_i \] ...........................................(10)

If \( z \) exceeds \( S \), the conclusion is the watermark is present.

The PSNR is defined as:

\[ PSNR = 10 \log_{10} \frac{255^2}{MSE} (db) \] ...........................................(11)

where mean-square error (MSE) is defined as:

\[ MSE = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (V'_{i,j} - V_{i,j})^2 \] ...........................................(12)

Where \( V_{i,j} \) and \( V'_{i,j} \) are the gray level of pixels in the host image and watermarked image respectively [Dugad et al, 1998].
Conflict over extraction method

Experimental Result
The Lena image is used in figure 4.3 (a) to test this implementation work. A Standard Normal Distributed pseudo-random number sequence \( \{X = x_1, x_2, \ldots, x_N\} \) where \( N \) is the length of the watermark is used as watermark. This watermark is embedded into the Lena image using equation (8) to produce watermarked image.

The original Lena image, the watermarked image, and their difference are shown in Figure-4.3. We see that the watermarked image is not distinguishable visually from the original image. The PSNR of figure 4.3(b) is 41.0992, which tells us that the watermarked image generated by this algorithm is imperceptible to human eyes. The difference image (figure-4.3(c)) basically represent embedded watermark in the spatial domain. Figure-4.3(c) shows that most of the watermark is added in edge regions of the image as stated before. For this, there is no need for any explicit visual masking, since each coefficient in the high frequency bands has an effect only on a spatially limited portion of the image. As a result, embedding the watermark to significant coefficients in the high frequency bands is equivalent to adding the watermark to only the edge areas of the image, which makes the watermark invisible to the human visual system.
Fig. 4.2 (a) Original Lena image, (b) Watermarked image (PSNR=41.0992)

The watermark is detected from the watermarked image without original image using Eq. 9 & Eq. 10. The following figure 4.4 depicts that several watermarks are randomly generated in the x-axis and corresponding correlation values (i.e. detector responses) between the watermarked image and randomly generated watermark are shown in the y-axis. The correlation with real watermark is located at 100 in the x-axis, since the original image has been watermarked with a seed of 100. That only the real watermark is located at 100 and fake watermarks are located at other points. Therefore, the correlation value is larger at 100 than the other fake watermark. So, it represents the existence of the watermark. The dashed line represents the threshold (Eq. 10). If the correlation value exceeds the threshold, then we conclude that the watermark is present.

Figure 4.4 Detector response for unattacked watermarked image.
Conflict over extraction method

Several common signal processing attacks such as Gaussian noise, salt & pepper noise, low pass filtering having window size 3x3, median filtering having window size 5x5, histogram equalization, JPEG compression, center cropping are applied to the watermarked image to evaluate robustness of this algorithm. The following figure 4.5 shows attacked watermarked image. The figure 4.5(a) shows watermarked image attacked by Gaussian noise with PSNR=21.9015, figure 4.5(b) shows watermarked image attacked by Salt & Pepper noise with PSNR=27.4014, figure 4.5(c) shows watermarked image attacked by low pass filtering with PSNR=38.5051, figure 4.5(d) shows watermarked image attacked by median filtering with PSNR=31.6133, figure 4.5(e) shows watermarked image attacked by histogram equalization with PSNR=18.0789, figure 4.5(f) shows watermarked image attacked by JPEG compression with 5% quality, PSNR value is 27.3343, figure 4.5(g) shows watermarked image attacked by center cropping with PSNR=6.5238.

(a) PSNR=21.9015  
(b) PSNR=27.4014  
(c) PSNR=38.5051  
(d) PSNR=31.6133  
(e) PSNR=18.0789  
(f) PSNR=27.3343  
(g) PSNR=6.5238

Figure 4.5: (a) attacked by Gaussian noise (mean=0, variance=0.006), (b) attacked by Salt & Pepper noise (mean=0, variance=0.006), (c) attacked by Low pass filtering (Window size =3x3), (d) attacked by median filtering (Window size =5x5), (e) attacked by Histogram Equalization, (f) attacked by JPEG compression (Quality=5), (g) attacked by cropping.
In following figures (from Fig.-4.6 to Fig.-4.12) we display the detector responses (correlation values (Eq.-9) on the y-axis and randomly generated watermark on the x-axis for attacked watermarked image. The original image has been watermarked with a seed of 100. Therefore, in each figure, the correlation with the real watermark is located at 100 on the x-axis, and the dashed line shows the value of the threshold.

Similarly for the attacked watermarked image, the correlation value of watermarked image and real watermark decreases somewhat, but the value is also larger than other. Therefore it represents the existence of our embedded watermark. The following figure-4.6 shows detector response after Gaussian noise addition, figure-4.7 shows detector response after salt & pepper noise addition, figure-4.8 shows detector response after low pass filtering, figure-4.9 shows detector response after median filtering, figure-4.10 shows detector response after histogram equalization, figure-4.11 shows detector response after JPEG compression; we compressed the image with 5% quality JPEG compression, figure-4.12 shows detector response after center cropping.
Making use of the principal advantages of Discrete Wavelet Transformation, this chapter implemented an image watermarking technique. The efficiency of this technique has been shown by a series of our experiment. From the above result, we see that this algorithm also works on geometric attack, such as center cropping.

Conclusions
This project introduces a discrete wavelet transform (DWT) digital watermark algorithm based on human vision characters. By using this technique watermarking signal is embedded into the high frequency band of wavelet transformation domain. With the help from Dugad’s paper the simulation results suggest that this watermarking system not only can keep the image quality well, but also can be robust against many common image processing operations of filter, sharp enhancing, adding salt noise, image compression and so on. This algorithm has strong capability of embedding signal and anti-attack. In future, we will combine DCT and DWT in digital image watermarking for increasing robustness.

References
Abhishek Goswami, “Introduction To Digital Watermarks And Classification Of Attacks”, Graduate Student, Department of Electrical and Computer Engineering, Stony Brook.


Chunlin Song, Sud Sudiman, Madjid Merabti, “Recent Advances and Classification of Watermarking Techniques in Digital Images” School of Computing and Mathematical Sciences, Liverpool John Moores University, UK.


Sviatolsav Voloshynovskiy, Shelby Pereira, and Thiery Pun, University of Geneva Joachim J. Eggers and Jonathan K. Su, University of Erlangen-Nuremberg, “Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks”.