DIGITAL WATERMARKING USING MULTIRESOLUTION WAVELET DECOMPOSITION

Deepa Kundur and Dimitrios Hatzinakos

10 King's College Road
Department of Electrical and Computer Engineering
University of Toronto,
Toronto, Ontario, Canada M5S 3G4
E-mail: deepa@comm.toronto.edu, dimitris@comm.toronto.edu

ABSTRACT

We present a novel technique for the digital watermarking of still images based on the concept of multiresolution wavelet fusion. The algorithm is robust to a variety of signal distortions. The original unmarked image is not required for watermark extraction. We provide analysis to describe the behaviour of the method for varying system parameter values. We compare our approach with another transform domain watermarking method. Simulation results show the superior performance of the technique and demonstrate its potential for the robust watermarking of photographic imagery.

1. INTRODUCTION

In recent years there has been growing interest in developing effective techniques to discourage the unauthorized duplication of digital data. The technology designed to make electronic publishing feasible has also increased the threat of intellectual property theft. One approach to address the problem is called *digital watermarking*. Digital watermarking is the imperceptible marking of multimedia data to "brand" ownership.

The process of digital watermarking involves the modification of the original multimedia data to embed a *watermark* containing key information such as authentication or copyright codes. The embedding method must leave the original data perceptually unchanged, yet should impose modifications which can be detected by using an appropriate extraction algorithm. Common types of signals to watermark are images, music clips and digital video. In this paper we concentrate on the application of digital watermarking to still images. The major technical challenge is to design a highly robust digital watermarking technique, which discourages copyright infringement by making the process of watermarking removal tedious and costly [7].

Current techniques described in the literature for the watermarking of images can be grouped into two classes: transform domain methods [2, 5] which embed the data by modulating the transform domain coefficients and spatial domain techniques [1, 6] which embed the data by directly modifying the pixel values of the original image. We propose a transform domain technique which shows greater robustness to common signal distortions. The fundamental advantage of our wavelet-based technique lies in the method used to embed the watermark in each of the resolution levels. The approach provides the simultaneous spatial localization and frequency spread of the watermark within the host image to provide robustness against widely varying signal distortions such as cropping and filtering. This work is a more practical extension of our previous research on watermarking using multiresolution wavelet data fusion [3, 4].

In the next section we introduce the proposed method for digital watermarking. In Section 3, we provide analysis to estimate the probabilities of false positives and false negatives for watermark detection as a function of the algorithm parameters. Simulation results are given in Section 4 in which a comparison is made with another multiresolution wavelet-based technique. We provide concluding remarks and a discussion in Section 5.

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada and by Communications and Information Technology Ontario (CITO).

2. PROPOSED WATERMARKING TECHNIQUE

In this section we discuss some motivating factors in the design of our approach to watermarking. Research into human perception indicates that the retina of the eye splits an image into several frequency channels each spanning a bandwidth of approximately one octave. The signals in these channels are processed independently. Similarly, in a multiresolution decomposition, the image is separated into bands of approximately equal bandwidth on a logarithmic scale. It is therefore expected that use of the discrete wavelet transform will allow the independent processing of the resulting components without significant perceptible interaction between them, and hence make the process of imperceptible marking more effective. For this reason, the wavelet decomposition is commonly used for the *fusion* of images. Fusion is a sensor-data-compressed information problem in which several signals are merged into one. Since digital watermarking involves the merging of a watermark with a host signal it follows that wavelets are attractive for the watermarking of images. The localization of the watermark at high resolutions provides the ability to identify distinct regions of the watermarked image which have undergone tampering, and the global spreading of the watermark at low resolutions within the host makes it robust to large-scale signal distortions. In addition, the technique is "unsupervised" since the host image is not required for watermark extraction. The combined result of these factors makes the proposed method promising for practical use.

2.1. The Architecture

We assume for simplicity that the binary watermark is of length N_w and consists of elements from the set $\{-1, 1\}$. We refer to the

0-7803-4428-6/98 $10.00 © 1998 IEEE

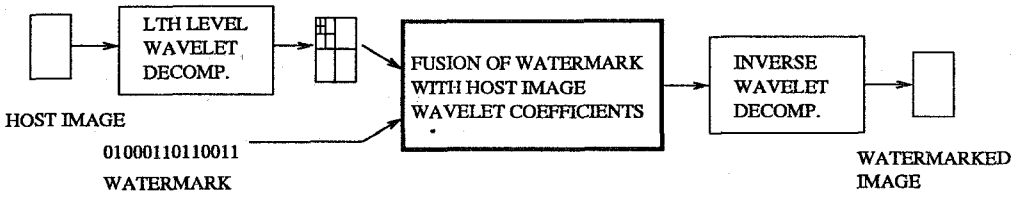

Figure 1: The proposed watermarking method.

original image f in which we embed the watermark as the *host image*. We embed the watermark into the detail wavelet coefficients of the host image with the use of a key. This key is randomly generated and is used to select the exact locations in the wavelet domain in which to embed the watermark. For each coefficient within the wavelet domain, the key has a corresponding value of one or zero to indicate if the coefficient is to be marked or not, respectively. The number of ones in the key must be greater or equal to the size of the watermark. The watermark values are repeatedly embedded in different coefficients selected by the key if the length of the watermark is less than the number of ones in the key.

The technique is comprised of the three stages described below:

Stage I: Compute the Lth-level discrete wavelet decomposition of the host image to produce a sequence of $3L$ *detail* images, corresponding to the horizontal, vertical and diagonal details at each of the L resolution levels, and a gross approximation of the image at the coarsest resolution level. We denote the kth detail image component at the lth resolution level of the host by $f_{k,l}(m,n)$ where $k = h, v, d$ (which stands for "horizontal", "vertical" and "diagonal", detail coefficients, respectively) and $l = 1, \ldots, L$. The gross approximation is represented by $f_{a,L}(m,n)$.

Stage II: Consider each resolution level l, and coefficient location (m,n). If the associated value of the key κ is one then proceed as follows. Otherwise do not embed a mark.

1. Sort the detail coefficients in ascending order so that $f_{k1,l}(m,n)$, $f_{k2,l}(m,n)$, and $f_{k3,l}(m,n)$ are coefficients such that

$$f_{k1,l}(m,n) \leq f_{k2,l}(m,n) \leq f_{k3,l}(m,n), \qquad (1)$$

   where $k1, k2, k3 \in \{h, v, d\}$ and $k1, k2, k3$ are distinct.

2. To embed the watermark, quantize $f_{k2,l}(m,n)$ as shown in Figure 2. The range of values between $f_{k1,l}(m,n)$ and $f_{k3,l}(m,n)$ is divided into bins of width

$$\Delta = \frac{f_{k3,l}(m,n) - f_{k1,l}(m,n)}{2Q-1}, \qquad (2)$$

   where Q is a user-defined variable. To embed a watermark bit of value one, $f_{k2,l}(m,n)$ is quantized to the nearest value shown with bold vertical bars in Figure 2. Alternatively, to embed a negative one, $f_{k2,l}(m,n)$ is quantized to the nearest value shown by dashed vertical lines.

Stage III: The corresponding Lth level inverse wavelet transform the fused image components is computed to form the watermarked image.

The method is shown in Figure 1. The parameter Q is user-defined and is set to establish an appropriate trade-off between the visibility and robustness of the watermark. A larger value of Q will make the quantization process of Stage II finer which reduces the possibility of visual degradation, but which can make the watermark extraction less accurate in the face of attack.

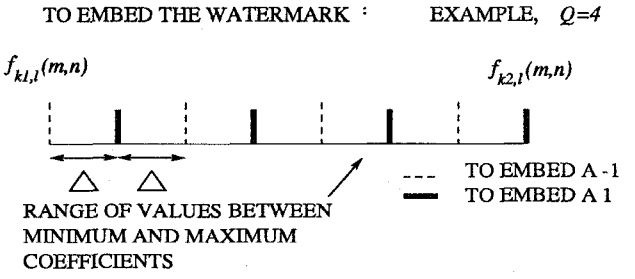

Figure 2: Quantization process to embed the binary watermark. The middle wavelet coefficient $f_{k2,l}(m,n)$ must be quantized to the nearest vertical bold bar line to embed a one and to the nearest dashed line to embed a negative one.

An attacker cannot easily determine the exact key given a watermarked image if the specific wavelet transform used in the decomposition of Step I is kept a secret and Q is unknown. Therefore, it is not be possible to use the relative value of the coefficients to determine the watermark locations and hence destroy th mark by randomly changing the coefficient values by small amounts. To destroy the watermark for certain, all of the coefficient values may have to be randomly changed which can cause visible degradation in the image and make the image useless to the attacker.

2.2. Watermark Extraction and Detection

The objective of the watermark extraction process is to reliably obtain an estimate of the original watermark from a possibly distorted version of the watermarked image. The detection process requires knowledge of the watermark $w(m,n)$ and the key $\kappa(m,n)$. We represent the image for which we wish to apply the extraction process by $r(m,n)$.

The first step involves applying an Lth level discrete wavelet decomposition on the image $r(m,n)$. We let $r_{k,l}(m,n)$ be the kth detail image component of the lth resolution level of $r(m,n)$.

We then make use of the key κ to find the locations in which the watermark was embedded for each resolution level l. We extract the watermark from these coefficients as follows:

1. Order the detail coefficients in ascending order as before so that

$$r_{k1,l}(m,n) \leq r_{k2,l}(m,n) \leq r_{k3,l}(m,n), \qquad (3)$$

   where $k1, k2, k3 \in \{h, v, d\}$ and $k1, k2, k3$ are distinct.

2. Estimate the watermark bit value from the relative position of $r_{k2,l}(m,n)$. Using the same constant Q as for embedding, a particular watermark bit is determined by finding the closest quantized value (shown in Figure 2) to $r_{k2,l}(m,n)$ and determining if this quantized value was used to embed a one or a negative one.
3. If the watermark had been embedded in different locations several times, then the most common bit value extracted is assigned for the estimated watermark. If an equal number of ones and negative ones were extracted, then a random guess is made to its value.

A given watermark is detected if the correlation of the extracted watermark with the given watermark is above a pre-specified threshold. More precisely, the watermark detection condition is given by

$$\rho(w,\tilde{w}) = \frac{\sum w(n)\tilde{w}(n)}{\sqrt{\sum w^2(n)}\sqrt{\sum \tilde{w}^2(n)}} \geq T, \quad (4)$$

where w is the given watermark, $\tilde{w}$ is the extracted one, and T is a pre-specified threshold. The quantity $\rho(w,\tilde{w})$ is know as the *correlation coefficient* between the given and extracted watermarks. Unless otherwise stated, all summations of Equation 4 and in the next section have index n and range from 1 to N_w.

3. ANALYSIS

We provide analysis to estimate the probability of a false positive (i.e., false watermark detection) and the probability of a false negative (i.e., failure to detect an existing watermark) for our proposed technique. We define the probability of false watermark detection as

$$P_{fp} = P\{\rho(w,\tilde{w}) \geq T | \text{no mark}\}, \quad (5)$$

where $P\{A|B\}$ is the probability of event A given event B. We can rewrite $\rho(w,\tilde{w})$ as

$$\rho(w,\tilde{w}) = \frac{\sum w(n)\tilde{w}(n)}{\sqrt{\sum w^2(n)}\sqrt{\sum \tilde{w}^2(n)}} = \frac{\sum w(n)\tilde{w}(n)}{N_w}, \quad (6)$$

since $w(n)$ and $\tilde{w}(n)$ are either one or negative one, and subsequently $w^2(n) = \tilde{w}^2(n) = 1$.

Let p_E be the probability of bit error during extraction. A bit error occurs when $\tilde{w}(n) \neq w(n)$ or more specifically, when $\tilde{w}(n) = -w(n)$ (since $w(n), \tilde{w}(n) \in \{-1, 1\}$). If we let $k(n) = w(n)\tilde{w}(n)$, then $k(n) = -1$ indicates a bit error and $k(n) = 1$ indicates no error. We may rewrite our expressions for ρ and P_{fp} in terms of $k(n)$ as

$$\rho(w,\tilde{w}) = \frac{\sum w(n)\tilde{w}(n)}{N_w} = \frac{\sum k(n)}{N_w}, \quad (7)$$

and

$$P_{fp} = P\{\sum k(n) \geq N_w T | \text{no mark}\}, \quad (8)$$

respectively. Since $k(n) \in \{-1, 1\}$, it can be shown that $\sum k(n)$ must take on discrete values from the set $\{-N_w, -N_w+2, -N_w+4, \ldots, N_w-4, N_w-2, N_w\}$, or $\sum k(n) = -N_w + 2m$, where $m = 0, 1, \ldots, N_w$. Thus, we find that

$$\begin{aligned} P_{fp} &= P\{\sum k(n) \geq N_w T | \text{no mark}\} \\ &= \sum_{m=\lceil N_w(T+1)/2 \rceil}^{N_w} P\{\sum k(n) = -N_w + 2m | \text{no mark}\}, \end{aligned} \quad (9)$$

where $P\{\sum k(n) = -N_w + 2m | \text{no mark}\}$ is the probability that the series $\{k(n)\}$ contains m ones and $N_w - m$ negative ones. Therefore,

$$P\{\sum k(n) = -N_w + 2m |_{\text{mark}}^{\text{no}}\} = \binom{N_w}{m} p_E^{N_w - m}(1-p_E)^m \quad (10)$$

where p_E is the probability that $k(n) = -1$ and $\binom{N_w}{m} = \frac{N_w!}{m!(N_w-m)!}$. Since we are given that no watermark is embedded, we can assume that the extracted mark $\tilde{w}$ consists of a series of random independent equally probable values from the set $\{-1, 1\}$. Thus, $p_E = 0.5$. Substituting into Equations (9) and (10),

$$P_{fp} = \sum_{m=\lceil N_w(T+1)/2 \rceil}^{N_w} \binom{N_w}{m} 0.5^{N_w}. \quad (11)$$

Similar analysis can be performed to compute the probability of false negative, which is defined as

$$P_{fn} = P\{\rho(w,\tilde{w}) < T | \text{watermark } w \text{ is embedded}\}. \quad (12)$$

If we can model the effects of image distortion on the extracted watermark as additive white Gaussian noise with variance σ^2, then we can approximate the probability of false negative as

$$\begin{aligned} P_{fn} \approx & \sum_{m=0}^{\lceil N_w(T+1)/2 \rceil - 1} \binom{N_w}{m} \left[\frac{2Q-1}{Q}\text{erfc}\left(\frac{\overline{\Delta}}{4\sigma}\right)\right]^{N_w - m} \cdot \\ & \left[1 - \frac{2Q-1}{Q}\text{erfc}\left(\frac{\overline{\Delta}}{4\sigma}\right)\right]^m, \end{aligned} \quad (13)$$

where $\text{erfc}(\cdot)$ is the standard complementary error function and $\overline{\Delta}$ is the average value of Δ over all the wavelet coefficients for a given image.

Given a desired probability of false alarm, we can set the threshold T using Equation (11). As the length of the watermark increases, the probability of false detection decreases for a fixed threshold. Similarly, increasing N_w will reduce the probability of false negative P_{fn}. It should be noted that there is a trade-off with respect to the choice of T. An increase in T will reduce P_{fp}, but will increase P_{fn}, and vice versa for a reduction in T. The choice of the threshold should be a function of N_w and must be application-dependent.

4. SIMULATION RESULTS

We demonstrate the robustness of the proposed watermarking algorithm by investigating the effects of image distortion on the correlation between the original and extracted watermark. We compare the performance of the algorithm with another wavelet-based technique by Ohnishi and Matsui [5]. The algorithm, which is similar to the proposed technique, was developed independently and embeds the watermark in different resolutions of the host image. The Haar wavelet is used to produce the coefficients. The most significant difference between the two methods lies in the merging stage of the watermark. In [5] the authors mark the host by forcing the modulo 2 difference between the largest and smallest wavelet coefficients for a particular position and resolution level to be one if $w(n) = 1$ and to be zero if $w(n) = -1$.

The image "barb" was watermarked with a 256 length randomly generated binary watermark using our algorithm with $Q = 4$ and the Daubechies 10-pt wavelet. For comparison the host image was also watermarked with the same 256 length binary watermark using the method in [5]. Neither process created visible artifacts or changes in the marked result. Both watermarked images were distorted in turn by mean filtering, median filtering, JPEG compression, additive white Gaussian noise, cropping and scaling. The results of $\rho(w, \tilde{w})$ upon watermark extraction for mean filtering as a function of the $M \times M$ filter size, JPEG compression as a function of compression ratio and additive noise as a function of signal-to-noise ratio (SNR) are shown in Figure 3.

To investigate the effect of embedding the watermark within the different resolution levels, two methods of extraction were employed. In the first instance, the watermark was extracted from all resolution levels and the most common bit value was assigned to the extracted watermark. In the second (novel) approach, the watermark was extracted from the coarsest resolution level only. The results for extraction from all resolution levels is shown using the solid lines in Figure 3 and the results of extraction from the lowest resolution level are presented with dashed lines. The plots with the "+" symbols are the results for the proposed technique and the remaining plots show the performance of the technique in [5]. In all cases, the correlation coefficient of the proposed technique is much higher than that for the method by Ohnishi and Matsui which suggests that the proposed algorithm is more robust to common signal distortions. The most dramatic difference between the two results occurred for distortion by additive white Gaussian noise. The simulations indicate that for specific types of distortion, extraction from the lowest resolution level gives superior results. Future work will involve the design of an appropriate receiver structure which weights the extracted watermark from each resolution level appropriately to make extraction more accurate.

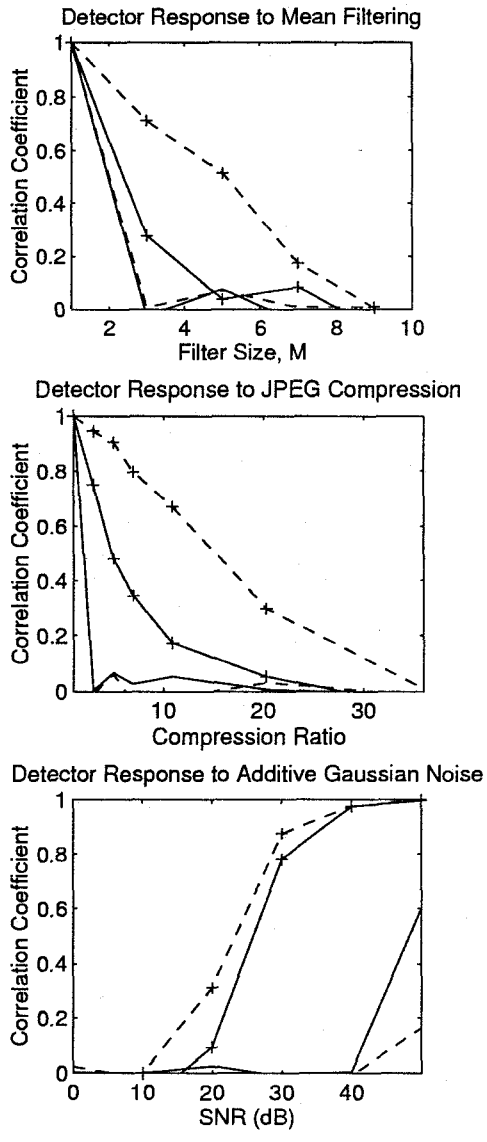

Figure 3: Results for linear mean filtering, JPEG compression and additive noise. The plots with the '+' symbols are the correlation results for the proposed method; the remaining results are for the method in [5]. The solid and dashed lines correspond to the extraction from all resolution levels and the lowest resolution for each of the methods, respectively.

5. CONCLUSIONS

We proposed a robust multiresolution wavelet-based digital watermarking method which shows superior performance to an existing technique of its class. Analysis is provided to compute the probability of false positive and false negative results; the expressions suggest that increasing the length of the watermark can reduce the probability of detection error. Simulation results demonstrate the robustness of the approach to common signal distortions.

Future work concentrates on coding strategies to reduce the extracted watermark bit error and modification of the embedding technique to combat the multiple watermarking problem.

6. REFERENCES

[1] A. G. Bors and I. Pitas, "Image watermarking using DCT domain constraints," in *Proc. IEEE Int. Conference on Image Processing*, vol. 3, pp. 231–234, 1996.

[2] I. J. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," Tech. Rep. 95–10, NEC Research Institute, 1995.

[3] D. Kundur and D. Hatzinakos, "A robust digital image watermarking method using wavelet-based fusion," *to appear in Proc. Int. Conference in Image Processing*, 1997.

[4] D. Kundur and D. Hatzinakos, "Digital Watermarking Based on Multiresolution Wavelet Data Fusion," *Proc. IEEE, Special Issue on Intelligent Signal Processing*, under review, 43 pages, 1997.

[5] J. Ohnishi and K. Matsui, "Embedding a seal into a picture under orthogonal wavelet transform," in *Proc. Int. Conference on Multimedia Computing and Systems*, pp. 514–521, June 1996.

[6] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Proc. Int. Conference in Image Processing*, vol. 2, pp. 86–90, 1994.

[7] J. Zhao, "Look, it's not there," *Byte*, January 1997.