Wavelet-based digital image watermarking

Houng-Jyh Mike Wang, Po-Chyi Su and C.-C. Jay Kuo

Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089-2564

houngjyh@sipi.usc.edu; pochyisu@sipi.usc.edu; cckuo@sipi.usc.edu

Abstract: A wavelet-based watermark casting scheme and a blind watermark retrieval technique are investigated in this research. An adaptive watermark casting method is developed to first determine significant wavelet subbands and then select a couple of significant wavelet coefficients in these subbands to embed watermarks. A blind watermark retrieval technique that can detect the embedded watermark without the help from the original image is proposed. Experimental results show that the embedded watermark is robust against various signal processing and compression attacks.

OCIS codes: (100.2000) Digital image processing; (999.9999) Digital watermarking

References

1. Houng-Jyh Wang and C.-C. Jay Kuo, "High fidelity image compression with multithreshold wavelet coding (MTWC)," SPIE's Annual Meeting - Application of Digital Image Processing XX, San Diego, CA July 27 - August 1 (1997)
2. A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," IEEE Trans. on Circuits and Systems for Video Technology, pp. 243-250, June (1996)
3. Xiang Gen Xia, Charles G. Boncelet and Gonzalo. R. Arce, "A multisolution watermark for digital images," International Conference on Image Processing, IEEE Signal Processing Society, Santa Barbara, CA, July (1997)

1. Introduction

With the rapid growth of Internet technologies and wide availability of multimedia computing facilities, the enforcement of multimedia copyright protection becomes an important issue. Digital watermarking is viewed as an effective way to deter content users from illegal distributing. In recent years, digital watermarking has been intensively studied to achieve this goal.

We can classify digital watermarking into two classes depending on the domain of watermark insertion, i.e. the spatial- and the frequency-domain watermarking. Spatial domain watermarking is easy to implement and requires no original image for watermark detection. However, it often fails under signal processing attacks such as filtering and compression. Besides, the fidelity of the original image data can be severely degraded since the watermark is directly applied on the pixel values. Frequency domain watermarking generally provides more protection under most of the signal processing attacks. But the existing frequency-domain watermark algrithms require the original image for comparison in the watermark retrieval process, which is not practical for a huge image database. Furthermore, the necessity of progressive transmission is one of the requirements for Internet distribution. The lack of progressive transmission property in existing spatial- and frequency-domain watermarking algorithms limits their Internet applications.

#7081 - $15.00 US Received October 28, 1998; Revised November 30, 1998

To solve the above problems associated with existing watermarking algorithms, we propose a new frequency-domain wavelet-based watermarking technique. The proposed method searches significant coefficients across subbands to embed the watermark. The watermark is adaptively weighted in different subbands to achieve robustness as well as high perceptual quality. Besides, a blind watermark retrieval algorithm (i.e. to retrieve watermark without the original image as a reference) can be applied. Experimental results show that the proposed algorithm can survive with various geometric, filtering, and compression attacks.

This paper is organized as follows. The procedure to search significant wavelet coefficients is described in Section 2. The adaptive watermark casting and retrieval algorithms are presented in Section 3. Experimental results are given in Section 4. Finally, concluding remarks are provided in Section 5.

2. Significant coefficient search

The procedure to search significant wavelet coefficients is motivated by the principle for the design of the multi-threshold wavelet codec (MTWC) [1]. The successive subband quantization (SSQ) scheme was adopted in MTWC to choose perceptually significant coefficients for watermark casting. These coefficients are sorted according to their perceptual importance. Furthermore, the current quantization threshold for each subband is used as the weighting function of the embedded watermark. This method gives a perceptual weighting for different significant wavelet coefficients, and sets a limit on the bound of fidelity loss after watermark casting.

After the wavelet transform, we assume that wavelet coefficients in each subband follow the Gaussian distribution. It was shown in [1] that, for a Gaussian distribution with zero mean and variance σ^2, the simplified rate-distortion model becomes

$$D = \begin{cases} T^2/12, & \sigma^2 > T^2/12, \\ \sigma^2, & \sigma^2 \leq T^2/12. \end{cases} \tag{1}$$

As given in (1), D is proportional to the square of the current threshold T if the corresponding variance σ is large. A larger value of D implies that this subband contains more energy and should be treated as a significant subband in comparison with other subbands. Thus, we can search significant subbands based on their maximum thresholds.

MTWC is a bit-plane coder. Each coefficient $C_s(x,y)$ in subband s can be represented by

$$C_s(x,y) = \text{sign} \times (a_0 \frac{T}{2^0} + a_1 \frac{T}{2^1} + ... + a_b \frac{T}{2^b} + ...), \tag{2}$$

where "sign" is the sign value (e.g. $+1$ for positive sign or -1 for negative sign) of coefficient $C_s(x,y)$, b is the bit-plane layer number ($b = 0$ indicates the most significant bit (MSB) plane layer), a_b is the binary bit at the bth bit plane and T is the initial threshold of subband s calculated via

$$T = C_{\max,s}/2, \tag{3}$$

where $C_{\max,s}$ is the maximum absolute coefficient value in subband s. The significant coefficient searching procedure can be summarized as follows.

1. Set the initial threshold T_s of each subband to one half of its maximum absolute value of coefficients inside the subband. Set all coefficients un-selected.

2. Select the subband (except the DC term) with the maximum value of $\beta_s T_s$, where β_s is the weighting factor of subband s. For the selected subband, we examine

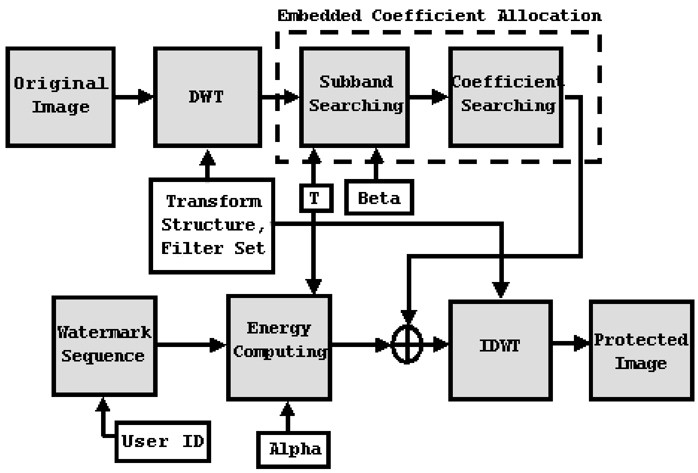
#7081 - $15.00 US
Received October 28, 1998; Revised November 30, 1998

all un-selected coefficients $C_s(x, y)$ and choose coefficients greater than the current threshold T_s as significant coefficients. Cast the watermark in these selected significant coefficients.

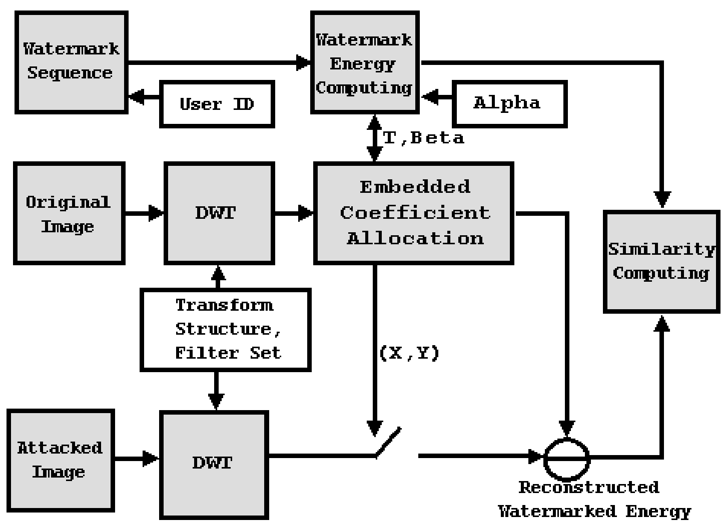
3. Update the new threshold in subband s via $T_s^{new} = T_s/2$.

4. Repeat Step 2 to Step 3 until all watermark symbols are cast.

3. Adaptive watermark casting and retrieval



(a)



(b)

Figure 1. The blockdiagram of invisible watermark embedding and detection (a)embedding (b)detection.

3.1 *Invisible Watermark Casting*

Figure 1(a) shows the blockdiagram of invisible watermark embedding. We perform watermark casting by using the spread spectrum technique. The watermark casting is performed as

$$C'_{s,k}(x,y) = C_s(x,y) + \alpha_s \beta_s T_s W_k, \tag{4}$$

where C' is the coefficient of the watermarked image, C is the original coefficient, α_s and β_s are scaling factors, T_s is the current threshold of subband s in the jth bit plane, and W_k is the kth watermark element in a watermark sequence of length N_w. W_k takes value between 1 and -1. The value of α_s is adjustable by users to increase (or decrease) the watermarked image fidelity and decrease (or increase) the security of watermark protection at the same time. It is chosen that $\alpha_s \in (0.0, 1.0]$. The error introduced by watermark insertion is

$$E_{s,k}(x,y) = \alpha_s \beta_s T_s W_k, \tag{5}$$

Given $|W_k| \leq 1.0$, the mean square error (MSE) introduced by the watermark can be computed as:

$$MSE \leq \frac{\sum_{s=1}^{N_s} \sum_{j=1}^{N_{bit}} H_s^j (\alpha_s \beta_s T_s)^2}{Height \times Width},$$

where H_s^j is the number of significant coefficients in subband s of the jth bit layer, $Height$ and $Width$ are the height and width parameters of the image, N_s is the total number of subbands (except for the lowest-frequency subband), N_{bit} is the total number of cast watermark bits per bit layer per subband. We would like to provide a watermarked image of high fidelity with a PSNR value greater than 35 dB. This implies that MSE should be less than 20.56 for an 8-bit gray level image. Thus, we have

$$MSE \leq \frac{N_w}{Height \times Width} (\alpha_{max})^2 (T_{max})^2 (\beta_{\max})^2 \leq 20.56, \tag{6}$$

where N_w is the size of the watermark sequence, T_{max} is the maximum value of T_s, $\alpha_{\max}$ is the maximum value of α_s, and $\beta_{\max}$ is the maximum value of β_s. By using the fact that $\beta_{max} = 1.0$, we can obtain a constraint on α_{max}, i.e.

$$\alpha_{max} \leq \frac{1}{T_{max}} \sqrt{\frac{20.56 \times Height \times Width}{N_w}}. \tag{7}$$

We see from (6) that a larger value of α will result in a watermarked image of lower fidelity.

3.2 Invisible Watermark Detection

When the watermarked image I' is distributed to the public, it could go through various attacks to result in an attacked image I^*. The difference between coefficient C^* of I^* and coefficient C of the original image I in the selected significant coefficient position (x, y) can be written as

$$E^*_{s,k}(x,y) = C^*_{s,k}(x,y) - C_s(x,y).$$

The similarity between C^* and C is calculated as

$$SIM(I^*, I) = N_w \frac{\sum_{k=1}^{N_w} E^*_{s,k}(x,y) \cdot E_{s,k}(x,y)}{\|E^*_{s,k}(x,y)\| \|E_{s,k}(x,y)\|}, \tag{8}$$

where $E_{s,k}(x,y)$ is the original watermark and $E^*_{s,k}(x,y)$ is the attacked watermark with respect to wavelet coefficient $C_s(x,y)$. The blockdiagram of watermark detection is given in Figure 1(b). As shown in (6), MSE between watermarked and original images is proportional to α^2_{max}. Thus, the root mean square error (RMSE) is propotional to α_{max}. If the distortion of the watermarked image's coefficient due to an attack is greater than RMSE, the watermark cannot be successfully retrieved via (8). Thus, robustness to attacks is highly dependent on α_s.

#7081 - $15.00 US
Received October 28, 1998; Revised November 30, 1998

3.3 Blind Watermark Detection

By blind watermark detection, we mean to retrieve the embedded watermark without any information from the original image. The most difficult problem associated with blind watermark detection in the frequency domain is to identify coefficients with the watermark inserted and the embedded watermark values. We develop a blind watermark detection algorithm by truncating selected significant coefficients to some specified value.

The main idea of blind watermarking comes from (2). Let $C_{s,b}(x,y)$ be the selected significant coefficient in the b^{th} bit plane of subband s, i.e. $T_{s,b} \leq C(x,y) \leq T_{s,b-1}$, then the watermarked version of $C_{s,b}(x,y)$ is:

$$C'_{s,b,k}(x,y) = sign \times \Delta_p(C_{s,b}(x,y)) + \alpha_s \beta_s T_{s,b} W_k. \tag{9}$$

where $sign$ is the sign value of $C_{s,b}(x,y)$. The operation Δ_p is defined as

$$\Delta_p(C_{s,b}(x,y)) = (1 + 2p\alpha_s)T_{s,b}, \tag{10}$$

where $T_{s,b}$ is defined in (3) with subband number s and bit plane number b, and p is an integer between 1 and $(2\alpha_s)^{-1}$. The distance $DIS_{s,b,p}(x,y)$ between $\Delta_p(C_{s,b}(x,y))$ and $C_{s,b}(x,y)$ is defined as

$$DIS_{s,b,p}(x,y) = \left|\Delta_p(C_{s,b}(x,y)) - |C_{s,b}(x,y)|\right|.$$

Then, we can obtain p by

$$p = \arg\min_{p'} DIS_{s,b,p'}(x,y).$$

After p is selected, we have

$$DIS_{s,b,p}(x,y) \leq 2\alpha_s T_{s,b}.$$

The blind watermark detection formula is basically the same as (8) with the replacement of $E_{s,k}(x,y)$ by $E^*_{s,k}(x,y)$, which is

$$E^*_{s,k}(x,y) = C^*_{s,b,k}(x,y) - sign \times \Delta_p C^*_{s,b,k}(x,y),$$

where $T^*_{s,b}$ is obtained from C^*. Since $T^*_{s,b}$ comes from the largest coefficient in subband s, it is not easily attacked. Besides, no watermark is embedded, so $T^*_{s,b}$ should be very close to $T_{s,b}$.

Given the constraint $\beta_s \leq 1.0$, if the distortion on $C'_{s,b,k}(x,y)$ is less than $\alpha_s T_{s,b}$, then the watermark on $C_{s,b}(x,y)$ can be perfectly detected. A larger α_s could provide more robustness to attacks but a poorer PSNR performance in the watermarked image. In order to provide a watermarked image with PSNR greater than 35.0 dB, we have

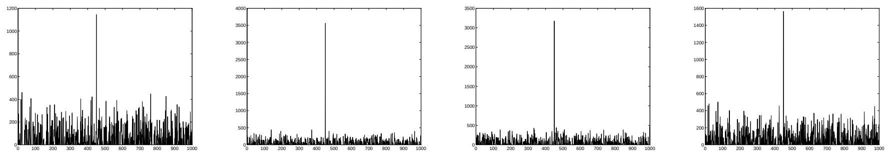
$$MSE \leq \frac{N_w}{Height \times Width}(2\alpha)^2(T_{max})^2(\beta)^2 \leq 20.56. \tag{11}$$

Thus, given that $\beta \leq 1.0$, the value of α_s is limited by

$$\alpha_{max} \leq \frac{1}{4T_{max}}\sqrt{\frac{20.56 \times Height \times Width}{N_w}}. \tag{12}$$

Compared with (7), we see that the performance of blind watermark detection is 4 times poorer than the watermark retrieval algorithm with the original image as a reference since the value of α is propotional to robustness against attacks.

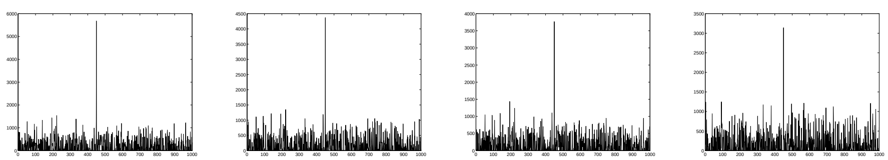
4. Experimental results



(a) (b) (c) (d)

Figure 2. Watermark retrieval from the watermarked 512×512 gray-level Lena image after (a) the 6×6 block mosaic attack, (b)the 50% uniform random noise attack, (c) the JPEG compression attack with 5% quality factor setting, and (d) the 512:1 compression attack with SPIHT.

We performed watermark protection on the Lena image of size 512×512 with $N_w = 8,192$ and $\alpha = 1.0$. The PSNR between the original image and the watermarked image is 37.20 dB. Watermark retrieval results after the attack of 6 by 6 block mosaic lowpass filtering and 50% uniform random noise attack were shown in Figures. 2 (a) and (b), respectively. We see clearly that the embedded watermark with ID number 450 is retrieved successfully. It is worthwhile to point out that most existing spatial domain watermark algorithms cannot survive under even a 4 by 4 block mosaic filtering attack. The proposed algorithm can survive well under a DCT-based JPEG compression attack. The result after the JPEG compression with the quality factor set to 5% is shown in Figure. 2(c). Moreover, Figure. 2(d) shows that our algorithm can survive even under a 512:1 compression attack by using a wavelet-based compression codec known as SPIHT[2].



(a) (b) (c) (d)

Figure 3. Blind watermark retrieval for the gray-level Lena image of size 512×512 after (a) no attack, (b) soften filter attack, (c) JPEG compression attack and (d) 64:1 compression ratio attack by SPIHT.

Figure. 3(a) shows the blind watermark retrieval result of a watermarked Lena image without any attack, where the weighting factor α was set to 0.125. The watermark was retrieved successfully without any information from the original image. The PSNR value between the original image and the watermarked image is 46.18dB. Figures. 3 (b), (c) and (d) show the watermark retrieval results after soften, default JPEG compression and SPIHT 64:1 compression attacks. As shown in these figures, the embedded watermark with ID number 450 was retrieved successfully under these attacks.

5. Conclusion

In this work, we developed a perceptual watermark casting scheme which searches the perceptually significant wavelet coefficients. The watermark sequence is cast into selected significant coefficients to provide a higher tolerance to various attacks. Moreover, the fidelity of the watermarked image can be adjusted by using the weighting factor α of the cast watermark energy. A blind watermark retrieval technique was also proposed and analyzed. It was demonstrated by experiments that the proposed algorithm can provide an excellent protection under various attacks.