Blind Wavelet-based Image Watermarking Based on HVS and Neural Networks

Hung-H. Tsai [1] Chi-C. Liu [1] Kuo-C. Wang [1]

[1] Dept. of Information Management, National Formosa University, Yulin, Taiwan 632, R.O.C.

Abstract

This work proposes a Blind Wavelet-based Image Watermarking (BWIW) technique, based on the human visible system (HVS) model and neural networks, for image copyright protection. A characteristic of the HVS, the Just Noticeable Difference (JND) profile, is employed in the watermark embedding so that the BWIW technique can make the watermark further imperceptible. The technique is developed in the wavelet domain. While embedding a watermark in an image, it adds the adaptive strengths to the wavelet coefficients of the image according to the JND thresholds. Moreover, an artificial neural network (ANN) is used to memorize the relationships between the wavelet version of an original image and its watermarked image. An advantage of the BWIW technique is that it utilizes the trained ANN to estimate the watermark without the original image to be applied in the calculation of the JND profile of the image. Finally, computer simulations demonstrate that both the transparency and the robustness of the BWIW technique are better than that of other proposed methods.

Keywords: Image Watermarking, Human Visual System, Neural Networks, Wavelets, Image Authorization.

1. Introduction

Due to the rapid growth of the Internet and the extensive evolution of digital technologies, the availability of digital multimedia content has sharply increased. Watermarking, which allows for the imperceptibly embedding information in an original multimedia data, has widely emerged for copyright protection and ownership identification [1]-[6]. Also, it can be used for tracing multimedia products that have been illegally distributed.

The theoretical models of the HVS have been vastly applied in image processing. A main purpose of exploiting the characteristics of the HVS is to effectively find the image information which can be removed without degrading the subjective image quality of the visual perception. A JND profile of an image is a key concept of the psycho-visual properties of the HVS. Although some proposed watermarking techniques employed the JND profile to enhance their transparency and robustness, they still suffer from two drawbacks. First, the JND profile of a wavelet-transformed image is not used in the design of their techniques during watermark embedding [1, 2]. Second, the original images are required for the calculation of the JND profile of the images during watermark extraction [3, 4]. In [5, 6], Wang *et al.* proposed a watermarking method, based on the JND profile, for JPEG and JPEG2000 images. The method has a difficulty to apply it to solve real-world problems unless it modifies the encoding and decoding algorithms of JPEG and JPEG2000 standards. Consequently, this paper proposes the BWIW technique, based on the HVS and neural networks, to overcome the limits mentioned above. In [6], they modified Waston's model to estimate the JND profile for Discrete Wavelet Transform (DWT) coefficients, and then used the JND profile to develop a watermarking method. Unfortunately, the method required original images while extracting watermarks. In [7], we employed Waston's model to derive the allowable visibility ranges of the JND thresholds for all DWT coefficients. Next, the BWIW technique hides a watermark in some coefficients of a wavelet-transformed image via the modifications to the coefficients according to the allowable visibility ranges of the JND thresholds. This is, the BWIW technique can adaptively control the modification strengths for all coefficients to be embedded so that it can make the watermark further transparent. Subsequently, the BWIW technique exploits an ANN with the multi-layer perceptrons (MLP) for memorizing relationships between an original DWT image and its watermarked image. During watermark extraction, the BWIW technique uses the Trained ANN (TANN) to estimate the watermark without the original image. In contrast, some methods require the original image to be applied in the JND-profile calculation during watermark extraction [3, 5]. Observing the computer simulations, the BWIW technique definitely outperforms other proposed

methods for both the transparency and the robustness.

The rest of this paper is organized as follows. Section 2 introduces the image denotations and the DWT. Next, Section 3 describes the BWIW technique. Subsequently, Section 4 shows experimental results. Finally, Section 5 gives conclusions.

2. Background

2.1. Image denotations and DWT

A gray-level image, X, with size $L \times K$ can be defined by $X = [x_\rho]_{L\times K}$, where $x_\rho \in \{0, 1, \ldots, 255\}$. Here x_ρ represents the pixel value located at position $\rho = (i, j)$ over X, where $i \in \{0, 1, \ldots, L\text{-}1\}$ and $j \in \{0, 1, \ldots, K\text{-}1\}$. Fig. 1 shows X is segmented into $\lfloor \frac{L}{8} \rfloor \times \lfloor \frac{K}{8} \rfloor$ non-overlapped blocks with size 8 × 8. For example, b_{21} stands for a block that the center pixel of the block is located at position (2, 1) on X. Let $b_{21}(r, c)$ denote the gray level of a pixel at the position (r, c) in the block b_{21}. As a result, these nonoverlapped blocks in X can be denoted by

$$\Phi = \left\{ b_{ij} \mid i=1,\cdots,\left\lfloor \frac{L}{8} \right\rfloor, j=1,\cdots,\left\lfloor \frac{K}{8} \right\rfloor \right\} \qquad (1)$$

where b_{ij} stands for a size 8 × 8 block.

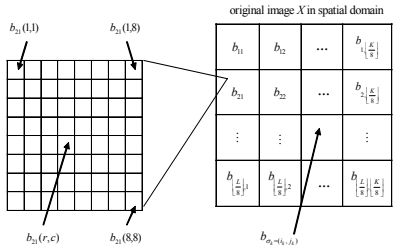

Fig.1: The segmented image in spatial domain.

An image X is decomposed into LL, HL, LH and HH subbands through DWT transformation based on the linear-phase 9/7 wavelets. Fig. 2 shows a block divided into 7 subbands with two levels DWT.

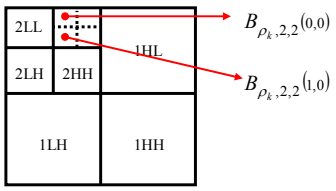

Fig. 2: Components are constituted of a wavelet image block of a size 8 x 8 at level 2.

3. The BWIW technique

3.1. Watermark representation

In the experiment of the paper, a 2D binary image serves as a watermark W. The binary 2D image can be expressed as a binary sequence in a row-major fashion. The watermark is denoted by

$$W = (w_1, w_2, \cdots, w_k, \cdots, w_m) \qquad (1)$$

where m denotes the size of W and $w_k \in \{-1, 1\}$ for each k.

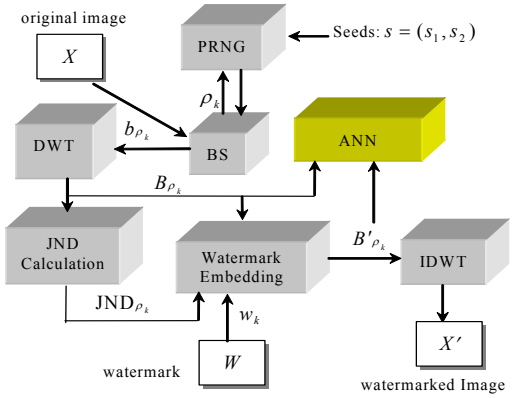

Fig 3: The block diagram of the watermark-embedding algorithm in the BWIW technique.

3.2. Watermark embedding

Fig. 3 illustrates the structure of the watermark-embedding algorithm of the BWIW technique. First, two seeds s_1 and s_2 are given, and then present the pseudo-random number generator (PRNG) component with these two seeds to generate a sequence of random positions,

$$\Gamma = \{\rho_1, \rho_2, \ldots, \rho_k \ldots, \rho_m\}, \qquad (2)$$

where $\rho_k = (i_k, j_k)$, $k = 1, \ldots, m$. In the BWIW technique, a set, Ψ, comprises m blocks which are randomly selected from Φ by using the PRNG scheme [9]. The set Ψ can be expressed as a form

$$\Psi = \left\{ b_{\rho_k=(i_k,j_k)} \mid \rho_k \in \Gamma \right\} \qquad (3)$$

where $k = 1, 2, \ldots, m$. That is, $|\Psi| = m$. Note that each b_{ρ_k} is of size 8 × 8. In the block diagram, B'_{ρ_k} denotes a wavelet block corresponding to the spatial block b_{ρ_k}.

The watermark-embedding algorithm of the proposed technique is given as follows.

Step 1. Input an original image X and watermark W.

Step 2. The original image X is segmented into 8×8 blocks.

Step 3. Present the PRNG scheme with two seeds (s_1, s_2) to generate the set Γ.

Step 4. According to Γ, find Ψ which includes the target blocks to be embedded.

Step 5. For each b_{ρ_k} where $\rho_k = (i_k, j_k) \in \{1, 2, \ldots, L/8\} \times \{0, 1, \ldots, K/8\}$.

Step 6. Compute B_{ρ_k} = DWT(b_{ρ_k}) where $\rho_k \in \Gamma$.

Step 7. Compute the JND thresholds, JND_{ρ_k} for b_{ρ_k} where $\rho_k \in \Gamma$.

Step 8. $B'_{\rho_k,2,2}(0,0) = B_{\rho_k,2,2}(0,0) + w_k\left(JND_{\rho_k,2,2}(0,0) + \alpha\right)$

Step 9. Compute b'_{ρ_k} = IDWT(B'_{ρ_k}) where B'_{ρ_k} represents the watermarked block in the wavelet domain.

Step 10. End for-loop.

Step 11. Output the watermarked image X'.

Following the completion of the embedding algorithm, the BWIW technique exploits an ANN for memorizing relationships between an original DWT image and its watermarked image [9]. Fig. 4 shows the architecture of the ANN which comprises three layers: the input layer with sixteen neurons, the hidden layer with twelve neurons, and the output layer with a single neuron. The ANN is so-called a 16-12-1 MLP. A back-propagation learning algorithm is applied to train the ANN by correcting its weights. The weights are adjusted to decrease the errors between the inputs and their corresponding target outputs. A physical output of the ANN is represented by $\hat{B}_{\rho_k,2,2}(0,0)$.

A set of training patterns Θ can be gathered and expressed as a form

$$\Theta = \left\{\left(\mathbf{B}'_k, B_{\rho_k,2,1}(0,0)\right) \mid k = 1, 2, \cdots, m\right\} \qquad (4)$$

where the vector $\mathbf{B}'_k$ denotes the input pattern of the kth training pattern in Θ, and is represented by

$$\mathbf{B}'_k = \left(B'_{\rho_k,2,1}(0,0), B'_{\rho_k,2,1}(0,1), \cdots, B'_{\rho_k,2,4}(2,2)\right). \qquad (5)$$

In addition, $B_{\rho_k,2,1}(0,0)$ represents the desired output of the kth training pattern.

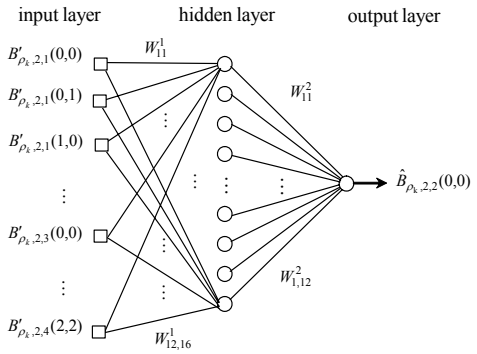

Fig. 4: The architecture of an ANN used in the BWIW technique.

3.3. Watermark extraction

Fig. 5 depicts the block diagram of the watermark extraction utilized in the BWIW technique. The watermark-extraction algorithm uses the TANN to estimate the watermark without original images which is employed to compute the JND thresholds for DWT coefficients. The algorithm is described as follows.

Step 1. Input a watermarked image X' and two seeds (s_1, s_2).

Step 2. Present the PRNG scheme with two seeds (s_1, s_2) to generate Γ.

Step 3. Find all watermarked blocks b'_{ρ_k} according to Γ.

Step 4. For k = 1 to m

Step 5. Compute B'_{ρ_k} = DWT (b'_{ρ_k}) where $\rho_k \in \Gamma$.

Step 6. Feed the TANN with $\mathbf{B}'_k$ to estimate $\hat{B}_{\rho_k,2,2}(0,0)$.

Step 7. If $\left(B'_{\rho_k,2,2}(0,0) - \hat{B}_{\rho_k,2,2}(0,0)\right) \geq 0$ then $\hat{w}_k$ = 1 else $\hat{w}_k$ = -1.

Step 8. End for-loop

Step 9. Output the estimated watermark $\hat{W}$.

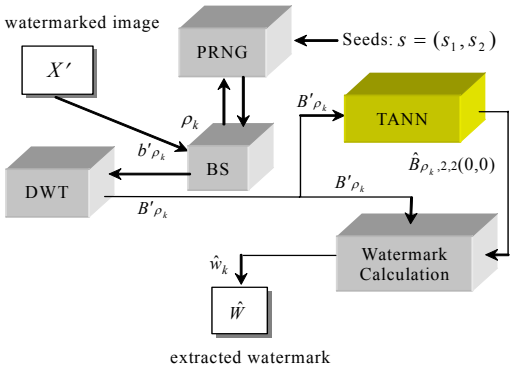

Fig. 5: The block diagram of the watermark-extraction algorithm in the BWIW technique.

4. Experimental results

In this experiment, a size 64 × 64 stamp binary image is taken to be the owner information. Five images, Peppers, Lena, Baboon, Couple, and Barbara, are used in the assessment for the performance of the BWIW technique. Fig. 6 shows the comparison results of the BWIW technique, Joo's [1], and Wang's methods [3] for the transparency. For the case of attack free, Fig. 7 displays the comparison results of these three methods. An observation is that the BWIW technique cannot outperform two proposed methods

for Baboon image. The reason is that the textures of Baboon image are highly complex. That is, it has considerable variations in the magnitudes of DWT coefficients of Baboon image. This reduces the generalization ability of an ANN to precisely memorize the relationship of original image and its watermarked image.

A Stirmark 4.0 benchmark [10] is used in the robustness investigation. Fig. 8 shows the comparison results in terms of the average of all NC values for JPEG compression, noising, median filtering, histogram equalization, sharpening, and cropping. In order to test the generalization ability of the BWIW technique, the weights of the TANN are the same as that used in the attack-free case. The above results demonstrate that the BWIW technique can effectively resist common image-processing attacks, and that it definitely outperforms other proposed methods.

5. Conclusion

This paper has been successfully incorporated the HVS model and the techniques of neural networks to devise the BWIW technique. Moreover, the technique employs the JND profile to embed watermarks so that it makes the watermark further imperceptible. Furthermore, an artificial neural network (ANN) is used to memorize the relationships between an original DWT image and its watermarked image. An advantage of the BWIW technique is that it utilizes the trained ANN to estimate the watermarks without the original image to be applied in the calculation of the JND profile of the image. Finally, computer simulations claim that both the transparency and the robustness of the BWIW technique are remarkably better than that of other proposed methods.

6. References

[1] S. Joo, Y. Suh, J. Shin, H. Kikuchi, and S.-J. Cho, "A new robust watermark embedding into wavelet DC componets," *ETRI Journal*, vol. 24, no. 5, pp. 401-404, 2002.

[2] Y. Wang and A. Pearmain, "Blind image data hiding based on self reference," *Pattern Recognition Letters*, pp. 1681-1689, 2004.

[3] X.-D. Zhang, J. Feng, and K.-T. Lo, "Image watermarking using tree-based spatial-frequency feature of wavelet transform," *Journal Visual and Communication Image Representation*, vol. 14, pp. 474–491, 2003.

[4] S.-H. Wang and Y.-P. Lin, "Wavelet tree quantization for copyright protection watermarking," *IEEE Transactions on Image Processing*, vol. 13, no. 2, pp. 154-165, Feb. 2004.

[5] H. W. Wong and C. Au, "A capacity for jpeg2000-to-jpeg2000 images watermarking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 746-752, Aug. 2003.

[6] H. W. Wong, Y. M. Yeung, and C. Au, "Capacity for jpeg2000-to-jpeg2000 images watermarking," *Proc. of 2003 IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 485-488, 6-9 Jul. 2003.

[7] C.-C Liu and H.-H. Tsai, "Robust image watermarking with visibility range estimation based on HVS," *the 16th Information Security Conference*, Taiwan, 2006.

[8] M. Blum and S. Micali, "How to generate cryptographically strong sequences of pseudo-random bits," *SIAM Journal of Computing*, vol. 13, no. 4, pp. 850-864, Nov. 1984.

[9] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2th ed., New York, Macmillan College Publishing Company, 1999.

[10] http://www.petitcolas.net/fabien/watermarking/stirmark/.

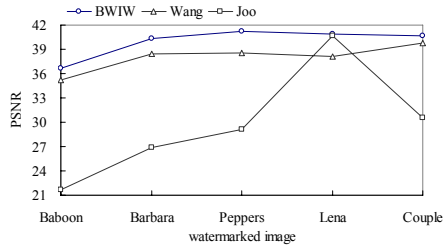

Fig. 6: Comparison results for the transparency.

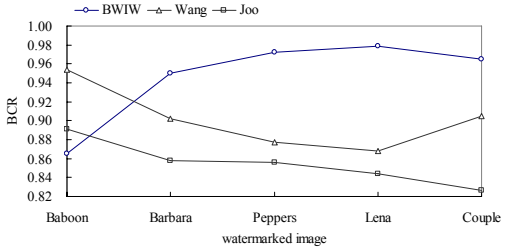

Fig. 7: Comparison results for the attack-free case.

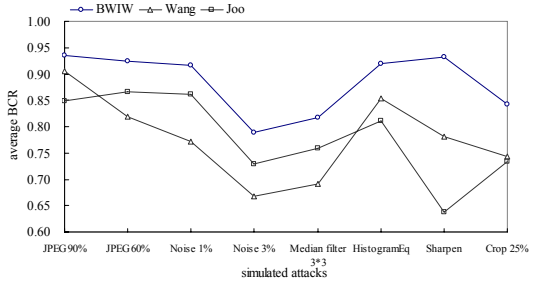

Fig. 8: Comparison results for the robustness examination.