Journal of Southeast University (English Edition) Vol. 23, No. 2, pp. 211 - 215 June 2007 ISSN 1003—7985

Digital watermarking algorithm based on neural network in multiwavelet domain

Wang Zhenfei[1,2] Song Shengli[3]

([1]College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)
([2] College of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)
([3]Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: A novel blind digital watermarking algorithm based on neural networks and multiwavelet transform is presented. The host image is decomposed through multiwavelet transform. There are four subblocks in the *LL*-level of the multiwavelet domain and these subblocks have many similarities. Watermark bits are added to low-frequency coefficients. Because of the learning and adaptive capabilities of neural networks, the trained neural networks almost exactly recover the watermark from the watermarked image. Experimental results demonstrate that the new algorithm is robust against a variety of attacks, especially, the watermark extraction does not require the original image.

Key words: digital watermarking; neural networks; multiwavelet transform

In the past decade, as the delivery and distribution of multimedia information by using computer networks has become easier, the difficulty of protecting digital images from illegal copies and reproduction has turned out to be more challenging. To overcome this major problem, digital watermarking is currently considered as the most convenient and promising technical solution[1]. Digital image watermarking is the process of embedding directly some digitized information into the host image by making small modifications to the image with a form imperceptible to the human eye. Such a watermark must be resistant against attack. Thus, the digital watermarking can be used to identify the rightful owner.

In general, digital watermarking schemes can be performed in a spatial domain and a transform domain, where the properties of the underlying domain can be exploited. Spatial domain schemes directly embed messages in pixels of an image. The least significant bit (LSB) scheme is the most common and the easiest method for embedding messages in an image. Celik et al. proposed a generalized LSB (g-LSB) embedding algorithm[2], which can introduce several additional points along the capacity-distortion curve. But the disadvantage is its lower security and proneness to distortion. For transform domain methods, images are transformed to a frequency domain, and then messages are embedded into the frequency coefficients, such as the discrete Fourier transform (DFT), the discrete cosine transform (DCT)[3], and the discrete wavelet transform (DWT)[4]. Usually, the transform domain scheme is more robust in resisting image processing attacks than the spatial domain scheme. Among the transform domain watermarking techniques, ones based on the DWT are gaining more popularity.

Dietl et al. used wavelet filter parameterization as a secret transform domain to improve the security of the watermarking method[4]. Wang and Lin proposed wavelet tree based watermarking algorithms[5]. A new wavelet based watermarking scheme for copyright protection of digital images is presented in Ref. [6]. Ghouti et al. proposed a robust watermarking algorithm using a balanced multiwavelet transform[7]. Compared with existing models based on scalar wavelets, this watermarking system clearly shows the capacity gains. It is robust against typical watermark attacks. Another way to improve the performance of watermarking schemes is to make use of neural network techniques. The digital watermarking problem can be viewed as an optimization problem. Therefore, it can be solved by neural networks. Applications of neural networks to settle watermarking problems have been researched. Yu et al. proposed a semi-public watermarking based neural network for color images[8]. Their method can greatly improve the performance of Kutter's technique[9].

Due to the fact that the watermarking scheme based on the discrete multiwavelet transform (DMWT) is robust against image processing attacks and the learning and adaptive capabilities of the neural net-

Received 2006-10-20.

Foundation item: The National Natural Science Foundation of China (No. 60473015).

Biography: Wang Zhenfei (1973—), male, doctor, associate professor, wzfei@ zzu. edu. cn.

works. This paper proposes an effective digital watermarking algorithm based on neural networks in the multiwavelet domain. Experimental results demonstrate that, compared with other techniques, the new algorithm is more effective. Moreover, the watermark extraction does not require the original image, so the application is more practical in real life for ownership verification.

1 Two-Dimensional (DMWT) of Image

As in the scalar wavelet case, the theory of the multiwavelet case is based on the idea of multiresolution analysis (MRA). They are very similar to wavelets but have some important differences. Wavelets have an associated scaling function $\phi(t)$ and a wavelet function $\psi(t)$, but multiwavelets have two or more scaling and wavelet functions. For notational convenience, the set of the scaling function can be written using the vector notion $\boldsymbol{\Phi}\{t) = \{\phi_1(t), \phi_2(t), \ldots, \phi_N(t)\}^{\mathrm{T}}$, where $\boldsymbol{\Phi}(t)$ is called the multiscaling function. Likewise, the multiwavelet function is defined from the set of wavelet functions as $\boldsymbol{\Psi}(t) = \{\psi_1(t), \psi_2(t), \ldots, \psi_N(t)\}^{\mathrm{T}}$. When $N = 1$, $\boldsymbol{\Psi}(t)$ is called a scalar wavelet, or simply a wavelet. While N can be arbitrarily large, the multiwavelets studied to date are primarily $N = 2$ in this paper.

The multiwavelet two-scale equations resemble those for scalar wavelets:

$$\boldsymbol{\Phi}(t) = \sum_{k=0}^{m-1} \boldsymbol{H}_k \boldsymbol{\Phi}(2t - k) \tag{1}$$

$$\boldsymbol{\Psi}(t) = \sum_{k=0}^{m-1} \boldsymbol{G}_k \boldsymbol{\Psi}(2t - k) \tag{2}$$

where $\boldsymbol{H}_k$ and $\boldsymbol{G}_k$ are matrix multifilters; $\boldsymbol{H}_k$ is a matrix lowpass filter and $\boldsymbol{G}_k$ is a matrix highpass filter. They are $N \times N$ matrices for each integer k, and m is the number of scaling coefficients. The matrix elements in these filters provide more degrees of freedom than a scalar wavelet. As in the scalar wavelet case, the multiwavelet decomposition of a one-dimensional signal is performed by the Mallat algorithm. However, because the lowpass filterbank and highpass filterbank are $N \times N$ matrices, the signal must be preprocessed to be a vector before decomposition. As for the decomposition of a two-dimensional image, it can be performed by the one-dimensional algorithm in each dimension. After a one cascade step, the result can be realized as the following matrix:

$$\begin{bmatrix} \boldsymbol{L}_1\boldsymbol{L}_1 & \boldsymbol{L}_2\boldsymbol{L}_1 & \boldsymbol{H}_1\boldsymbol{L}_1 & \boldsymbol{H}_2\boldsymbol{L}_1 \\ \boldsymbol{L}_1\boldsymbol{L}_2 & \boldsymbol{L}_2\boldsymbol{L}_2 & \boldsymbol{H}_1\boldsymbol{L}_2 & \boldsymbol{H}_2\boldsymbol{L}_2 \\ \boldsymbol{L}_1\boldsymbol{H}_1 & \boldsymbol{L}_2\boldsymbol{H}_1 & \boldsymbol{H}_1\boldsymbol{H}_1 & \boldsymbol{H}_2\boldsymbol{H}_1 \\ \boldsymbol{L}_1\boldsymbol{H}_2 & \boldsymbol{L}_2\boldsymbol{H}_2 & \boldsymbol{H}_1\boldsymbol{H}_2 & \boldsymbol{H}_2\boldsymbol{H}_2 \end{bmatrix}$$

Note that a typical block $\boldsymbol{H}_2\boldsymbol{L}_1$ contains lowpass coefficients corresponding to the first scaling function in the horizontal direction and highpass coefficients corresponding to the second wavelet in the vertical direction.

An example using multiwavelet decomposition is shown in Fig. 1(a). Compared with the scalar wavelet decomposition shown in Fig. 1(b), there are four subblocks in the LL-level of the multiwavelet domain, while there is only one in that of the scalar wavelet domain.

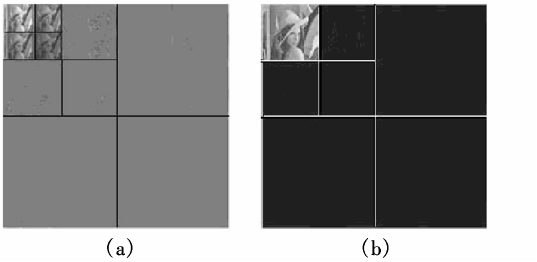
(a) (b)

Fig. 1 The difference between the multiwavelet domain and the scalar wavelet domain. (a) Decomposition by the multiwavelet; (b) Decomposition by the scalar wavelet

2 New Watermarking Scheme

The proposed method embeds the watermark to the low frequency coefficients of the multiwavelet domain. The watermark bits are added to the selected coefficients without any perceptual degradation of the host image. The watermark used for embedding is a binary logo image and a pseudo-random binary sequence, which is very small compared to the size of the host image. During the watermark recovery, the trained neural network is employed to extract the watermark.

The watermark used in our method is denoted as

$$\begin{aligned} \boldsymbol{W} = \boldsymbol{H}_{p\times q} + \boldsymbol{L}_{m\times n} = \\ \{h(1), \ldots, h(pq)\} + \{l(1), \ldots, l(mn)\} = \\ \{w(1), \ldots, w(pq + mn)\} \end{aligned} \tag{3}$$

where $\boldsymbol{H}_{p\times q}$ and $\boldsymbol{L}_{m\times n}$ are a $p \times q$-bit sequence and a $m \times n$-bit binary sequence, respectively. Moreover, "+" denotes the concatenation operation. Note that $\boldsymbol{H}_{p\times q}$ is a pseudo-random according to the secret key $\boldsymbol{K}$. It is extra information known in the beginning of watermark recovery to enhance the correctness of extracting $\boldsymbol{L}_{m\times n}$, and $\boldsymbol{L}_{m\times n}$ are pixels of a binary logo image. The main purpose of creating $\boldsymbol{H}_{p\times q}$ is to construct the training patterns for a neural network to effectively memorize the characteristics of the relationship between $\boldsymbol{W}$ and $\bar{\boldsymbol{I}}$ (the multiwavelet coefficients of the watermarked image).

2.1 Watermark embedding

The algorithm for embedding a binary watermark is formulated as follows:

Step 1 Decompose the host image by L-levels using DMWT. The low frequency sub-blocks are respectively denoted as $\boldsymbol{I}_1, \boldsymbol{I}_2, \boldsymbol{I}_3, \boldsymbol{I}_4$.

Step 2 Numerical values are sorted in descending order for the low frequency coefficients of the multiwavelet to find the threshold weight.

$$\boldsymbol{T} = S(p \times \boldsymbol{G}) \tag{4}$$

where $S()$ are the sorted low frequency coefficients, p is the percentage of multiwavelet coefficients in which watermark is embedded and $\boldsymbol{G}$ is the size of mulitiwavelet low frequency coefficients. The coefficients, which have numerical values greater than the threshold value $\boldsymbol{T}$, are considered as significant coefficients.

Step 3 Add watermark bits to significant coefficients using Eq. (5).

$$\bar{\boldsymbol{I}}(i,j) = \boldsymbol{I}(i,j) + \alpha(2w(k) - 1) \tag{5}$$

where $\boldsymbol{I}(i, j)$ are significant coefficients and the constant α is the watermark embedding strength. The greater the constant α is, the more robust the watermarked image is and the watermark is less imperceptible .

Step 4 After embedding the watermark bits, the multiwavelet transform of the image is inversed by the L-level, and the watermarked image is obtained.

2.2 Neural network training

It is well known that neural networks perform a highly adaptive nonlinear decision function from training samples. We establish the relationship among the wavelet coefficients by using the back-propagation neural networks (BPNN) model. First, the watermarked image is L-level multiwavelet decomposed and significant coefficients are obtained based on numerical values the same as in the embedding algorithm. This indicates that each significant coefficient hides the message of the watermark. Next, we use the extra information to train a neural network to make sure that it possesses the capability of memorizing the characteristics of the relationship between $\boldsymbol{W}$ and $\bar{\boldsymbol{I}}$.

In this paper, we naturally select the first $p \times q$ significant coefficients corresponding sub-blocks as the training patterns. We construct three layers BPNN with 3, 5 and 1 neurons in the input, hidden and output layers, respectively, and the tangent sigmoid and sigmoid transfer functions are used for recognition. For example, for a selected significant coefficient, $\boldsymbol{I}_1(i, j)$, the network is trained with its corresponding sub-blocks, i.e., let $\{\boldsymbol{I}_2(i,j), \boldsymbol{I}_3(i,j), \boldsymbol{I}_4(i,j)\}$ be the input vectors and the value $\tilde{\boldsymbol{I}}(i,j) = \bar{\boldsymbol{I}}(i,j) - \alpha(2w(k) - 1)$, $1 \leqslant k \leqslant pq$ be the output value. Here $w(k) \in \boldsymbol{H}_{p \times q}$ and $\boldsymbol{H}_{p \times q}$ is generated by the secret key $\boldsymbol{K}$, the same as the embedded algorithm. The neural network is trained using the training sample until it achieves convergence.

2.3 Watermark extracting

The trained neural network performs a highly adaptive nonlinear decision function f. Therefore, based on the physical output of the trained neural network, the estimated watermark can be obtained by

$$\boldsymbol{T}_{\text{in}}(i,j) = \{\boldsymbol{I}_2(i,j), \boldsymbol{I}_3(i,j), \boldsymbol{I}_4(i,j)\}_{pq+1}^{pq+mn} \tag{6}$$

$$\boldsymbol{T}_{\text{out}}(i,j) = \{\tilde{\boldsymbol{I}}(i,j)\}_{pq+1}^{pq+mn} \tag{7}$$

$$\bar{w}_{pq+t} = \bar{l}_t = \begin{cases} 1 & \text{if } \bar{\boldsymbol{I}}(i,j) \geqslant \tilde{\boldsymbol{I}}(i,j) \\ 0 & \text{otherwise} \end{cases} \quad t = 1, \ldots, mn \tag{8}$$

where $\bar{\boldsymbol{I}}(i,j)$, $\tilde{\boldsymbol{I}}(i,j)$ are the multiwavelet coefficients of the watermarked image and the output of BPNN, respectively. Hence, the estimated logo $\bar{L}$ is obtained.

3 Experimental Results

In the following experiments, two gray-level images with a size of 512 × 512, Lena and boat are used as the test images. The "panda logo" as the signature L of the watermark $\boldsymbol{W}$ and the logo is a binary image with a size of 64 × 64. The watermark $\boldsymbol{W}$ can be formed as a bit sequence $\boldsymbol{H}_{64 \times 5} + \boldsymbol{L}_{64 \times 64}$. We choose $\alpha = 0.45$ and $p = 0.25$ to balance the tradeoff between robustness and imperceptibility. The 2-level decomposition of GHM-multiwavelet[10] is used. The watermarked Lena image and the watermarked boat image are shown in Figs. 2 (a) and (b), respectively. The watermarked Lena image and the watermarked boat image have PSNR values of 42.3 dB and 42.1 dB, respectively. If the original and the watermarked images are observed, we cannot find any perceptual degradation.

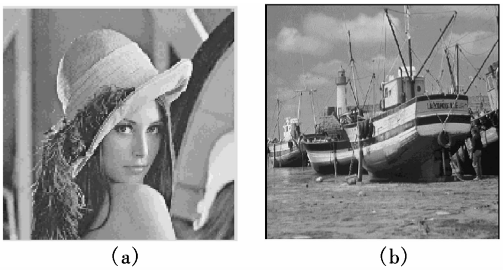
(a) (b)

Fig. 2 Test images. (a) Watermarked Lena (PSNR 42.3 dB); (b) Watermarked fishingboat (PSNR 42.1 dB)

Here the results are presented for a grayscale 8-bit Lena image of size 512 × 512. The watermarked Lena image is tested for cropping attack, and the proposed method shows better robustness. This can be observed in Fig. 3. Fig. 4 illustrates various attacks on the watermarked image by a variety of image processing operations including averaging, median filtering and additive noise. Logos extracted after applying 9 × 9 average and median filtering are shown in Figs. 4(a) and (b). After applying these filters, the images are very much degraded

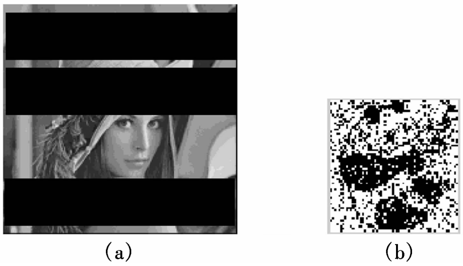

Fig. 3 Robustness to cropping. (a) 40% remained watermarked Lena image by cropping; (b) Extracted logo

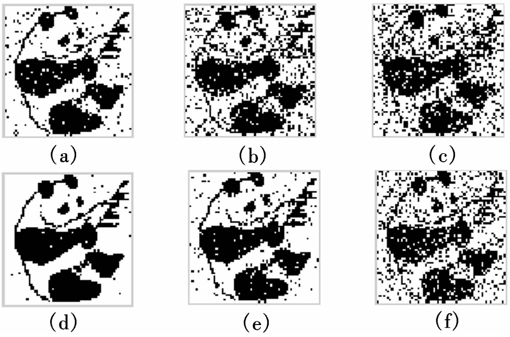

Fig. 4 Extracted logo. (a) 9 ×9 average filtering; (b) 9 ×9 median filtering; (c)50% noise adding; (d) 50% brightness increase; (e) Gaussian blurring; (f) Mosaic

and lots of data are lost but extracted logos are still recognizable. To test the robustness against adding noise, adding salt and pepper noise randomly degrades the watermarked image. The logo extracted from a 50% noise-added watermarked image is shown in Fig. 4(c). The extracted logo is noisy, but it is recognizable. The watermarked Lena image is also tested for increased and decreased brightness attacks. The extracted logo from the image with a 50% brightness increase is shown in Figs. 4 (d). Fig. 4 (e) shows the corresponding extracted watermarks after Gaussian blurring (radius = 5). Obviously, they are all recognizable and meaningful to human eyes. The results in Fig. 5 show the normalized correlation (NC) value responses with different percentages of salt and pepper noise attack. Watermark is detected up to 40% noise with our method and up to 32% noise with Liu's method[1]. Tab. 1 exhibits comparisons in terms of NC for evaluating the performance of our method, Liu ' s method and Reddy' s method[6]. The performance of our scheme is superior to that of other methods via the quantitative measures.

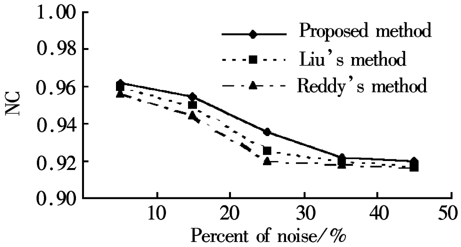

Fig. 5 Results of NC comparison with other schemes after added noise attack

Tab. 1 Comparison in terms of NC between our method and other schemes

Method	Average filtering	Median filtering	Noise adding	Brightness increase	Gaussian blurring	Mosaic
Our method	0.932	0.917	0.920	0.996	0.943	0.915
Liu' s method	0.912	0.915	0.917	0.937	0.931	0.855
Reddy' s method	0.926	0.903	0.918	0.945	0.928	0.906

Then, we test the robustness by a JPEG compression operator. Fig. 6 shows the extracted watermarks from the JPEG compressed version of the watermarked images with various compression qualities. On the other hand, the proposed method is compared with Liu' s scheme[1] and Reddy' s scheme[6] in different quality factors under JPEG compression. Fig. 7 shows the watermark NC' s response curve. Obviously, the watermark of our scheme possesses a high similarity with the original watermark under different quality factors.

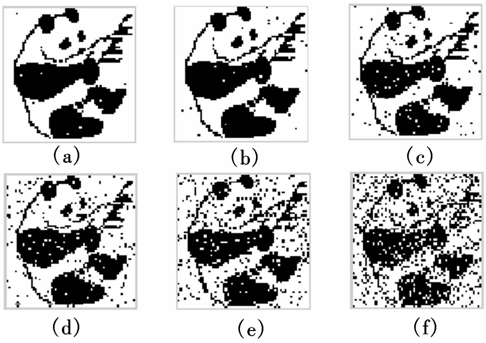

Fig. 6 Robustness to JPEG compression (quality). (a) 70%; (b) 60%; (c) 50%; (d) 40%; (e) 30%; (f) 20%

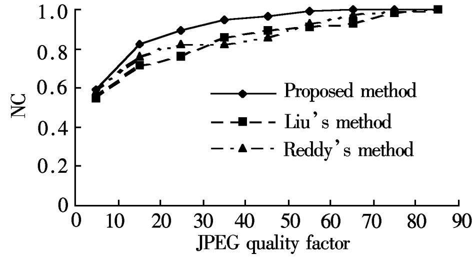

Fig. 7 Results of NC comparison with other schemes after JPEG compression

4 Conclusion

In this paper, a new blind digital watermarking scheme based on neural networks in a multiwavelet domain is proposed. The watermark is embedded into the low-frequency coefficients. We have successfully fused neural networks with watermarking to enhance the performance of conventional watermarking techniques.

Due to neural networks possessing the learning and adaptive capabilities, the trained neural networks almost exactly recover the watermark from the watermarked image against image processing attacks. Extensive experiments and comparisons with other well-known methods have been made. The proposed digital watermarking technique can be robust against many different types of attacks.

References

[1] Liu Jiang-Lung, Lou Der-Chyuan, Chang Ming-Chang, et al. A robust watermarking scheme using self-reference image[J]. *Computer Standards & Interfaces*, 2006, **28**(3): 356 - 367.

[2] Celik M U, Sharma G, Tekalp A M, et al. Lossless generalized-LSB data embedding[J]. *IEEE Transactions on Image Processing*, 2005, **14**(2): 253 - 266.

[3] Cox I J, Kilian J, Leighton T, et al. Secure spread spectrum watermarking for multimedia[J]. *IEEE Transactions on Image Processing*, 1997, **6**(12): 1673 - 1687.

[4] Dietl W, Meerwald P, Uhl A. Protection of wavelet-based watermarking systems using filter parameterization[J]. *Signal Processing*, 2003, **83**(10): 2095 - 2116.

[5] Wang Shih-Hao, Lin Yuan-Pei. Wavelet tree quantization for copyright protection watermarking[J]. *IEEE Transactions on Image Process*, 2004, **13**(2): 154 - 165.

[6] Reddy A A, Chatterji B N. A new wavelet based logo-watermarking scheme[J]. *Pattern Recognition Letters*, 2005, **26**(7): 1019 - 1027.

[7] Ghouti L, Bouridane A, Ibrahim M K, et al. Digital image watermarking using balanced multiwavelets [J]. *IEEE Transactions on Signal Processing*, 2006, **54**(4): 1519 - 1536.

[8] Yu Pao-Ta, Tsai Hung-Hsn, Lin Jyh-Shyan. Digital watermarking based on neural networks for color images[J]. *Signal Processing*, 2001, **81**(3): 663 - 671.

[9] Kutter M, Jordan F, Bossen F. Digital watermarking of color images using amplitude modulation[J]. *Journal of Electronic Image*, 1998, **7**(2): 326 - 332.

[10] Geronimo J S, Hardin D P, Massopust P R. Fractal functions and wavelet expansions based on several functions [J]. *Journal of Approximation Theory*, 1994, **78**(3): 373 - 401.

基于神经网络和多小波变换的数字水印算法

王振飞[1,2]　　宋胜利[3]

([1]华中科技大学计算机科学与技术学院,武汉 430074)
([2]郑州大学信息工程学院,郑州 450001)
([3]华中科技大学控制科学与工程系,武汉 430074)

摘要:基于图像多小波域低频系数子块的相似性,利用神经网络的学习特性,提出了新的盲数字水印算法.将宿主图像变化为多小波域,把水印加入到宿主图像多小波变化后的低频系数中.通过后向传播算法的神经网络训练出宿主图像与嵌入的水印信号之间的关系特征,利用神经网络具有学习和自适应的特性,训练后的神经网络能够完全恢复嵌入到宿主图像中的水印信息.仿真实验表明,该算法针对各种攻击具有很好的鲁棒性,特别是在水印检测时不需要原始图像.

关键词:数字水印;神经网络;多小波变换

中图分类号:TP391