Proceedings of the 2009 IEEE
International Conference on Mechatronics and Automation
August 9 - 12, Changchun, China

An Improved Watermarking Method Based on Neural Network for Color Image

Qianhui Yi and Ke Wang
School of Communication Engineering
JiLin University
Changchun, JinLin, 130022, China
yqh9835@sohu.com

Abstract **- In this paper, a novel digital watermarking scheme is devised based on improved Back-Propagation neural network (BPN) for color image. The watermark is embedded into the discrete wavelet domain of the original image and extracted by training BPN, which can learn the characteristic of the image. For improving the performance of traditional BPN, we consider the adding of momentum coefficient to reduce the error and improve the rate of the learning. The watermark can be successfully extracted by training the improved BP neural network, and the watermarking algorithm is good at defending many kinds of common attacks. The experimental results demonstrate that the proposed algorithm has good visual effect and high robustness to general image processing techniques and geometric distortions.**

Index Terms - Digital watermarking; Discrete wavelet transform; Neural network; Color image.

I. INTRODUCTION

With the rapid growth of the Internet technologies and the immediate availability of multimedia computing resources, the digital multimedia reproduction and distribution are becoming extremely easier and faster. A very crucial issue for copyright protection noticeably emerged from the development of multimedia systems [1]. One technical way to make law enforcement and copyright protection for digital media possible and practical is digital watermarking, which is defined as a process of embedding an invisible digital signature into multimedia data (image, audio, and video). In recent years digital watermarking has been intensively used to the following purposes: copy protection, fingerprinting, copyright protection, broadcast monitoring and data authentication [2].

The digital watermarking, as a technique of embedding and transmitting small amounts of data imperceptibly in the host data, has found many applications. Each watermarking application has its own special requirement. The general requirements are perceptual invisibility, capacity of the watermark, robustness, security [3]. For image data authentication, the embedded watermark has to be invisible to a human observer otherwise it may be altered by any intentional modification. The requirements on the watermarking are all related to each other. Hence, a tradeoff should be considered between the different requirements so that an optimum watermarking for each application can be developed.

Any distortion applied to the watermarked image has to be seen as an attack, this attack is an operation on the watermarked image, which attempts to corrupt or remove a digital watermark from a host signal. So a good watermarking scheme must withstand the intentional or malicious attacks. In order to prove the validity of the algorithm, different kinds of attacks will be considered in our experiments.

In the following sections, first we will introduce the related theories of the digital watermarking, such as discrete wavelet transform and Back-Propagation neural network; then we will present the structure of the watermarking scheme for color images; at last, through the experiments we will demonstrate the security and robustness of the algorithm proposed in the watermarking system.

II. RELATED THEORIES

A. Discrete Wavelet Transform

Since the inception of digital watermarking around the early 1990s, there have been a variety of methods proposed in the literature. Watermarking methods can be classified into two main categories. The first class of techniques is based on embedding data in the spatial domain [4]. Spatial domain methods usually modify the least significant bits of the host image, and are easily affected by signal processing operations. The second class of techniques is based on transform domain. The discrete cosine transform (DCT) and the discrete wavelet transform (DWT) are two of the frequently used transformations.

During the last decade, multi-resolution representation using discrete wavelet transforms has emerged as a strong alternative to the DCT. Many applications of the wavelet transform, such as compression, signal analysis and signal processing have been found [5, 6]. With the appearance of image compression standard, JPEG2000, It seems only natural to take advantage of the superior performance and modeling properties of the wavelet transform for watermarking purposes as well. The wavelet transform has a number of advantages over DCT:

1) The wavelet transform has multi-resolution description characteristic. The decoding can be processed sequentially from a low resolution to the higher resolutions. It is helpful in managing a good distribution of message in the original data and enhancing the robustness of the watermark.

2) The wavelet transform is closer to the human visual system than the DCT. The multi-resolution characteristic can

978-1-4244-2693-5/09/$25.00 ©2009 IEEE

be well matched with human visual system. The multi-resolution decomposition way accords with sensitivity characteristic of human visual nerve on horizontal and vertical directions. The watermark can be adaptively embedded into wavelet coefficients by using human visual model thresholds.

3) The main different between JPEG2000 and the traditional JPEG is that the latter primarily adopts DCT sub-area encoding method, and JPEG2000 uses multi-resolution encoding method of the wavelet transformation, which has a good ability of resistance on filtering and compression processing. Even if regarding nonstationary process, it can deal well.

A function or signal can be viewed as composed of a smooth background and details. The distinction between the smooth part and the details is determined by the resolution. In wavelet domain, a signal is divided into two parts, usually high frequency and low frequency. The edge components of the signal are largely confined in the high frequency part, the low frequency part is split again into two parts of high and low frequency. This process is continued until the signal has been entirely decomposed or stopped before by the application at hand.

In the wavelet domain, the low frequency band owns much more energy than the high frequency band. That means most energy of image exists in the low frequency, while the energy of the image edge exists at high frequency. So, in the experiment of this paper, the image is decompose by the 3rd DWT function firstly, the sub-images have no any overlapped blocks. After that, we embed watermark to each sub-image.

B. Back-Propagation Neural Network

The artificial neural network (ANN) is a powerful tool that provides an optimization procedure using high-speed computation associated with huge memories. The general sorts of ANN are illustrated in Fig.1. Unlike traditional computers, which are programmed to perform specific task, neural network have to be trained. Once trained, the network can perform the specific task for which it was constructed. Due to the abilities of learning, generalization and nonlinear approximation, ANN has been recently applied to a wide range of areas [7], such as: classification, clustering, associative memory, control and function approximation.

The back-propagation neural network is a type of supervised learning neural network; this network is a neural network in which the output of a neuron is not only determined by the inputs and weights but also the previous output. The architecture of a network consists of a description of how many layers a network has, the number of neurons in each layer, each layer's transfer function, and how the layers connect to each other. In this section, we describe a back-propagation algorithm for training three-layer feed-forward neural network, which is composed of input layer, hidden layer and output layer, and the architecture is shown in Fig.2. The BP neural network has superior abilities to learn the complex nonlinear relationships for inputs and outputs, so it has gained wide acceptance through academic and industrial applications [8].

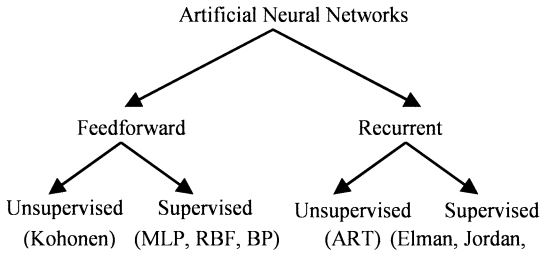


Fig.1 The general class of ANN

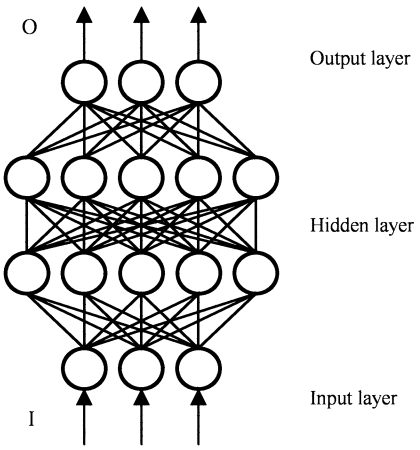


Fig.2 Architecture of three-layered neural network

The unique feature of the back-propagation model is the training procedure, which will be discussed in detail in this section. Through setting training sample, it can effectively optimizes an approximation to a given network target.

The back-propagation algorithm performs two steps of data flow. The first step is that the inputs are ordinarily propagated forward from input to output layer, and then it produces an actual output. The error from the difference between target values and actual values are propagated backward from output layer to the previous layers to update their weights. The connection weights are modified by the difference between the calculated and desired output [9].

We defined that there are m nodes in the input layer, l nodes in the hidden layer, and n nodes in the output layer, $I=(i_1,i_2,\ldots,i_m)^T$, $O=(o_1,o_2,\ldots,o_n)^T$and the matrix of all weight vectors $W=(w_1,w_2,\ldots,w_l)^T$.

The squared error function of training step t is defined as

$$E(t)=\frac{1}{2}\sum_{t=1}^{n}\left(F(net_t)-o_t\right)^2 \tag{1}$$

where $net(t)$ is the activation value of the node, F is the activation function. By using gradient descent method, the weights in the hidden to output connections are updated as

$$\Delta W_{ji}=-\eta\cdot\frac{\partial E}{\partial w_{ji}}(W)=\eta(T_j-O_i)I_i=\eta\delta_j I_i \tag{2}$$

where T_j is the jth component of the target output, O_j is the jth element of the actual output, I_i is the ith element of the input patterns and the constant η is called the learning parameter. ΔW_{ji} is the change to be made to the weight from the ith and jth unit, $\delta_j=T_j-O_j$. Then we define δ_{oi} as the error signal and its double signal, o_i means the ith node in the output layer

$$\delta_{oi} = -\frac{\partial E}{\partial net_i} \tag{3}$$

The error measure between the desired output and actual outputs is used to adjust the weights of the network. The modified weight can be calculated as follows

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_j O_j \tag{4}$$

In order to speed the convergence to an optimal weight assignment, we introduce momentum coefficient which modifies the computation of $\Delta W(t)$ by incorporating a momentum, and the weights are updated as

$$\Delta W(t) = \eta \delta I + \alpha W(t-1) \tag{5}$$

where the arguments t and t-1 are used to indicate the current and the most recent previous training step respectively, and α is the momentum coefficient, $\alpha \in (0, 1)$.

The introduction of momentum coefficient can cut down the times of learning, and efficiently prevent the network trap into the local optimum and at the same time ensure the veracity of the evaluation result.

There is no general prescription for selecting the appropriate learning rate and momentum rate, so success is dependent on a trial and error process. In practical applications, the learning rate is chosen as large as possible without leading to oscillation for fast learning. One way to increase the learning rate without leading to oscillation is to modify the back-propagation learning rule to include a momentum coefficient; the steps of the improved BP algorithm are as follows:

1) Initialize parameters: weights $W(0)$ with small random values, α , η , training sample P.

2) Apply the tth input pattern to the input layer.

3) Propagate the signal forward through the network and compute the actual output.

4) Compute the error value and error signals δ_i for the output layer.

5) Propagate the errors backward to update the weights and compute the error signals δ_i .

6) Check whether the whole set of training data has been cycled once. If t<P, then t=t+1 and go to 2); otherwise go to 7).

7) Check whether the current error is acceptable: if E<E_{max}, terminate the training process and get the final weights; otherwise, E=0, t=1, and initiate the new training epoch by going to 2).

In our experiments, the hidden layer uses sigmoid function, and the output layer uses pure-line function. The training method is adding the momentum rule. The training error is set to 0.001 and the number of maximum learning iteration is set to be 5000. The training is finished when either training error is smaller than 0.001 or the iteration is reached to the maximum iteration number. Fig.3 shows the training error in each step when the BPN is trained using a 256×256 Lena colour image.

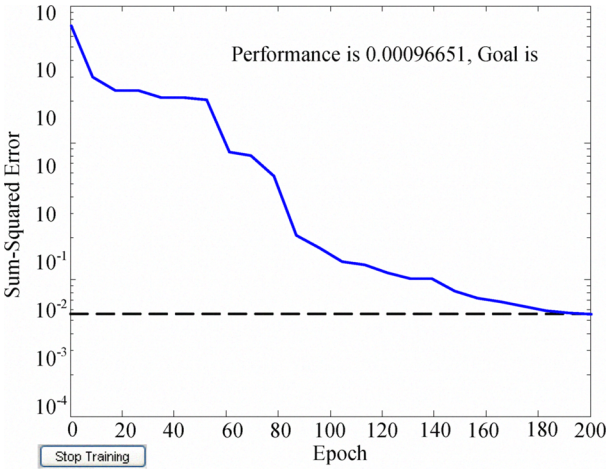


Fig.3 Training curve of BP neural network

III. WATERMARKING EMBEDDING AND EXTRACTION

In this paper, a logo image is embedded as a watermark, the measurements performed with the watermarking algorithm that we have implemented. The embedding parameters are chosen following the description in the paper, at the same time, the watermarks are still invisible and can not be detected by a human observer familiar to the test images.

A. *Watermarking Embedding*

The watermark embedding process transforms the original image into the wavelet domain. The tests are performed on the 256×256 Lena and Baboon images as the original watermark images I. The proposed watermark embedding algorithm is summarized as follows:

1) Divide the original image I into blocks with the size of 8×8 and perform the three-level DWT transformation on each block. Select the position of watermark embedding coefficient using random sequence with a secret key, and quantize the DWT coefficient $C(i, j)$ by Q, and use that value as the input T of BPN.

2) Create the improved BP neural network and initialize its parameters, we use T as supervised signal and the average value of each block is used as the desired output value of the BPN to train BPN. In this experiment, the training error is set to be 0.001.

3) Training improved BP neural network using input and output values, and the watermark is embedded into the wavelet domain using the trained BPN.

4) Perform the inverse discrete wavelet transform (IDWT) on the coefficients where the watermark is embedded to obtain the watermarked image.

B. *Watermarking Extraction*

The watermark extracting procedure is inverse procedure of watermark embedding, the steps are as follows:

1) Perform the three-level DWT transformation on the watermarked image blocks.

2) Select the position of coefficient $C(i, j)$, where watermark is embedded with the same secret key, which is used in watermark embedding sequence. Quantize the DWT

coefficient by Q, and use it as input value of the trained BPN to get the output T'.

3) Extract the watermark according to (6) below using the output T' and coefficient $C(i, j)$

$$W' = \begin{cases} 1 & if \quad C(i,j) > T' \\ 0 & else \end{cases} \tag{6}$$

4) Calculated the correlation between the original watermark and the extracted watermark to detect the existence of the watermark. The quality of watermark extracted from embedded image is measured by the normalized correlation (NC). The NC between the embedded watermark $W(i, j)$ and the extracted watermark $W'(i, j)$ is defined as

$$NC = \frac{\sum_{i=1}^{H}\sum_{j=1}^{L} W(i,j) \times W'(i,j)}{\sum_{i=1}^{H}\sum_{j=1}^{L} [W(i,j)]^2} \tag{7}$$

The image quality is based on the peak signal to noise ratio (PSNR), it is given by

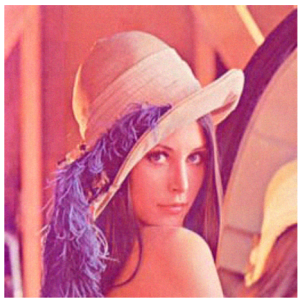
$$PSNR(dB) = 10 \log_{10} \frac{255^2}{MSE} \tag{8}$$

The mean squared error (MSE) will be used

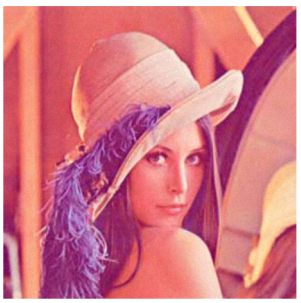
$$MSE = \frac{1}{N} \sum_{i=0}^{N} (I_i' - I_i)^2 \tag{9}$$

IV. Experimental Results

For the entire test results, Matlab version 7.0 is used. A novel digital watermarking technique based on improved BP neural network for color images has been proposed in this paper. In our experiments, we use the 256×256 Lena and Baboon color images as the original images for all the tests shown in Fig.4 (a) and (c), then (b) and (d) are the watermarked images. The watermark is a 56×59 logo image, which is shown in Fig.6 (a).



(a)



(b)

(c)

(d)

Fig.4 (a) (c) Original image, (b) (d) watermarked image

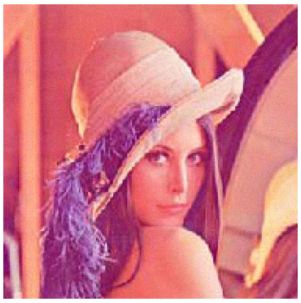
During transmission and distribution of the watermarked image, the watermarking algorithm must be robust enough to withstand not only degradations brought about by unavoidable signal processing operations, geometric distortions, and so on, but also intentional attacks to remove existing watermarks [10, 11]. In succession, we will do some different kinds of attacks experiments, these attacks include JPEG compression, noise addition, filtering, chopping, and rotation.

The watermarked Lena images under the attacks of JPEG compression and Gaussian noise attack are shown in Fig.5, and the extracted watermark from the watermarked Lena image is shown in Fig.6 (b). In order to compare the different types of distortion caused by JPEG compression, we will vary the quality factor in a scale of 20 to 100, 100 correspond to highest quality and least degradation due to compression. The results of the correlation between the embedded and the extracted watermark on a PSNR scale is shown in TABLE I.

TABLE I
NC Values Against JPEG Compression (%)

Quality Factor	Images	
	Lena	Baboon
JPEG Q=20	92.39	94.56
JPEG Q=40	95.41	96.30
JPEG Q=60	96.72	97.41
JPEG Q=80	98.56	99.23
JPEG Q=90	99.34	99.92
JPEG Q=100	100	100

(a)



(b)

Fig.5 (a) Under Gaussian noise attack, (b) under JPEG compression attack

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

Fig.6 (a) Original watermark, (b) extracted watermark from JPEG (Q=60), (c) extracted watermark from Gaussian noise, (d) extracted watermark from chopping.

The embedded watermarks from other attacks are shown in Fig.6 (c) and (d). We can see that the watermarks can be successfully detected after different attacks. The robustness is the property of a watermarking scheme to withstand image distortion. To evaluate the robustness of the proposed watermarking scheme, the traditional BP algorithm is simulated as comparison under the same test conditions. The results against several attacks are shown in TABLE II.

TABLE II
COMPARISON RESULTS AGAINST ATTACKS (MEASURED BY NC %)

Attacks	Traditional BP scheme	The proposed scheme
PSNR[dB]	40.21	42.53
JPEG(Q=60)	94.02	96.72
Guassian noise	92.61	94.37
Filtering	96.57	95.36
Rotation(30°)	92.74	93.02
Chopping	95.31	95.24

The comparison illustrates that the proposed scheme has the better performance of robustness against common attacks than the traditional BP neural network scheme.

V. CONCLUSIONS

In this paper, a watermarking algorithm has been proposed on improved Back-Propagation neural network for color image. BPN model has the ability to learn the characteristics of the image, and the watermark is embedded and extracted by the trained BPN. Updating the weights with adding momentum coefficient to train the BP neural network, we can achieve an optimum approximation to a given network target. We embed the watermark into DWT using the improved BP neural network, which can reduce the error and improve the rate of the learning, and the watermark can be well extracted. Experimental results show that the proposed method has good imperceptibility and high robustness to common image processing such as JPEG compression, noise adding, chopping, and rotation.

REFERENCES

[1] I.J. Cox, J. Kilian, F.T. Leighton, T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions Image Process*, vol.6, no.12, pp.1673-1687, December 1997.

[2] M.D. Swanson, M. Kobayashi, A.H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, vol.86, no.6, pp.1064-1087, June 1998.

[3] Masataka Ejima, Akio Miyazaki, and Taku Saito, "A wavelet-based watermarking for digital images and video," In *proceedings of the IEEE International Conference on Image Processing, ICIP'00,* Vancouver, Canada, September 2000.

[4] J.R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-Domain watermarking techniques for still image: Detector performance analysis and a new structure," *IEEE Trans. Image Processing,* vol.9, pp.55-68, January 2000.

[5] Nadarajah, Saralees, "A comment on asymptotically optimal detection for additive watermarking in the DCT and DWT domains," *IEEE Trans. on Image Processing*, vol.16, pp.1182-1183, 2007.

[6] Joong-Jae Lee, Won Kim, and Na-Yong Lee, "A new incremental watermarking based on dual-tree complex wavelet transform," *The journal of supercomputing*, vol.33, pp.133-140, 2005.

[7] Pao-Ta Yu, Hung-Hsu Ysai, Jyh-Shyan Lin, "Digital watermarking based on neural networks for color images," *Signal Processing*, vol.81, pp.663-671, 2001.

[8] Y.C. Fan, W.L. Mao, H.W. Tsao, "An artificial neural network-based scheme for fragile watermarking," *IEEE Int. Conf. Consumer Electronics*, 2003, pp.210-211.

[9] Jian Zhao, Qin Zhao, Ming-quan Zhou, and Jian-shou Pan, "A Novel Wavelet Image Watermarking scheme Combined with Chaos Sequence and Neural Network, Advances in Neural Networks," *Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, vol.3174, pp.663-668, 2004.

[10] Tefas, A. Nikolaidis, N. Solachidis, V. Tsekeridou, S. Pitas, "Performance Analysis of Correlation Based Watermarking Schemes Employing Markov Chaotic Sequences, *IEEE Trans. signal processing*, vol.51, no.7, pp.1979-1994, 2003.

[11] H. Alexandre, Paquet, K. Rabab, Ward, and Ioannis Pitas, "Wavelet packets-based digital watermarking for image verification and authentication," *Signal Processing*, vol.83, no.10, pp.2117-2132, 2003.