Robust Digital Image Watermarking Algorithm Using BPN Neural Networks

Cheng-Ri Piao, Wei-zhong Fan, Dong-Min Woo, and Seung-Soo Han

Department of Information Engineering & NPT Center, Myongji University,
Yongin, Kyunggi, 449-728, South Korea
shan@mju.ac.kr

Abstract. This paper proposes a new watermarking scheme in which a logo watermark is embedded into the spatial domain of the original image using Back-Propagation neural networks (BPN). BPN will learn the characteristic of the image, and then watermark is embedded and extracted by the trained BPN. The image is divided into 8×8 blocks and the average pixel value of each block is used as the desired output value of the BPN. The quantized DC coefficient of discrete cosine transform (DCT) domain of each block is used as input value of the BPN to be trained. After the BPN is trained using those input/output values, watermark is embedded into the spatial domain using the trained BPN. The trained BPN also used in watermark extracting process. Experimental results show that the proposed method has good imperceptibility and high robustness to common image processing.

1 Introduction

With the rapid development of computer and communication networks, the digital multimedia reproduction and distribution are becoming extremely easier and faster. However, these advances also afford unprecedented opportunities to pirate copyrighted digital multimedia products. As a result, the watermarking technique, which embeds a watermark into digital multimedia contents for detecting and tracing copyright violations, has recently become a very active area of multimedia security [1]. The watermarking techniques can be classified into two classes depending on the domain of watermark embedding, i.e. a spatial domain and a frequency domain. Among the spatial domain watermark embedding methods, Schyndel *et al*. proposed a watermark embedding technique by changing the least significant bit of some pixels in an image [2]. Bender *et al*. described a watermarking approach by modifying a statistical property of an image called 'patchwork' [3]. On the other hand, there are many algorithms for watermark embedding in frequency domain. Cox *et al*. described a method where the watermark is embedded into the large discrete cosine transform (DCT) coefficients using ideas borrowed from spread spectrum in communications [4]. Xia *et al*. proposed a frequency domain method of embedding the watermark at all the subbands except LL subband, using discrete wavelet transform (DWT)[5].

J. Wang et al. (Eds.): ISNN 2006, LNCS 3973, pp. 285 – 292, 2006.
© Springer-Verlag Berlin Heidelberg 2006

Recently, quantization index modulation (QM) [6-8] technique is widely used in watermarking area, and this technique is very robust to various attacks such as JPEG compression, noise insertion, image resize and so on. And Mei, *et. al*. [9] proposed a watermarking technique using neural network. It is a very robust method, but a non-blind method that requires original image to extract the watermark.

In this paper, a new watermark embedding/extracting algorithm using error back-propagation neural network (BPN) is introduced. In this algorithm, an image is divided into 8×8 blocks, then calculate the average pixel value for each block, make it the desired output for BPN, and start the BPN learning procedure. After that, embed and extract the watermark by the trained BPN.

The experimental results show that the watermarked image has at least 43 dB in peak signal-to-noise ratio (PSNR). Also, the results are compared among the presented scheme, the method that uses quantization index modulation (QM) technique on the DC coefficient of DCT domain, and the neural network method applied on DCT domain [9].

2 Related Theories

2.1 The Error Back-Propagation Neural Network (BPN)

The BPN is a kind of supervised learning neural network. It is one of the most frequently used learning techniques in neural networks. The principle behind the BPN involves using the steepest gradient descent method to reach a small approximation. A general model of the BPN has an architecture like that depicted in Fig. 1. There are three layers including input layer, hidden layer, and output layer. Two nodes of each adjacent layer are directly connected to one another, which is called a link. Each link has a weighted value, which represents the relational degree between two nodes. A training process described by the following equations updates these weighted values:

$$\begin{aligned} net_j(t) &= \sum_i \alpha_{i,j} o_i(t) - \theta_j \\ o_j(t+1) &= f_{act}(net_j(t)) \end{aligned} \tag{1}$$

where $net_j(t)$is the activation value of the j^{th} node in iteration t, $o_j(t+1)$is output of the j^{th} node in iteration $t+1$, $f_{act}(x)$ is called the activation function of a node, which usually is a sigmoid function in hidden layers and a pureline function in output layer. Generally, all initial weight values $\alpha_{i,j}$ are assigned using random values. In each iteration process, all $\alpha_{i,j}$ are modified using the delta rule according to the learning samples. The trained neural network can memorize the characteristics of the learning samples, and predict a new output due to the adaptive capability of it. BPN will be used to learn the characteristics of image for improving the performance of the proposed watermarking scheme in the section 3.

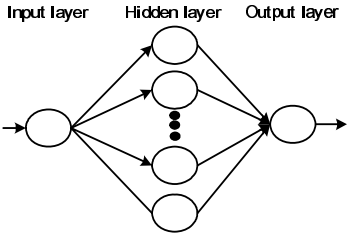

Fig. 1. Architecture of a BPN

2.2 The BPN Neural Network Training Procedures

The BPN neural network training procedures are as Fig. 2 below.

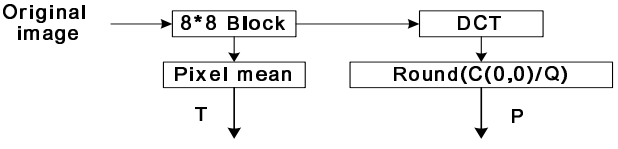

Fig. 2. BPN training procedures

In the figure, *C(0,0)* is the DC coefficient of the discrete cosine transform (DCT) domain of the original image, and *Q* is the quantization value. The output of the equation *Round(C(0,0)/Q)*, which is represented as *P*, is used as an input value for the BPN, and *T*, which is the average pixel value of each block, is used as a desired output value for the BPN neural network. The structure of BPN network presented in the paper is 1-120-1, which is obtained as an optimal neural network structure from many experiments. The hidden layer uses sigmoid function, and the output layer uses pureline function. The training method is Levenberg-Marquardt rule. And the training error is set to 0.01 and the number of maximum learning iteration is set to 5000.

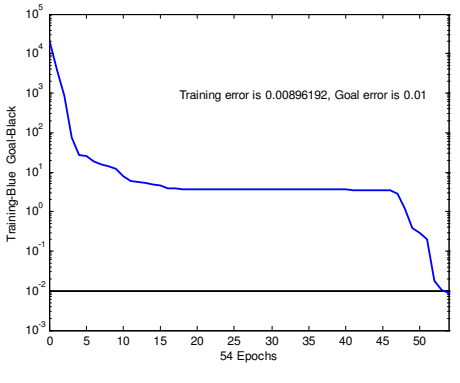

Fig. 3. The training process of BNP

The training is finished when either training error is smaller than 0.01 or the iteration is reached to the maximum iteration number. Fig. 3 shows the learning error in each step when the BPN is trained using a 256×256 size Lena image. The BPN trained by this way will be used to embed and extract the watermark.

3 Watermark Embedding and Extracting

Generally, a watermark algorithm includes 3 steps: watermark generation, embedding, and extraction. In this paper, a logo image, which can be easily distinguished by human eyes, is embedded as a watermark. The proposed watermark technique is a kind of blind watermarking algorithm, which embed/extract the watermark into/from the spatial domain.

3.1 Watermark Embedding

Fig. 4 below is the block diagram of the watermark embedding procedure. The trained BPN in the figure is discussed in section 2.2. Q is the quantization value.

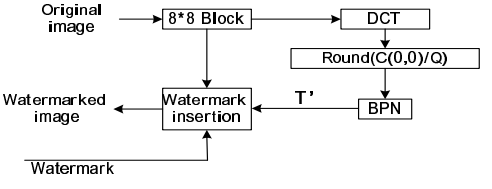

Fig. 4. Watermark embedding procedure

The watermark embedding procedures are as follows:

Step1: Divide the original image to be 8×8 blocks and make DCT transformation.
Step2: Quantize the DC coefficient $C(0,0)$ of the DCT by Q, and use that value as the input value of BPN then get the output T'.
Step3: Embed the watermark according to the equation 2 and 3 below which use the output value T' and the pixel values of the corresponding block.

$$\Delta P_{blcok\,1} = \frac{1}{4} * 8 * Q + 64 * T' - \sum_{i,j=0}^{7} p_{i,j}$$
$$\Delta P_{blcok\,0} = -\frac{1}{4} * 8 * Q + 64 * T' - \sum_{i,j=0}^{7} p_{i,j} \tag{2}$$

$$p'_{i,j} = \begin{cases} p_{i,j} + round\ (\Delta P_{blcok\ 1} / \sum_{i,j}^{7} p_{i,j}) & if \quad w_{m,n} = 1 \\ p_{i,j} + round\ (\Delta P_{blcok\ 0} / \sum_{i,j}^{7} p_{i,j}) & if \quad w_{m,n} = 0 \end{cases} \tag{3}$$

Where $p_{i,j}$ is the pixel value of the original image, $p'_{i,j}$ is the pixel value in which watermark is embedded. w is the watermark and Q is the quantization value. ΔP_{blok1} is the change of sum of all pixels in a block when watermark is 1, and ΔP_{blok0} is the change of sum of all pixels in a block when watermark is 0. $\sum_{i,j=0}^{7} p_{i,j}$ represents the sum of pixel value in a block, and $64\times T'$ is also the sum of pixel value of a block, but calculated by BPN. A change $\frac{1}{4}*8*Q$ should be given to each block because $Round(C(0,0)/Q)$ must have no change after embedding a watermark, so that there will be no change in $Round(C(0,0)/Q)$ when extracting the watermark. With many experiments, we concluded that the embedded watermark is most robust to attacks when the change is $\frac{1}{4}*8*Q$.

3.2 Watermark Extracting

The watermark extracting procedures are the converse procedures of watermark embedding, as shown in Fig.5. The BPN here is trained neural network, which is discussed in section 2.2, and Q is the quantization value.

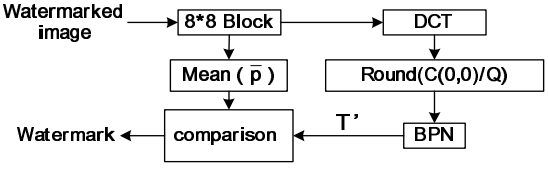

Fig. 5. Watermark extraction procedure

The watermark extraction procedures are as follows:

Step1: Divide the watermarked image into 8×8 blocks and calculate the average pixel value $\overline{p}$ for each block, and make a DCT transformation.

Step2: Quantize the DC coefficient $C(0,0)$ of the DCT by Q, and use it as input value of the BPN to get the output T'.

Step3: Extract the watermark using the equation 4 below, which is in use of output T' and $\overline{p}$. w' of equation 4 is extracted watermark.

$$w'_{m,n} = \begin{cases} 1 & if \quad \overline{p} > T' \\ 0 & else \end{cases} \tag{4}$$

Step 4. The correlation between the original watermark and the extracted watermark is calculated to detect the existence of the watermark. The similarity between the original watermark w and the extracted watermark w' is quantitatively measured by the bit correlation ratio (BCR), defined as follows:

$$BCR = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\overline{(w_{i,j} \otimes w'_{i,j})}}{\sum_{i=1}^{m}\sum_{j=1}^{n}\overline{(w_{i,j} \otimes w_{i,j})}} \times 100\% \tag{5}$$

Where $w_{i,j}$ is the original watermark bit, $w'_{i,j}$ is the extracted watermark bit, and $\otimes$ is the exclusive OR.

4 Experimental Results

In this paper, 8 bit 256×256 sized Lena, cameraman, baboon, boat images are used as cover (original) images, and a 32×32 sized logo image, which can be easily distinguished by human eyes, is used as watermark. Fig.6 shows the images used in this experiment and the watermark in use.

Fig.7 shows the relationship between the embedding quantization step-size Q and the peak signal-to-noise ratio (PSNR). We can see that the PSNR of the watermarked image is decreasing with increasing Q value, but in any case the PSNR is more than 43dB. This shows that the watermarked image has a good PSNR.

We tested the robustness of the proposed method with several typical images attacked by JPEG compression. The watermarks extracted from JPEG compressed versions of the watermarked image with various compression quality factors, and

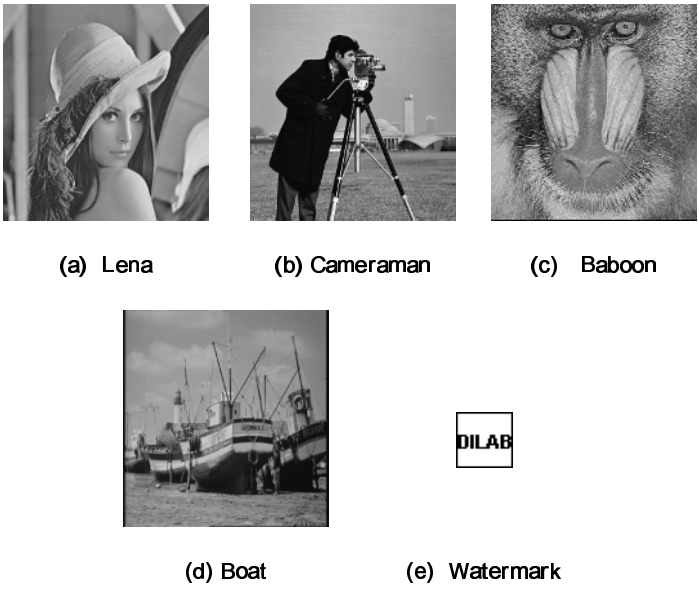

Fig. 6. Experiment images and watermark

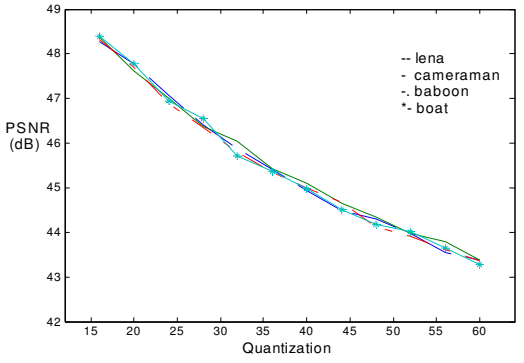

Fig. 7. The relationship between Q and PSNR

Table 1. BCR values of the extracted watermarks (%)

Images / PSNR / Quality Factor	Lena	Baboon	Cameraman	Boat
	44.48 [dB]	44.65 [dB]	44.64 [dB]	44.51 [dB]
100	100	100	100	100
90	100	100	100	100
80	99.99	100	99.70	100
70	100	99.80	99.21	99.99
60	99.95	99.60	98.53	100
50	99.68	98.53	97.36	99.80
40	96.68	97.06	95.87	96.09

Table 2. Comparison(BCR) results among QMs' method, nerual netwrok method described in [9] , and presented method

	QM method	Method [9]	The proposed method
PSNR [dB]	44.16	44.12	44.53
Embedded	100	100	100
JPEG (Quality factor = 40)	89.45	88.04	96.42
Noise (Gaussian)	87.50	85.86	89.04
Resize (225×225)	82.71	84.92	86.23

their corresponding BCR values are listed in Table 1. According to Table 1, we can see that the extracted watermark is still recognizable when the compression quality factor reaches 40, which means the proposed method has good robustness to JPEG compression.

Finally, we compared the performance of the proposed method with that of the one proposed in QM method, and with the performance of the scheme using neural network on DCT domain. All the three methods have the same test conditions

including the same test image "Lena" (256×256), the same amount of embedded information (1024 bits i.e. a 32×32 binary pattern watermark). Comparison results are listed in Table 2. According to this table, the BCR of the extracted watermarks using the proposed method are always higher than the others. These results prove that the proposed method has superior performance.

5 Conclusions

A new watermarking method for image has been proposed. It has following characteristics: First, a logo watermark is embedded into the spatial domain of the image using Back-Propagation neural networks (BPN). The embedding scheme can result in good quality of the watermarked image. Second, a BPN model is used to learn the characteristics of image. Due to the learning and adaptive capabilities of the BPN, the embedding /extracting strategy can greatly improve the robustness to several attacks. Experimental results illustrate that the performance of the proposed technique is superior to that of the similar one in the literature.

Acknowledgement

This work was supported by the ERC program of MOST/KOSEF (Next-generation Power Technology Center).

References

1. Hartung F., Kutter, M.: Mulimedia Watermarking Techniques. Proceedings of the IEEE (1999) 1079-1094
2. Schyndel, R.G., Tirkel, A.Z., Osborne, C.F.: A Digital Watermarking. Int. Conf. on Image Processing, **2** (1994) 86-90
3. Bendor, W., Gruhl, D., Morimoto, N., Lu, A.: Techniques for Data Hiding. IBM Systems Journal **3&4** (1996) 313-336
4. Cox, I.J., Kilian, J., Leighton, T., Shamoon, T.: Secure Spread Spectrum Watermarking for Multimedia. IEEE Trans. on Image Processing **12** (1997) 1673-1687
5. Xia, X., Boncelet, C.G., Arce, G. R.: A Multiresolution Watermark for Digital Images. Proc. IEEE ICIP **3** (1997) 548-551
6. Chen, B., Wornell, G.W.: Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding. Information Theory, IEEE Transactions on, **4**, (2001)1423 - 1443
7. Jang, Y., Kim, I., Kang, H., Kim, K., Han, S.: Blind Watermarking Algorithm Using Complex Block Selection Method. Lecture Notes in computer science, Vol. **2195.** Springer (2001) 996-1001
8. Inoue, H., Miyazaki, A., Araki, T., Katsura, T.: A Digital Watermark Method Using the Wavelet Transform for Video Data. IEICE Trans. Fundamentals. **E83-A** (2000) 90-95
9. Mei, S., Li, R., Dang, H., Wang, Y.: Decision of Image Watermarking Strength Based on Artificial Neural Networks. Neural Information Processing, ICONIP '02. Proceedings of the 9th International Conference on, Vol. 5. (2002) 2430 - 2434