J. Vis. Commun. Image R. 32 (2015) 130–143

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Blind image watermarking via exploitation of inter-block prediction and visibility threshold in DCT domain ☆

Ling-Yuan Hsu [a], Hwai-Tsu Hu [b,*]

[a] Department of Information Management, Junior College of Medicine, Nursing and Management, I-Lan, Taiwan
[b] Department of Electronic Engineering, National I-Lan University, I-Lan, Taiwan

A R T I C L E I N F O

Article history:
Received 12 January 2015
Accepted 30 July 2015
Available online 7 August 2015

Keywords:
Blind image watermarking
Inter-block prediction
Discrete cosine transform
Backward-propagation neural network
Just-noticeable difference
Relative modulation
Robustness
Imperceptibility

A B S T R A C T

In this paper, the backward-propagation neural network (BPNN) technique and just-noticeable difference (JND) model are incorporated into a block-wise discrete cosine transform (DCT)-based scheme to achieve effective blind image watermarking. To form a block structure in the DCT domain, we partition a host image into non-overlapped blocks of size 8×8 and then apply DCT to each block separately. By referring to certain DCT coefficients over a 3×3 grid of blocks, the BPNN can offer adequate predictions of designated coefficients inside the central block. The watermarking turns out to be a process of adjusting the relationship between the intended coefficients and their BPNN predictions subject to the JND. Experimental results show that the proposed scheme is able to withstand a variety of image processing attacks. Compared with two other schemes that also utilize inter-block correlations, the proposed one apparently exhibits superior robustness and imperceptibility under the same payload capacity.

1. Introduction

With the rapid development of multimedia and Internet technology, it is quite common for people to make collections of multimedia (e.g., images, audios and videos) stored on Internet-attached servers and accessible to a vast number of others through the Internet. Furthermore, due to the availability of convenient software tools, tampering such multimedia data has become very easy. As a solution to this issue, digital watermarking is a promising measure to prevent intellectual property infringement.

Digital watermarking has attracted considerable attention because of numerous potential applications such as content authentication, copyright protection and secret communication. In general, a digital image watermarking technique must satisfy five requirements including imperceptibility, robustness, security, capacity and non-detectability [1,2]. First, the embedded data (called watermark) are supposed not to yield perceptible distortion to the host image. Second, the watermark needs to be robust against common signal processing attacks. Third, it ought to be difficult for an attacker to access embedded data. Forth, the payload capacity has to be high enough to contain sufficient information. And fifth, it shall be undeletable by other algorithms. Apart from the above five properties, blind watermarking imposes an extra challenge that watermark retrieval is performed without the presence of original data.

Blind image watermark technology can be mainly classified into spatial and transform domains. Spatial domain watermarking is usually done by directly adjusting image pixels in accordance with the watermark. Methods such as least significant bit (LSB) [3] and vector quantization (VQ) [4] belong to this category. Spatial domain methods have the advantages of low complexity and easy implementation but suffer the disadvantage of weak resistance against malicious attacks. In contrast, the watermarks embedded by transform domain methods generally show better robustness and imperceptibility, but the required computation is relatively high as compared to that demanded by spatial domain methods. The merits of transform domain methods consist in the capability of converting spatial data to a representation more compatible to human perception. Commonly used transforms for digital watermarking are discrete wavelet transform (DWT) [5–10], discrete cosine transform (DCT) [11–16], discrete Fourier transform (DFT) [9,17,18], and singular value decomposition (SVD) [19–24], etc. Among these transforms, the DCT holds the advantage of excellent energy compaction for highly correlated image data. Imposing invisibility constraints is comparatively easy when working in the DCT domain [25].

There are generally two ways to embed binary watermarks into images. One is to map the chosen coefficients to a designated range

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author. Fax: +886 3 9369507.
E-mail address: hthu@mail.niu.edu.tw (H.-T. Hu).

http://dx.doi.org/10.1016/j.jvcir.2015.07.017
1047-3203/

according to the binary values. The quantization index modulation (QIM) proposed by Chen and Wornell [26] is the most common technique. The other way is to manipulate the coefficients in paired groups so that each binary bit is manifested by the relationship of the paired groups [15,21,27–31]. While using two antithetic relationships to characterize binary information, further improvement may be gained by exploiting the correlation between chosen coefficients. In [15], Wang and Pearmain proposed a data hiding technique via self-reference of specific coefficients in DCT domain. They divided a host image into blocks of size 8×8 pixels, each of which was converted to a DCT representation. The DC coefficients in a 3×3 grid of DCT blocks were formulated to render an estimate of an AC coefficient in the central block. The corresponding AC coefficient magnitude was adaptively increased or decreased depending on whether the watermark bit was '0' or '1'. The attempt of using adaptive AC coefficient prediction for blind image watermarking can also be found later in [30,31]. This kind of reference schemes had the advantage of efficiency, but they suffered the drawback that the image quality might be considerably distorted due to insufficient estimation accuracy. In [32], Das et al. attempted to explore the correlation between DCT coefficients drawn from adjacent DCT blocks. They picked a pair of DCT coefficients from two neighboring blocks and adjusted one coefficient according to the other. One direct advantage of this approach lay in the improvement of imperceptibility as the DCT coefficients were only moderately modified. A probable disadvantage was that the embedded watermark might still not withstand severe attacks.

In this study, we consider using artificial neural networks (ANN) to further exploit the coefficient correlations so that binary information can be better described by distinct criteria. An ANN is a type of computing that emulates human brain functions. The basic architecture of an ANN comprises layers of neurons linked through weights and biases. The well-known back-propagation method is employed to train the neural networks. By giving the ANN a set of training data, the biases and weights will be tuned correspondingly to attain a best match between inputs to desired outputs [33]. It is our belief that, with the assistance of back-propagation neural network (BPNN), we may search out suitable relations among the DCT coefficients in different locations to carry out effective watermarking. Moreover, since many perceptual properties can be explored in DCT domain, we intend to use an adequate just-noticeable difference (JND) model to guide image watermarking. The use of a JND model can not only remove the redundancy due to statistical correlation but ensure perceptual quality of watermarked images [34].

The remaining sections of this paper are organized as follows. Following this introduction, Section 2 discusses relevant techniques. The proposed blind image watermarking scheme is subsequently explicated in Section 3. Section 4 presents our experimental results. Finally, Section 5 summarizes the contributions of this paper.

2. Relevant techniques

Three techniques serve as the groundwork while developing our watermarking scheme.

2.1. Discrete cosine transform (DCT)

DCT has owned considerable attention in various signal processing applications including image coding and watermarking. The main idea behind the DCT is to transform an image from the spatial domain to another presentation in the frequency domain. By exploring the energy compactness property of DCT, an image can be reconstructed from very few DCT coefficients with adequate accuracy. In general, DCT-based watermarking could be implemented using spatially local or global transforms. A commonly used block size for DCT watermarking is a square matrix of 8×8 as shown in Fig. 1. This is the same size adopted in the JPEG compression standard. The two-dimensional DCT with respect to a matrix of $N \times N$ can be expressed as follows [35]:

$$F^{m,n} = \frac{1}{\sqrt{2N}}\phi(m)\phi(n)\sum_{u=0}^{N-1}\sum_{v=0}^{N-1} f^{u,v}\cos\left[\frac{(2u+1)m\pi}{2N}\right]\cos\left[\frac{(2v+1)n\pi}{2N}\right]; \quad (1)$$

$$f^{u,v} = \frac{1}{\sqrt{2N}}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\phi(m)\phi(n)F^{m,n}\cos\left[\frac{(2u+1)m\pi}{2N}\right]\cos\left[\frac{(2v+1)n\pi}{2N}\right], \quad (2)$$

where $f^{u,v}$ denotes the $(u,v)^{th}$ image pixel value in the spatial domain and $F^{m,n}$ represents the $(m,n)^{th}$ DCT coefficient in the frequency domain. $\phi(k)$ is a multiplication factor given as

$$\phi(k) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{if } k = 0; \\ 1, & \text{if } k > 0. \end{cases} \quad (3)$$

2.2. Back-propagation neural network (BPNN)

A BPNN algorithm is an approach intended to minimize an error function in weight space using gradient descent. The combination of weights is considered the solution of a learning problem. A BPNN process normally comprises three parts: (1) a forward propagation, (2) an error calculation, and (3) a weight update. A training pattern is propagated through the neural network to produce activation signals. The input and output of each processing unit (or "neuron") was expressed by Eqs. (4) and (5), through which the summation $\varphi_{\hat{j}}$ is taken as the input of an activation function that uses the sigmoidal transformation.

$$\varphi_{\hat{j}} = \sum_{\hat{i}}\varpi_{\hat{i}\hat{j}}x_{\hat{i}} + \theta_{\hat{j}}; \quad (4)$$

$$y_{\hat{j}} = f(\varphi_{\hat{j}}) = \frac{1}{1+e^{-\varphi_{\hat{j}}}}, \quad (5)$$

where $\varphi_{\hat{j}}$ represents output of unit $\hat{j}$, $\varpi_{\hat{i},\hat{j}}$ represents the weight on connection from unit $\hat{i}$ to $\hat{j}$, $x_{\hat{i}}$ represents the input to unit $\hat{j}$ from unit $\hat{i}$ and $\theta_{\hat{j}}$ is the bias on unit $\hat{j}$. Since the BPNN uses the gradient descent method, one needs to calculate the derivative of the squared error function with respect to the weights of the network. Assuming one output neuron, the squared error E is calculated as

$$E = \frac{1}{2}\sum_{\hat{k}}(D_{\hat{k}} - O_{\hat{k}})^2, \quad (6)$$

where $D_{\hat{k}}$ and $O_{\hat{k}}$ respectively denote the target and actual outputs for the $\hat{k}^{th}$ training sample. The weight update is carried out through

$$\varpi_{\hat{j},\hat{k}} = \varpi_{\hat{j},\hat{k}} - \eta\frac{\partial E}{\partial \varpi_{\hat{j},\hat{k}}}, \quad (7)$$

where η represents a learning rate.

2.3. Just-noticeable difference (JND)

JND refers to the minimum perceptible threshold when visual contents are altered. It is the result deduced from the psychophysical and physiological characteristics of the human visual system (HVS). The DCT-JND is a well-cited scheme for DCT thresholds