TRANSPARENT ROBUST INFORMATION HIDING FOR OWNERSHIP VERIFICATION

Dan Yu, Student Member, IEEE, Farook Sattar, Member, IEEE and Sirajudeen Gulam Razul
School of Electrical and Electronic Engineering
Nanyang Technological University, Nanyang Avenue, Singapore 639798

ABSTRACT

For copyright protection, the robustness of a watermarking scheme against various attacks is an essential requirement. Many proposed robust watermarking schemes may achieve good robustness but sacrifice the good quality of the watermarked image. This paper, therefore, proposes a transparent robust watermarking scheme, which embeds the watermark (or the secret information) adaptively in the Discrete-Cosine Transform (DCT) domain. The proposed scheme is blind as the original image is not needed, and only a key containing the watermark locations and the scaling factors is required for watermark extraction. This adaptive replacement embedding technique can guarantee in preserving the good visual quality of a watermarked image. Comparing with other existing DCT-domain watermarking methods, the proposed watermarking method can achieve higher robustness performance, while retaining better quality of the watermarked image in terms of PSNR.

1. INTRODUCTION

Digital watermarking technology has evolved rapidly in recent years. Watermarking technique [1] is to hide secret information into the cover media which is able to be retrieved later. The most popular application of watermarking is to solve the problem of ownership and content authentication of digital media (e.g., audio, image, and video). An effective data hiding system must balance the requirements of three parties: *imperceptibility* – *robustness* – *capacity* [1]. *Imperceptibility* requires the marked data and the original data should be perceptually indistinguishable. *Robustness* requires that the embedded information should be reliably detectable or retrievable when the marked data is altered. *Capacity* refers to the amount of the information that is being embedded into the host cover data. If the amount of information to be embedded is decided, there always exists a trade-off between the visual quality of the marked data and robustness of the watermark. In general, the higher the embedding strength, the better robustness a watermarking system can achieve, however, at the same time it may result a poorer visual quality of the marked image.

In an ownership verification system for copyright protection, the robustness of the watermarking scheme against attacks is the most important requirement. Systematic benchmarking, such as, Stirmark [2], provides a common criteria for robustness performance evaluation against attacks to show which techniques work better than others. However, the evaluation should also include distortion measurement of the watermarked data introduced during watermark embedding, for a fair performance evaluation.

Thus the objective of this paper is to propose a new *blind robust* DCT-based information hiding scheme to further improve the performance in terms of the robustness and the imperceptibility, so that it is more efficient for ownership verification to enforce copyright protection.

2. PROPOSED TRANSPARENT ROBUST WATERMARKING SCHEME

2.1. Proposed Adaptive Watermark Embedding Scheme

The proposed embedding technique is to use *adaptive replacement* operation rather than *additive* modification. That is, to replace the DCT coefficients of the original image adaptively by the DCT coefficients of a watermark. The grayscale images such as some pattern, logo and signature images are adopted as watermarks. Suppose we have two grayscale images which have similar energy distribution of the DCT coefficients. The mismatch between the DCT coefficient of the original image and that of the watermark image can be very low for the nearest match. Thus a grayscale watermark image can be embedded more transparently. This type of watermarks can also provide unique ownership verification and can be easily recognized visually rather than by using an objective similarity measure.

Let $\mathbf{I}$ be the original image of size $N1 \times N2$, and $\mathbf{W}$ be the grayscale watermark image of size $M1 \times M2$. Figs. 1(a) and 1(b) give an original Lena image and a grayscale text image, respectively, as an illustration. The proposed watermark embedding procedure includes the following steps:

Step 1 : Perform 2-D whole DCT for the original image $\mathbf{I}$ and the watermark $\mathbf{W}$ to obtain the 2-D DCT coefficients $\mathbf{C_I}$ and $\mathbf{C_W}$, respectively.

Step 2 : The zig-zag scanned sequence of $\mathbf{C_I}$ is denoted as S_I. The coefficients within the range – [$Initial$, End], where $Initial$ denotes the starting index and End is the ending index, are chosen for embedding. The watermark's DCT coefficient sequence, S_W, is obtained by column-wise scanning of $\mathbf{C_W}$.

Step 3 : The embedding algorithm searches for the nearest match of each watermark coefficient $S_W(i)$ in the

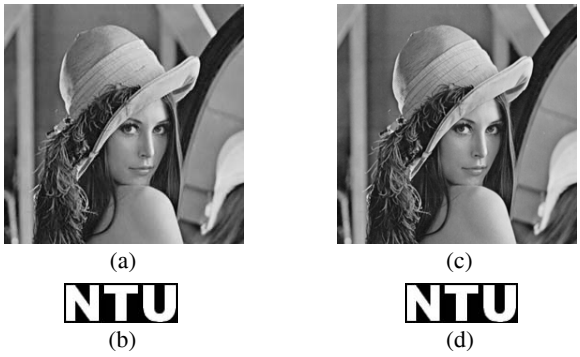


Fig. 1. (a) The original Lena image (256×256 pixels), (b) the grayscale watermark (38×111 pixels, 'NTU' is the abbreviation form of 'Nanyang Technological University') to be embedded, (c) the watermarked Lena image (PSNR=61.00dB) and (d) the extracted watermark ($r \doteq 1.0000$).

selected embedding range, in terms of their absolute amplitudes. To achieve good robustness, low to middle frequency components (i.e., more significant coefficients) are chosen for embedding. The detailed adaptive watermark embedding algorithm for the *i*-th component of the watermark's coefficients, $S_W(i)$, is shown in Fig. 2.

The location of the nearest match of S_I is stored as a vector – $location(i)$. Let define the difference, d, between the two nearest matches as

$$d(i) = ||S_I(location(i))| - |S_W(i)||, i = 1, 2, ..., M1 \times M2, \quad (1)$$

where $S_I(location(i))$ is the DCT coefficient of the original image which is the nearest match of the *i*-th watermark coefficient. A threshold T_0 is set to control the quality of the watermarked image. If the difference d is smaller than T_0, $S_I(location(i))$ is replaced by the watermark coefficient $S_W(i)$ directly. Otherwise the coefficient $S_I(i)$ is kept unchanged; instead a scaling factor, $scale(i)$, between the $S_I(location(i))$ and $S_W(i)$ is generated for a perfect watermark recovery in the watermark extraction. Mathematically, the DCT sequence of the watermarked image, $S_X(j)$ (where $\mathbf{X}$ denotes the watermarked image and $j = 1, 2, ..., N1 \times N2$), is expressed as

$$S_X(j) = \begin{cases} \text{sign}\{S_I(j)\} \cdot |S_W(j)| & \text{if } j \in location(i), \\ & i = 1, 2, ..., M1 \times M2 \\ & \text{and } d(j) < T_0 \\ S_I(j) & \text{else} \end{cases}; \quad (2)$$

and the scaling factor – *scale* – is defined as follows:

$$scale(i) = \begin{cases} 1 & \text{if } d(i) < T_0 \text{ and} \\ & S_I(location(i))S_W(i) \geq 0 \\ -1 & \text{if } d(i) < T_0 \text{ and} \\ & S_I(location(i))S_W(i) < 0 \\ \frac{S_I(location(i))}{S_W(i)} & \text{if } d(i) \geq T_0 \end{cases}. \quad (3)$$

The scaling factor is either +1 or -1 if the difference d is smaller than the threshold T_0, depending on whether the two nearest matched candidates are of the same sign. If the candidates are of the same sign, $scale$ equals 1; otherwise, $scale$ equals -1. Only when d is larger than T_0, the scaling factor equals the ratio of these two candidates for the nearest match. Therefore, T_0 is employed as a control parameter for watermarked image quality, and *scale* is an important refinement parameter for good watermark retrieval.

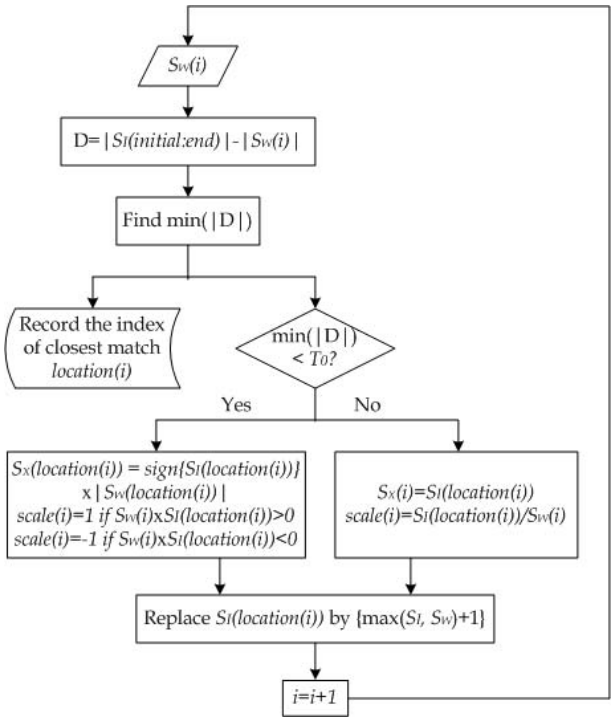


Fig. 2. The adaptive watermark embedding algorithm for the *i*-th component of the watermark DCT coefficient.

Step 4 : The dual-key of the proposed watermarking system includes the watermark embedding locations – *location* and the scaling factors – *scale*. Note that the dual key of the algorithm is image-dependent, which implicates only the person, who knows the key, is able to do the ownership verification for the particular image content.

Step 5 : The watermarked image $\mathbf{X}$ is obtained by an inverse zig-zag scanning of the embedded sequence S_X followed by a 2-D inverse discrete cosine transform.

2.1.1. Embedding Range Selection

One would like to hide the watermark into the coefficients that are less sensitive to JPEG compressions. The embedding range is then selected after investigating the JPEG compression effects over the DCT coefficients of the original image. In particular, JPEG compression with quality factor 60% is used as a reference. The sensitivity of DCT coefficients against compression is measured as the relative change between the compressed coefficient $C_{compressed}$ and its original $C_{original}$, RC, defined by

$$\text{RC} = \frac{C_{compressed} - C_{original}}{C_{original}}. \quad (4)$$

As shown in Fig. 3, the frequency components of Lena image – roughly the first 8,000 most significant coefficients,

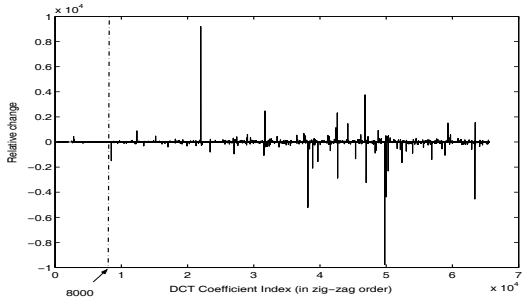


Fig. 3. The range selection for Lena image.

are more resistive against compression. Therefore, the embedding range for Lena image is set between $Initial$ index of 101 and End index of 8,000. Note that $Initial$ index is set as 101, because one does not want to disturb those most significant coefficients of the original image (the first 100 coefficients in our case) to avoid the easily perceived modifications of the watermarked image.

2.1.2. Threshold Setting

The criterion to set the threshold T_0 is by considering the total distortions allowed in a watermarked image. Generally, to make the watermarked image visually indistinguishable from the original image $\mathbf{I}$, the PSNR of the watermarked image $\mathbf{X}$ needs to be maintained above 40dB [3]. The requirement of threshold setting T_0 can be derived from the following relationships:

$$\begin{aligned} \text{PSNR} &= 10\log\left(\frac{255^2 \times Total\ image\ pixel\ no.}{\Delta E}\right) \geq 40\text{dB}; \\ \text{and } T_0 &\leq \sqrt{\frac{\Delta E}{Total\ no.\ of\ watermark\ samples}}, \end{aligned} \tag{5}$$

where ΔE is the total energy difference allowed between the watermarked image and the original image. For the given embedding example, the threshold T_0 is obtained as

$$\begin{aligned} T_0 &\leq \sqrt{\frac{\Delta E}{Total\ no.\ of\ watermark\ samples}} \\ &= \sqrt{\frac{255^2 \times 256^2}{10^4 \times (38 \times 111)}} \\ &\doteq 10.0514. \end{aligned} \tag{6}$$

Fig. 1(c) shows a watermarked Lena image (PSNR= 61.00 dB) by using T_0=10.0514.

2.2. Proposed Blind Watermark Extraction Scheme

The watermark extraction does not require the original image, and only the key containing the watermark embedding location and scaling factor is needed. The watermark extraction process consists of the following steps:

Step 1 : Perform 2-D DCT transform for the received watermarked image $\mathbf{X}'$, which may or may not be the same as $\mathbf{X}$. The received watermarked image could be altered during transmission through the channel or manipulated deliberately due to attacks.

Step 2 : The DCT coefficients, $C_X{}'$, are then re-arranged into a vector sequence of $S_X{}'$ by zig-zag scanning.

Step 3 : Based on the key information of the watermark embedding locations (in vector *location*) and scaling factors (in vector *scale*), the vector form of the DCT coefficients of the watermark, $S_W{}'$, can be retrieved by

$$S_W{}'(i) = \frac{S_X{}'(location(i))}{scale(i)}, i = 1, 2, ..., M1 \times M2. \tag{7}$$

Step 4 : The watermark's DCT coefficients, $C_W{}'$, in 2-D form, is obtained by rearrangement of $S_W{}'$ in column-wise sequencing. The final extracted watermark, $\mathbf{W}'$, is obtained by a 2-D inverse DCT transform of $S_W{}'$. Fig. 1(d) shows the extracted watermark from the watermarked Lena image in Fig. 1(c).

3. PERFORMANCE EVALUATION

The performance of our proposed watermarking scheme is evaluated by using the watermarked Lena image (in Fig. 1(c)). The simulations are performed in the MATLAB 6.5 software environment. For a fair evaluation, performance is examined and compared with other existing DCT-based watermarking schemes from two issues – the imperceptibility of an embedded watermark and the robustness of the embedded watermark. Specially, our proposed scheme is compared with Cox's [4] and Tsai's [5] methods. Cox's scheme [4] embeds the pseudo-random binary watermark into the most perceptually significant DCT coefficients by a non-linear fashion, but the extraction requires the original image. Tsai *et al.* [5] proposed a robust 8×8 block DCT based watermarking scheme, by adopting gray-level watermarks. An adjusting quantization table is required for embedding and extraction. The middle frequency coefficients of the original image are *randomly* selected to be replaced by the quantized DCT coefficients of the watermark, therefore, the watermarked image could not achieve very high quality measure.

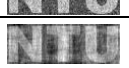
3.1. Performance Measures

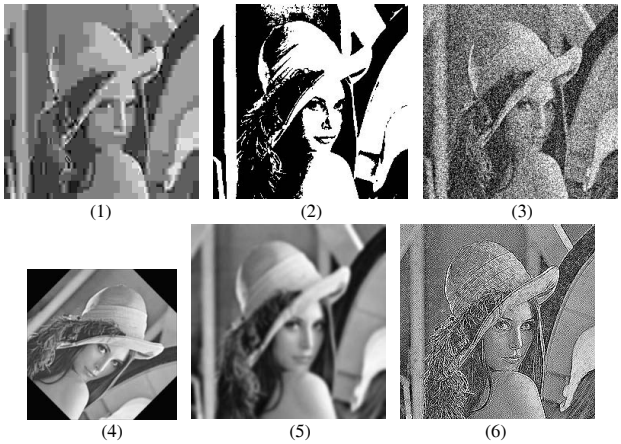
The imperceptibility of a watermark is measured by the watermarked image quality in terms of *Peak-Signal-to-Noise Ratio* (PSNR) (in dB) [1]. The robustness performance of watermark extraction is evaluated by *normalized correlation coefficient*, *r*, of the extracted watermark $\mathbf{W}'$ and the original watermark $\mathbf{W}$ (of $M1 \times M2$ pixels) [1]

$$r = \frac{\sum_{i=1}^{M1}\sum_{j=1}^{M2} \widetilde{\mathbf{W}}(i,j) \times \widetilde{\mathbf{W}}'(i,j)}{\sqrt{\sum_{i=1}^{M1}\sum_{j=1}^{M2} \widetilde{\mathbf{W}}^2(i,j) \times \widetilde{\mathbf{W}}'^2(i,j)}}, \tag{8}$$

where $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{W}'}$ are, respectively, the normalized original and extracted watermark by subtracting its corresponding mean value. The magnitude range of r is [0, 1], and the unity holds if the extracted image perfectly matches the original one.

Table 1. Watermark extraction results against various attacks.

Attacks	PSNR (dB)	Extracted watermark	r
JPEG compression 50%	33.88		0.9996
JPEG compression 1%(1)	22.79		0.9482
Color quantization to a binary image(2)	5.77		0.8889
Additive Gaussian noise(3)	13.61		0.9123
Rotation 45°, cropping and scale down by 40%(4)	–		0.8299
Blurring(5)	25.66		0.9858
Sharpening(6)	9.59		0.9339



3.2. Robustness Tests

The imperceptibility is investigated to include the transparent condition in the robustness performance evaluation. The PSNR values of the watermarked image used for evaluation in Cox's and Tsai's schemes are 40.00dB (an approximate value) and 39.65dB, respectively. Note that the watermarked image, which is obtained by our proposed scheme and used for performance evaluation, has the best PSNR measure of 61.00dB. The watermark extraction results against various attacks such as JPEG compression, quantization, additive noise, geometric distortions and filtering, are shown in Table 1. Experimental results have shown its excellent resistance against a wide range of attacks. It has been demonstrated that the proposed scheme has better robustness performance than Cox's and Tsai's schemes as well as other DCT-based watermarking schemes.

4. DISCUSSION AND CONCLUSION

A novel transparent robust watermarking technique has been proposed in this paper. Meaningful gray-level image rather than binary sequence is used as watermark. The adaptive replacement embedding method in DCT domain can hide the gray level watermark transparently. It is also found that the replacement technique is more robust against additive embedding (as in Cox's method). The selection of low to middle frequency range for embedding enhances the excellent robustness performance. At the same time, the good perceptual quality of the marked data is achieved by *adaptive replacement* with a threshold setting. One disadvantage of the proposed scheme could be the heavy computational load due to the whole DCT transform when the image size is large. This could be compromised by dividing the original image and the watermark into smaller processing blocks. Moreover, the automatic selection of the DCT coefficient range used in embedding needs to be investigated to replace our current heuristic selection by experiments.

Furthermore, the use of dual key provides a higher level of security in the practical applications for ownership verification. In an ownership verification system [6], the legal authority stores the watermark as well as one of the two keys. Only when the owner present the other key, the final watermark detection can be done by the legal authority and the ownership is then verified. Consider the scenario that one Work is created jointly by multiple owners, which arises the issue of joint ownership verification [7]. A dual-key watermarking system could be an advantage in such cases. Suppose there are two owners sharing the ownership of the Work jointly. The key can be easily separated in two parts – one part contains the watermark embedding location and the other part is the scaling information. The rightful ownership can be verified, only when both owners present their keys. These features make the proposed information hiding scheme very effective in ownership verification applications, particularly, when the quality of the watermarked data is required to be high. The practical applications for ownership verification of biomedical images using our proposed scheme, will be further investigated.

5. REFERENCES

[1] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Boston: Artech House, 2000.

[2] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," *Electronic Imaging*, vol. 3657, pp. 87-90, Jan. 1999.

[3] M. Ramkumar, A.N. Akansu, "Information theoretic bounds for data hiding in compressed images," *Proc of IEEE Second Workshop on Multimedia Signal Processing*, pp. 267-272, Dec. 1998.

[4] I. J. Cox, J. Kilian, T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1073-1087, 1997.

[5] C.-S. Tsai and C.-C. Chang, "Embedding robust graylevel watermark in an image using discrete cosine transformation," *Distributed Multimedia Databases: Techniques and Applications*, Timothy Shih (Ed.), Idea Group Publishing, Jan. 2002.

[6] Y. Wang, J. F. Doherty, and R. E. V. Dyck, "A wavelet-based watermarking algorithm for ownership verification of digital images," *IEEE Trans. on Image Processing*, vol. 11, no. 2, pp. 77-88, Feb. 2002.

[7] H. Guo and N. D. Georganas, "Digital image watermarking for joint ownership verification without a trusted dealer," *Proc. IEEE ICME'03*, vol 2, pp. 497-500, July 2003.