Watermarking Damage to Risk Multimedia in Intellectual Property Systems[&]

Hung-Jui Ke[1] Shiuh-Jeng Wang[*] Wen-Ya Chiang[+]
[*,+]Department of Information Management
Central Police University, Taoyuan, Taiwan 333
sjwang@mail.cpu.edu.tw
[1]Information Office at Changhua Police Bureau
Changhua County, Taiwan

Abstract

In this paper, we propose two algorithms to confuse the determination of intellectual property rights, when a watermark is the only method of authentication. We attempted to discover the potential drawbacks of Wang's scheme [10] and highlighted the potential risks. Based on observation of Wang's algorithms, we found the embedded watermark could be tampered without any knowledge of original secret information, such as the key or the pixel-block size. Our algorithms can make the embedded watermark "chaotic" to confuse intellectual property rights authentication. Both of our algorithms proposed in this paper can cause serious confusion without damaging the original image. The experiment results prove that our algorithms are useful to damage Wang's scheme.

Keywords： procedures, watermark embedding, intellectual rights, information hiding

1. Introduction

Using a digital watermark to claim the ownership of intellectual property has been an important issue recently. The digital watermarking scheme is applied to sound, video and still image for hiding secret information. When it comes to the digital watermark or information hiding, there are two key points: imperceptibility and robustness [1,6,7,8,9].

Imperceptibility means that the watermark cannot be detected by the human sense. The PSNR (Peak Signal Noise Ratio) and NC (Normalized Correlation) are often used to evaluate the imperceptibility of the image. The equations of PSNR and NC are showed as follows [2,3,5].

[&] This work was supported in part by National Science Council in R.O.C. under Grant No. NSC 97-2221-E-015-001-
[*] Whom corresponce

- PSNR:

$$PSNR = 10\log_{10}\frac{255^2}{MSE}dB \qquad (1)$$

where MSE (Mean Square Error) is computed as the form of

$$MSE = (\frac{1}{m \times n})\sum_{i=0}^{m-1}\sum_{j=0}^{m-1}(I_{ij} - I'_{ij})^2 \qquad (2)$$

In (2), I_{ij} denotes the $(i,j)^{th}$ pixel value of the host-image on the 2-dimentional coordinate and I'_{ij} denotes the $(i,j)^{th}$ pixel value of the stego-image on the 2-dimentional coordinate. The parameters *m* and *n* are the length and width of the image, respectively. We calculated PSNR between original image and the watermarked image. The higher the PSNR shows the better quality of watermarked image. So, if we have bigger PSNR, it shows least difference between original and watermarked one. When the PSNR value is greater than 30 dB [6], it will be very difficult to tell the difference between the two images. Robustness means the embedded watermark is not easy to be removed , but the watermark must still be clear, even after regular image processing, such as filtering, JPEG compression, cropping and so on. In this way, intellectual property rights can be guaranteed.

- NC:

$$NC = \frac{\sum_i\sum_j W(i,j)W'(i,j)}{\sum_i\sum_j [W(i,j)]^2} \qquad (3)$$

where $W_{(i,j)}$ is the pixel value of the original watermark image in the location (i,j), while $W'_{(i,j)}$ is the pixel value of the altered watermark image in (i,j) . NC can evaluate the differences between an original watermark and an extracted watermark. If the NC value is 1, this means both images are identical; if they are not identical, the NC value

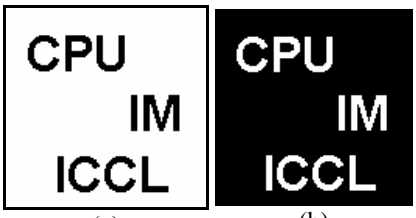
would be lower.

However, the PSNR and NC can not always represent the real conditions of an image and the watermark. For example, while the eye in the 'Lenna' image has removed, the PSNR value is still 33.77 dB, as shown in Fig. 1. We also can see the similar situation to occur in the NC value. As the extracted watermark becomes inverted, the NC value equals 0. We can still make the watermark clear enough, however, as shown in Fig. 2.

In 2002, Wang [10] proposed a watermark scheme, with the characteristics of imperceptibility and robustness. Wang's scheme is based on spatial domain and block-oriented modulo calculation to embed and extract the watermark. The experiment showed Wang's scheme has high robustness and imperceptible, the extracted watermark is still clear after filtering, requantization and JPEG compression. Hence, they claimed the technique not only superior to Lee et al. [4], but also safer. In addition, Wang's scheme can use DES-like encryption and pseudo random number to increase security.

(a) (b)

Fig. 1 (a) The original 'Lenna' image (b) A tampered 'Lenna' image, with part of the left eye destroyed



(a) (b)

Fig. 2 (a) Logo designed as watermark (b) Inverted logo, in contrast to (a)

In this paper, our goal is to uncover the potential drawbacks of Wang's scheme: the unauthorized attacker are able to confuse an embedded watermark without damaging image quality, and has no knowledge of the relevant secret information, such as the modulo number, pseudo random seed number, encryption key or size of the pixel-block.

The rest parts of this paper are presented as follows. The Sec. II, a review to the past work in [10]. Then our work is given in Sec. III, in which we propose two chaotic algorithms to confuse the embedded watermark on the basis of block-oriented information hiding system. The experiments related to our proposed algorithms are discussed in Sec. IV. Finally, we make some comments and conclude this paper in Sec. V.

2. Review of Wang's scheme

In the pre-processing procedure of Wang's scheme [10], watermark image is arranged in a one-bit listing, the original bit listing being transformed into a new one by means of the DES-like encryption and permutation functions. In the watermark embedding procedure, the original image is divided into a number of non-overlapping blocks. One pixel-block is embedded in a bit produced in the watermark transformation. The pseudo-random number generator (PRNG) is in charge of choosing the pixel-block into which data will be inserted. Finally, according to the watermark bit value and the defined parameters, the pixel values of the selected pixel-block are adjusted to embed the watermark. We summarize the techniques used in [10] as follows:

Watermark Embedding Procedure

Input: A host-image O, the bit-string associated with the bit-pattern watermark W_b, a seed key S_B and the parameter of threshold T.

Output: A stego-image with the embedded watermark

Step 1. Divide the image O into a number of non-overlapping sub-images, so that each sub-image is the size of $r \times c$, where the parameters are dependent on the size of the embedded watermark.

Step 2. Pick up a pixel-block $O_B(i)$ of $r \times c$ from O by the pseudo random number generating procedure, using the seed key S_B.

Step 3. Compute the mean, g_{mean}, for all the pixels in $O_B(i)$:

$$g_{mean} = \frac{1}{r \times c} \sum_{x=0}^{r-1} \sum_{y=0}^{c-1} b(x, y)$$

where $b(x,y)$ is the pixel in the 2-dimensional coordinate.

Step 4. Compute $g_{remainder} = g_{mean} \bmod T$, where T is a threshold.

Step 5. Compute two parameters, g_{q0} and g_{q1} as follows:

Case I: The embedded bit '0', g_{q0} is generated via the following rule:

1. IF $0 \le g_{remainder} \le \left\lfloor \frac{3T}{4} \right\rfloor$ and $g_{remainder} \neq \left\lfloor \frac{T}{4} \right\rfloor$ THEN $g_{q0} = (- g_{remainder}) + \left\lfloor \frac{T}{4} \right\rfloor$.

2. IF $g_{remainder} = \left\lfloor \frac{T}{4} \right\rfloor$ THEN $g_{q0} = (g_{remainder})$.

3. IF $\left\lfloor \frac{3T}{4} \right\rfloor < g_{remainder} < T$ THEN $g_{q0} = T + (-g_{remainder}) + \left\lfloor \frac{T}{4} \right\rfloor$. IF $(x < \frac{r}{2}$ and $y < \frac{c}{2})$ or $(x \ge \frac{r}{2}$ and $y \ge \frac{c}{2})$ THEN $b(x,y)' = b(x,y) + g_{q0} + \delta$

ELSE $b(x,y)' = b(x,y) + g_{q0} - \delta$, where δ is a variant number.

Case II: The embedded bit '1', g_{q1} is generated as follows:

1. IF $0 \le g_{remainder} \le \left\lfloor \frac{T}{4} \right\rfloor$ THEN $g_{q1} = (- g_{remainder}) - T + \left\lfloor \frac{3T}{4} \right\rfloor$.

2. IF $\left\lfloor \frac{T}{4} \right\rfloor < g_{remainder} < T$ THEN $g_{q1} = (-g_{remainder}) + \left\lfloor \frac{3T}{4} \right\rfloor$. IF $(x < \frac{r}{2}$ and $y < \frac{c}{2})$ OR $(x \ge \frac{r}{2}$ and $y \ge \frac{c}{2})$ THEN $b(x,y)' = b(x,y) + g_{q1} + \delta$.

ELSE $b(x,y)' = b(x,y) + g_{q1} - \delta$.

In the above procedure, $b(x,y)'$ denotes a new pixel, with the variant number δ being in the interval of $-2 \le \delta \le 2$.

□

Watermark Extraction Procedure

Input: A stego-image with the embedded watermark, a seed S_B and a threshold T

Output: Extracted watermark

The extract bit $W'_b(i)$ is computed according to the following rule:

$$W'_b(i) = \begin{cases} 1 & \text{if } g'_{remainder} > \left\lfloor \frac{T}{2} \right\rfloor, \\ 0 & \text{otherwise.} \end{cases}$$

□

We can learn from the above algorithms, this technique adopts the average value of the image block pixels and the modulo operation in order to embed the watermark. The level of robustness and visual quality are determined by the T value. In Wang's scheme, T was proposed to be in the interval of $3 < T < 26$ and T=12 was suggested for better hiding results. In the watermark extraction procedure, the bit pattern was determined by the threshold value $\left\lfloor \frac{T}{2} \right\rfloor$. If it is greater than $\left\lfloor \frac{T}{2} \right\rfloor$, the watermark bit value is 1; otherwise, the bit value is set as 0. To further demonstrate Wang's scheme, we set T=12 and use a 'Lenna' image size of 512×512. The logo size for the watermark sample was 128×128. These benchmarks are shown in Fig. 3. According to the experimental results, we can embed and extract a watermark by using this scheme, as shown in Fig. 4.

(a) (b)

Fig. 3 (a) 'Lenna' sized 512x512
(b) Binary pattern of watermark sized 128x128

(a) (b)

Fig. 4 (a) Stego-image, PSNR=34.37dB
(b) Extracted watermark in (a), NC = 1.

3. Two Chaotic Algorithms

In Sec. 2, we can observe that the embedded watermark step is based on the mean of a pixel-block. As a result, we maliciously break the scheme by adjusting the mean of a pixel-block. The ideas are presented below.

First of all, we define a number ε, which will be added to each pixel in the proposed algorithms. As a result, the extracted bit from the selected pixel-block is likely to be different from the original embedded bit. In Wang's scheme, for example, if $g_{remainder} < \left\lfloor \frac{T}{2} \right\rfloor$, then the extracted bit $W_b(i)=0$. This rule can be broken if we compute the number of $g'_{mean} = g_{mean} + \varepsilon$, then $g'_{remainder} = g_{remainder} + (\varepsilon \bmod T)$ according to the basic procedure in [10]. This will result in the complementary bit $W'_b(i)=1$, because the watermark extraction was based on the computation of $\left\lfloor \frac{T}{2} \right\rfloor$. If one can use the parameter ε to affect the computation of $g_{remainder}$, the bit output of the extracted watermark will result in both being complementary. In other words, if the parameter ε is chosen in $\left[\left\lfloor \frac{T}{4} \right\rfloor + \lfloor \delta \rfloor, \left\lfloor \frac{T}{2} \right\rfloor \right]$ and added to each pixel in the stego-image, the extracted watermark will be completely complementary to the original one. The two chaotic algorithms, ***CWOA* (*Complementary-watermark-output Algorithm*) and *DWOA* (*Destroyed-watermark-output Algorithm*)** are depicted as follows:

CWOA: Complementary-watermark-output Algorithm

Input: A stego-image $O=\{P(i,j)|P(i,j)$, which is the pixel value of (i,j) on the 2-dimensional coordinate$\}$, a threshold T and δ, where δ is in the interval of $-2 \le \delta \le 2$.

Output: Watermark W.

Step 1. Define $\varepsilon \in \left[\left\lfloor \frac{T}{4} \right\rfloor + \lfloor \delta \rfloor, \left\lfloor \frac{T}{2} \right\rfloor \right]$.

Step 2. Compute $P(i,j)=P(i,j)+\varepsilon$.

Step3. Extract the watermark, using the procedure in Wang's scheme.

□

Following the idea proposed in *CWOA*, we will add the parameter ε in each pixel-block. In this way, the extracted watermark will be inverted. In other words, the original watermark has become "chaotic" and it will come to confuse the ownership.

DWOA:Destroyed-watermark-output Algorithm

Input: A stego-image $O=\{P(i,j)|P(i,j)$, which is the pixel value of (i,j) on the 2-dimensional coordinate$\}$, a threshold T and δ, where δ is in the interval of $-2 \le \delta \le 2$.

Output: An extracted Watermark

Step 1. Define $\varepsilon \in \left[\left\lfloor \frac{T}{4} \right\rfloor + \lfloor \delta \rfloor, \left\lfloor \frac{T}{2} \right\rfloor \right]$.

Step 2. Divide O into a number of non-overlapping pixel-blocks, where each pixel-block is sized of $n \times n$.

Step 3. Choose $\left\lfloor \frac{n \times n}{2} \right\rfloor$ pixel numbers in each pixel-block by pseudo random generator. Let the chosen pixel, $R(i.j)$ is set as the form, $R(i,j)=P(i,j)$.

Step 4. Compute $P(i,j)'=R(i,j)+\varepsilon$.

Step 5. Extract the watermark, according to Wang's scheme in [10].

4. Experiment and Analysis

In our experiment, we examined the stego-image, shown as Fig. 4 (a). We defined ε=5 and added ε to each pixel value in the stego-image, as shown in Fig. 5. The PSNR of the confused stego-image was still 34.15 dB after our *CWOA* had been applied. Even the extracted watermark image was almost complementary, the NC for this extraction being only 0.05. Nevertheless, it is still recognizable by the human eyes, as representing the original logo. This is because the bit string of an extracted watermark is determined by

comparing the values of $g_{remainder}$ and $\left\lfloor \frac{T}{2} \right\rfloor$. If the extracted bit string of the watermark can be adjusted, the watermark output is predictable, as in the above-mentioned case. Another example is given, using *DWOA*. Assume we divided the stego-image into 2×2 pixel-blocks and defined ε to be 5. Afterwards, we added ε to the pixels, according to Step 4 in *DWOA*. As shown in Fig. 5, the PSNR of this confused stegto-image remains at 37.16 dB, the same as the basic stego-image. The NC of the extracted watermark is still 0.59, which is not so bad. Unfortunately, this chaotic watermark output gives no information of the watermark logo. Besides, the PSNR is 32.89dB when the confused stego-image and the host-image are compared. This means that the confused stego-image is extremely similar to the stego-image. Thus, the chaotic algorithms proposed in our scheme can really threaten modulo watermarking systems, such as those in [4, 10].

Although the threshold T chosen in [10] is in the interval of $3 \leq T \leq 26$, this range can easily be conjectured. Therefore, we implemented chaotic analysis for the stego-image. If the crucial parameter ε is in the interval range of $5 \leq \varepsilon \leq 11$, the intentional destruction of the embedded watermark in the stego-image will succeed. This fact is further proven from the results of experiments using the *CWOA* and *DWOA,* proposed in this scheme.

(a) **(b)**

Fig. 5 (a) Confused stego-image under *CWOA*，PSNR= 34.15dB(b) Choatic watermark extracted from (a), NC=0.05

(a) **(b)**

Fig. 6 (a) Confused stego-image under *DWOA*, PSNR= 37.16dB (b) Chaotic watermark extracted from (a), NC= 0.59

5. Concluding Remarks

In this paper, we have proposed the concerns of risking at intellectual property rights of watermark-based systems in the case of Wang's scheme [10]. In such a way that using watermark to claim ownership will be challenged. In our scheme, we are able to confuse the embedded watermark without any prior information; even the quality of the stego-image remained almost the same. In general, if the embedded watermark in a stego-image is destroyed, the intellectual property rights associated with this image will become controversial. This is because the extracted watermark is not the same as the original. In our scheme, two algorithms were proposed to confuse the embedded watermark. The first algorithm *CWOA* complements the original watermark. According to our experiment, the extracted watermark is clearly recognizable, but obviously different from the original watermark. The extracted watermark, conducted by this paper, resembles the original one, the inverted watermark bring a wrong message in the authentication of ownership. In a business transaction, the holder of a new image can be deceived by the complementary watermark, in the course of intellectual property transfer. In *DWOA*, the second algorithm in our scheme, the embedded watermark is destroyed when the watermark is extracted in step 5. In other words, the original watermark is lost. This can result in the loss of ownership of the image. Both algorithms proposed in this paper can cause serious confusion in the intellectual property rights, when the digital watermarking is the only way of protection against copyright infringements. We will develop more ideas to guarantee the reliability in authenticating legal ownership in the future.

References

[1] C.C. Chang, K.F. Hwang, and M.S. Hwang, "Robust Authentication Scheme for Protection Copyrights of Images and Graphics," IEE Proc.-Vis. Image Signal Process., Vol. 149, No. 1, pp. 43-50, February 2002.

[2] C.T. Hsu and J.L. Wu, "Hidden Digital Watermarks in Images," IEEE Transactions Image Processing, Vol. 8, No. 1, pp. 58-68, January 1999.

[3] M. Kutter and F.A.P. Peticolas, "A Fair Benchmark for Image Watermarking Systems," Electronic Imaging '99. Security and Watermarking of Multimedia Contents, Vol. 3657, Sans Jose, CA, USA, January 1999.

[4] C.H. Lee and Y.K. Lee, "An Adaptive Digital Image Watermarking Technique for Copyright Protection," IEEE Transactions on Consumer Electronics, Vol. 45, No. 4, pp. 1005-1015, November 1999.

[5] I. Nasir, Y. Weng and J. Jiang, "Novel Multiple Spatial Watermarking Technique in Color Images," Fifth International Conference on Information Technology: New Generations, pp. 777-782, 2008.

[6] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Information Hiding - A Survey," Proceedings of the IEEE, Vol. 87, No. 7, pp. 1062 -1078, July 1999.

[7] K. Rabah, "Steganography-The Art of Hiding Data," Information Technology Journal Vol. 3, No. 3, pp. 245-269, 2004.

[8] H. Wang and S. Wang, "Cyber Warfare: Steganography vs. Steganalysis," Communications of the ACM, Vol. 47, No. 10, pp. 75-81, 2004.

[9] S.J. Wang, "Steganography of Capacity Required Using Modulo Operator for Embedding Secret Image," Applied Mathematics and Computation, Vol. 164, pp. 99-116, May, 2005.

[10] S.J. Wang, "Information Hiding at Oblivious Watermarking Scheme upon Alternative Threshold in Spatial Domains," Journal of Discrete Mathematical Science & Cryptography, Vol. 10, No. 3, pp. 359-384, 2007, Taru publications, India.