Copyright Protection and Authentication of Grayscale Images by Removable Visible Watermarking and Invisible Signal Embedding Techniques: A New Approach*

Pei-Ming Huang (黃培銘) and Wen-Hsiang Tsai (蔡文祥)

Department of Computer & Information Science
National Chiao Tung University
E-mails: { gis90541, whtsai }@cis.nctu.edu.tw

Abstract

A novel method for copyright protection and authentication of grayscale images is proposed. The involved techniques include a human visual model, least significant bit replacement, and reversible pixel color adjustment. These techniques are integrated effectively to embed a visible watermark and authentication signals into the single grayscale channel of the image. The copyright of the image can be proved by extracting the embedded watermark. And the integrity of the image can be proved by verifying the existence of the embedded authentication signals. Besides, the embedded visible watermark can be removed, if desired, so that only the watermarked image need be kept, providing an advantage of storage space saving. Some experimental results are shown to prove the feasibility of the proposed method.

Keywords: digital watermarking, grayscale image, visible watermark, copyright protection, image authentication, integrity verification.

1. Introduction

With the rapid growth of digital processing techniques, digital images may be duplicated and tampered easily, resulting possibly in unauthorized uses or modifications of them. Therefore, it is necessary to develop appropriate methods to protect the copyright of images and verify their integrity. One approach is to use digital watermarking techniques. Such techniques can be categorized into two types: visible watermarking and invisible watermarking. Invisible watermarking [1-8] embeds a watermark in an image in a way that yields imperceptible results in the resulting *stego-image* under normal observations. The watermark is extracted from the stego-image for copyright claiming when necessary. On the contrary, visible watermarking [9-13] creates an observable watermark on the top of the image to convey an immediate claim of the ownership of the image.

When using a conventional visible watermarking technique, a user must keep both the original image and the watermarked one. This results in a waste of storage space because of the requirement of a double amount of storage. Consequently, a new type of watermark, called *removable visible watermark*, is proposed in this study for grayscale images. Accordingly, a user can remove the embedded visible watermark from a

* This work was supported partially by the NSC 92-2422-H009-010 and the MOE Program for Promoting Academic Excellency of Universities under the grant number 89-1-FA04-1-4.

watermarked image so that only the watermarked image need be saved.

The proposed method is based on three techniques, namely, the least significant bit replacement technique, a human visual model, and a reversible pixel color adjustment scheme. Unlike a full-color image that has three channels of colors, namely, red (R), green (G), and blue (B), a grayscale image has only one channel, the grayscale channel. In this study, effective integration of the three techniques is designed to embed various involved signals into the single grayscale signal without mutual interference. More specifically, techniques capable of embedding a visible watermark into a grayscale image and removing it from the stego-image to restore the original image are proposed. Also proposed are techniques for embedding authentication signals into the image and checking their existence to verify the integrity and fidelity of the image.

In the remainder of this paper, the human visual model employed in this study for authentication signal embedding and extraction is first described in Section 2. In Section 3, the proposed copyright and integrity protection techniques is presented. In Section 4, several experimental results are illustrated. Finally, some conclusions are made in Section 5.

2. Review of A Human Visual Model

The human visual model (HVM), proposed in [14] and modified in [15], is employed in this study for authentication signal embedding and extraction. The model guarantees that modifications of images are imperceptible. By considering the eight surrounding pixels of a 3×3 image block as the background of the central pixel of the block, the standard deviation σ of the background is first calculated. The HVM regards σ as a parameter and takes it as an input to classify each 3×3 image block into one of four classes, from smooth areas to edged ones, and divide accordingly the grayscale range (from 0 to 255) into a certain number (denoted as n) of quantization levels by a *contrast function* $f(\sigma)$ defined in the following:

$$f(\sigma)=n=\begin{cases}32 & \text{when } \sigma \le 2.4 \text{ (class 1 : smoothest);}\\ 24 & \text{when } 2.4 \le \sigma \le 3.6 \text{ (class 2 : smooth);}\\ 16 & \text{when } 3.6 \le \sigma \le 4.8 \text{ (class 3 : edged);}\\ 12 & \text{when } 4.8 \le \sigma \text{ (class 4 : most edged).}\end{cases}$$

Let g be the gray value of the central pixel of an image block, which falls within one of the quantization levels, say L. Let the two boundary values of L be denoted as g_{min} and g_{max}, respectively, so that $L = [g_{min}\ g_{max}]$. Then the HVM guarantees that all gray values in L against backgrounds with an identical standard deviation σ will have the same sensitivity to the human vision, implying that a modification of the image block content by replacing g with a gray value within the level L will cause imperceptible effect.

3. Proposed techniques for Copyright and Integrity Protection

In this section, the proposed method for embedding a visible watermark and certain authentication signals in a grayscale image will be described as a *watermarking process*. The visible watermark can be extracted to verify the copyright and can be removed, if necessary, without referring the original image. Furthermore, by checking the embedded authentication signals, indirect copyright protection and tampering detection can be achieved.

Each visible watermark used for image ownership protection is assumed to be a binary image in this study. Pixels in the watermark with values 1 will be called *black watermark pixels* and those with values 0 *white ones*.

A. Proposed watermarking process

The inputs to the proposed watermark embedding process are a grayscale image C and a binary watermark W. The output is a stego-image S. The process can be briefly expressed as an algorithm as follows.

1. (**Division of input image into two parts**) Divide C and W into non-overlapping 3×3 blocks, and partition C into a watermark area C_w and a non-watermark area C_n according to a pre-selected position in C where a semi-transparent visible copy W' of W is to be created.
2. (**Insertion of visible watermark**) Adjust the pixel value p of each pixel P in each 3×3 block in C_w according to the following reversible adjustment function f_r to create the above-mentioned W' on the top of C:

$$p' = f_r(p) = p \times (1 - w \times k), \qquad (1)$$

where p' denotes the new value of p, w is the binary value (0 for white and 1 for black) of the corresponding watermark pixel imposed on P, and k is a pre-selected *embedding ratio* with its value being within the range of 0 and 1, which controls the darkness of the created visible watermark on the output watermarked image. That is, replace the value p of P with p' defined above.

3. (**Labeling of watermark signals**) Set the least significant bit (LSB) of the value p' of each of the eight surrounding pixels of each block B in C_w to be w as labels of the visible watermark, which facilitate later watermark extraction.
4. (**Embedding of authentication signals**) Embed authentication signals according to the HVM in the following way.
   4.1. For each 3×3 image block B_w in C_w, find g_{min} with the HVM, and replace the central pixel value g of B_w with the authentication signal $g_{min} + \alpha$ if $w = 0$; or with $g_{min} + \beta$, otherwise, where α and β are two pre-selected constants.
   4.2. For each 3×3 image block B_n in C_n, find g_{min} with the HVM, and replace the central pixel value g of B_n with the authentication signal $g_{min} + \gamma$, where γ is a pre-selected constant.
5. (**Creation of the watermarked image**) Take the final result as the desired stego-image S.

It is mentioned here that the three constants α, β, and γ are selected in such a way that the authentication signals, $g_{min} + \alpha$, $g_{min} + \beta$, and $g_{min} + \gamma$, are not larger than g_{max}. In addition to being used for authentication of images, these signals can also be used to label watermark pixels for deciding whether a 3×3 block is in the watermark area or in the non-watermark area, as described next.

B. Visible Watermark Extraction Process

The input to the proposed visible watermark extraction process includes just a stego-image S. The output is a *reconstructed watermark image* W which is embedded presumably in S. The extraction process is described as an algorithm as follows.

1. (**Classification of stego-image blocks**) Classify each 3×3 image block B of S in the following way: find the value of g_{min} of B according to the HVM, and if the central pixel value g of a 3×3 block B is $g_{min} + \alpha$ or $g_{min} + \beta$, then decide B to be in the watermark area; if g is $g_{min} + \gamma$, then decide B to be in the non-watermark area; otherwise, decide B to be neither in the watermark area nor in the non-watermark one, i.e., regard it as a tampered block.
2. (**Reconstruction of embedded watermark**) For each 3×3 image block B of the watermark area, perform the following operations.
   2.1. For each pixel P of the eight surrounding ones of B, if the LSB value b of P is 1, set the

corresponding watermark pixel to be black; otherwise, white.

2.2 For the value g of the central pixel P of B, if $g = g_{min} + \alpha$, then set the value of the pixel in the desired reconstructed watermark W corresponding to P to be white; otherwise, black.

3. Take the final result as the desired reconstructed watermark image W.

C. Original Image Recovery Process

We use an inverse function f_r' corresponding to the reversible adjustment function f_r mentioned previously in (1) to remove the embedded watermark to recover the original non-watermarked image. The input to the proposed original image recovery process is a stego-image S. The output is a *recovered image* R. The recovery process is described as an algorithm as follows.

1. (**Classification of stego-Image blocks**) Classify each 3×3 image block B of S in the following way: find the value of g_{min} of B according to the HVM, and if the central pixel value g of a 3×3 block B is $g_{min} + \alpha$ or $g_{min} + \beta$, then decide B to be in the watermark area; if g is $g_{min} + \gamma$, then decide B to be in the non-watermark area; otherwise, decide B to be neither in the watermark area nor in the non-watermark one, i.e., regard it as a tampered block.
2. (**Removing Embedded Watermark**) For each 3×3 image block B of the watermark area, perform the following operations.

2.1 For each pixel P of the eight surrounding ones of B of the watermark area, check the LSB b of the pixel value p' of P. If $b = 1$, define a value denoted by w to be 1; otherwise, 0. And replace p' with a new value p computed by the use of the inverse function f_r' in the following way:

$$p = f_r'(p') = p'/(1 - w \times k), \quad (2)$$

where k is the embedding ratio selected before in the watermark embedding process.

2.2 For the value g of the central pixel P of B, if g is equal to $g_{min} + \alpha$, keep g unchanged; otherwise, replace g with the average of the values of the eight surrounding pixels of P.

3. (**Creation of the recovered image**) For the non-watermark area, keep all the pixel values of each 3×3 block unchanged. Take the final result as the desired recovered image R.

D. Image Authentication Process

The input to the proposed image authentication process is a suspicious image U. The output is an *authentication image* A with its blocks being marked tampered or unaltered. The proposed authentication algorithm is expressed as an algorithm as follows.

1. (**Classification of Stego-Image Blocks**) Classify each 3×3 image block B of S in the following way: find the value of g_{min} of B according to the HVM, and if the central pixel value g of a 3×3 block B is $g_{min} + \alpha$ or $g_{min} + \beta$, then decide B to be in the watermark area; if g is $g_{min} + \gamma$, then decide B to be unaltered; otherwise, tampered.
2. (**Creation of the authentication result**) Mark tampered blocks in red colors, and unaltered ones in white as the desired authentication result.

4. Experimental Results

Some experimental results of applying the proposed method are shown here. A binary image of size 256×256 as shown in Figure 1 is used as the visible watermark. The embedding ratio k is assigned to be 0.4. Figures 2(a) and (b) show two grayscale images both with size 512×512. And the

corresponding stego-images after embedding the visible watermark are shown in Figures 2(c) and (d), respectively. Two recovered images from Figures 2(c) and (d) are shown in Figures 3(a) and (b), respectively. The corresponding PSNR values are shown in Table 1. The high values in the table show that the quality of the recovered images is still good. Finally, two tampered and cropped images are shown in Figures 4(a) and (b). And Figures 4(c) and (d) show the resulting authentication images. The red parts indicate the correctly detected tampered areas.

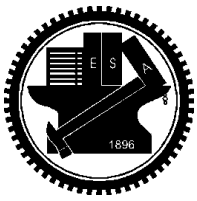


Figure 1 A watermark image of size 256× 256.

5. Conclusions

In this paper, we have proposed a method for embedding a visible watermark into a grayscale image. A visible watermark can be shown semi-transparently on the top of a grayscale image by adjusting the pixel values of the grayscale image. And with the LSB replacement technique and the use of an HVM, we can label watermark signals and embed authentication signals in the image. The proposed method includes four functions. The first is to embed a visible watermark to claim the image copyright of the owner. The second is to extract the visible watermark from the stego-image. The third is to remove the visible watermark to avoid a waste of storage space. The recovered image can be considered as the original image because the quality of the recovered image is almost the same as the original one. The last function is to authenticate the stego-image. With the help of extracted authentication signals, tampered blocks can be located to achieve the verification of image integrity and fidelity, so protecting further the image copyright indirectly.

References

[1] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," *Proc. IEEE Int. Conf. Image Processing,* Vol. Ⅱ, 1994, pp. 86–90.

[2] G. Voyatzis, I. Pitas, "Applications of toral automorphisms in image watermarking," *Proc. IEEE Int. Conf. on Image Processing (ICIP'96)*, Vol. Ⅱ, Lausanne, Switzerland, Sept. 1996, pp. 237–240.

[3] J. Fridrich, "Robust bit extraction form images," *Proc. IEEE ICMCS'99 Conf.*, Florence, Italy, June 1999.

[4] W. Bender, N. Morimoto, and D. Gruhl, "Method and apparatus for data hiding in images," *U. S. Patent*, No. 5689587, 1997.

[5] I. J. Cox, J. Kilian, F. T. Leighton and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, Vol. 6, No. 12, pp. 1673–1687, 1997.

[6] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, "A DCT-domain system for robust image watermarking," *Signal Processing*, Vol. 66, pp. 357–372, 1998.

[7] C. T. Hsu and J. L. Wu, "DCT-Based watermarking for video," *IEEE Trans. on Image Processing*, vol. 8, pp. 58–68, 1999.

[8] S. D. Lin and C. F. Chen, "A Robust DCT-Based Watermarking for copyright Protection," *IEEE Trans. on Consumer Electronics*, Vol. 46, No. 3, pp. 415–421, August 2000.

[9] Y. J. Cheng and W. H. Tsai. "Double copyright

protection and authentication of bitmap images by embedding removable visible watermarks and irremovable invisible authentication signals in multiple color channels," accepted and to appear in *IEEE Signal Processing Letters.*

[10] Mohan S. Kankanhalli, Rajmohan, and K. R. Ramakrishnan, "Adaptive visible watermarking of images." *Proc. 1999 IEEE Int. Conf. on Multimedia Computing and Systems*, Vol. 1, pp. 568 –573, 1999.

[11] Saraju P. Mohanty, K.R. Ramakrishnan, and Mohan S. Kankanhalli, "A DCT domain visible watermarking technique for images." *Proc. 2000 IEEE Int. Conf. on Multimedia and Expo*, Vol. 2, pp. 1029 –1032, 2000.

[12] Pei-Min Chen, "A visible watermarking mechanism using a statistic approach," *Proc. 5th Int. Conf. on Signal Processing*, Vol. 2, pp. 910 –913, 2000.

[13] Yongjian Hu and Sam Kwong, "Wavelet domain adaptive visible watermarking," *Electronics Letters*, Vol. 37, pp. 1219 –1220, Sept. 2001.

[14] C. H. Kuo and C. F. Chen, "A prequantizer with the human visual effect for the DPCM," *Signal Processing: Image Communication*, Vol. 8, pp. 433-442, 1996.

[15] D. C. Wu and W. H. Tsai, "Embedding of any type of data in images based on a human visual model and multiple-based number conversion," *Pattern Recognition Letters*, Vol. 20, pp. 1511-1517, 1999.

(a)

(c)

(b)

(d)

Figure 2 Input grayscale images, and output stego-images with the visible watermark of Figure 1. (a) Grayscale image "Lena". (b) Grayscale image "Jet". (c) and (d) Stego-images after embedding the visible watermark, respectively.

(a)

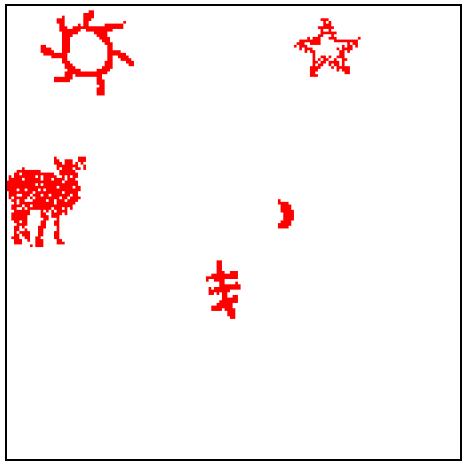
(b)

Figure 3 Recovered images. (a) Recovered image "Lena". (b) Recovered image "Jet".

Table 1 The PSNR values of the recovered images after removing the visible watermark.

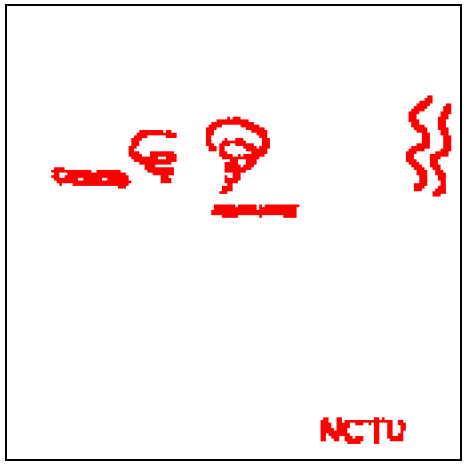
	Lena	Jet
PSNR	42.3	42.4

(a)



(c)

(b)



(d)

Figure 4 Some tampered images and authentication results. (a) Tampered image “Lena”. (b) Tampered image “Jet”. (c) and (d) authentication results of (a) and (b), respectively.