BI-LEVEL IMAGE WATERMARKING FOR IMAGE AUTHENTICATION SURVIVING JPEG LOSSY COMPRESSION

Jagdish C. Patra, Chong How Tan and Ee-Luang Ang

School of Computer Engineering
Nanyang Technological University
Singapore
E.mail: aspatra@ntu.edu.sg, chonghow@hotmail.com, aselang@ntu.edu.sg

ABSTRACT

Techniques to establish authenticity and integrity of digital images are essential for secure transactions through distributed and unreliable systems, such as the Internet. Here, we propose two novel semi-fragile watermarking schemes that are designed to survive JPEG lossy compression for ownership identification and content authentication of digital images. First, we propose to embed bi-level images into the high-frequency block-DCT coefficients of the host image. The watermarked image is authenticated by marking the region of tampering (RoT) with crosses, if any tampering was made. Next, we modify the first scheme such that instead of reflecting the RoTs on the watermarked image, they will be reflected on the bi-level images directly. These schemes provide satisfactory results and can distinguish JPEG lossy compression from malicious content modification.

1. INTRODUCTION

Success of the Internet resulted in the wide distribution of digital media because it is inexpensive and delivery is almost instantaneous. However, content providers also see a high risk of piracy. As a result, digital watermarking of digital media has become an active research area over the past several years. The growth of new powerful image editing applications has made it easier to tamper with digital images in ways that are difficult to detect. With the ease of editing in digital domain, content authentication of digital images has become an important concern.

The importance of content authentication has led to the development of image authentication techniques. These techniques are designed to verify the integrity of the images by detecting malicious content modification.

Wolfgang and Delp [1] developed an authentication method which reshapes a bipolar m-sequence into two-dimensional watermark blocks. These blocks are then added on a block-wise basis. This method can localize manipulation on a block-wise basis and is moderately robust with respect to linear and nonlinear filtering and small additive noise. Fridrich [2] proposed the idea of a robust watermarking technique for authentication. The image is divided into blocks of 64 by 64 and embedded with quasi-VQ codes in each block using the spread spectrum method. This technique is robust to manipulations but results in more error for JPEG compression. Lin and Chang [3],[4] proposed a scheme that makes use of the invariant properties of DCT coefficients before and after JPEG compression. This technique is able to withstand manipulations such as JPEG lossy compression and reasonable brightness adjustment on the watermarked image up to a pre-determined quality factor. Some of other image authentication techniques and survey papers may be found in [5]-[8].

This paper focuses on semi-fragile watermarks that are designed to survive JPEG lossy compression. We propose two image authentication schemes which are extensions of the system used in [3],[4]. As opposed to the system in [3],[4], instead of generating the authentication bits from the low-frequency DCT coefficients, we propose to use bi-level images to generate our authentication bits. In addition, our watermarking region is limited to 32 high-frequency DCT coefficients (Fig. 1). Furthermore, these coefficients are shuffled before they are embedded with the authentication bits using the exclusive OR (XOR) operation to enhance the security of our system. The blocks are also shuffled to change the order in which the authentication bits are embedded into the host image. In all the shuffling steps, secret seeds belonging to the owner are used to associate the image and the watermark with the original owner to achieve ownership identification.

In the first scheme, we propose to embed the authentication bits generated using two bi-level images into the high-frequency block-DCT coefficients of the host image using operation. During authentication, any identified region of tampering (RoT) is marked as crosses on the watermarked image. In the second scheme two bi-level images are embedded into the host image in such a way that the RoT will be reflected on the extracted bi-level images, if the watermarked image is tampered.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

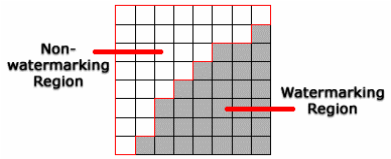


Fig. 1. Position of the two disjoint regions.

2. THE FIRST SCHEME: AUTHENTICATION USING TWO BI-LEVEL IMAGES

In the first scheme, we propose to insert two bi-level images as watermarks into the host image. Consider the host image, X, of size M by N pixels as a 256 level grayscale image. We partition X into non-overlapping blocks, X_r, of size 8 by 8 pixels. In each X_r, the whole space of 64 DCT coefficients is divided into two disjoint regions: non-watermarking region and watermarking region as shown in Fig. 1.

Let $A^{(1)}$ be a bi-level image of size I x J pixels ($I < M$ and $J < N$). By tiling, $A^{(1)}$ is converted into another image $W^{(1)}$ comprising of $S_i/2$ copies of $A^{(1)}$, where S_i is the number of authentication bits to be embedded in each block. In a similar manner, another bi-level image $W^{(2)}$ comprising of $S_i/2$ copies of $A^{(2)}$ is generated. As $W^{(i)}$, i=1, 2, are bi-level images, each of their pixels can be encoded by a single bit. To generate the authentication bits S_i, the pixel values of $W^{(i)}$, i = 1, 2, are read in a raster scan manner but in an interleaving fashion.

2.1. Watermark Embedding Process

During the embedding process (Fig. 2), the host image X, is partitioned into non-overlapping blocks X_r, of size 8 by 8 pixels. The blocks X_r are then converted to their DCT domain to form the set of DCT blocks x_r. By shuffling the order of x_r according to secret seed 1, a set of shuffled blocks, x'_r is obtained. This reordering of x_r is to ensure that the order of authentication bits will be hidden after they are embedded into x'_r. For each shuffled block x'_r, 32 high frequency coefficients C, from the watermarking region are selected in an order determined by secret seed 2. Next, The shuffled coefficients C', the authentication bits S_i and a quality threshold q, are fed into the watermark embedder. The threshold q, is used to specify the level of JPEG compression that the watermark should survive. It is defined in terms of quality factor (QF) as [3]

$$\begin{aligned} & q = (2 - QF*0.02) && \text{if } QF \geq 50 \\ \text{or} \quad & q = 50/QF && \text{if } QF < 50. \end{aligned} \qquad (1)$$

From the set of C', m coefficients are selected to embed one authentication bit. If the number of authentication bits to be embedded in each block is denoted by S_i, then $m = 32 \div S_i$. Each of the m coefficients is subsequently divided by its corresponding quantization factor $Q[i]$ (obtained from the quantization table Q_{50} in [3]) and the quality threshold q, and rounded to the nearest integer.

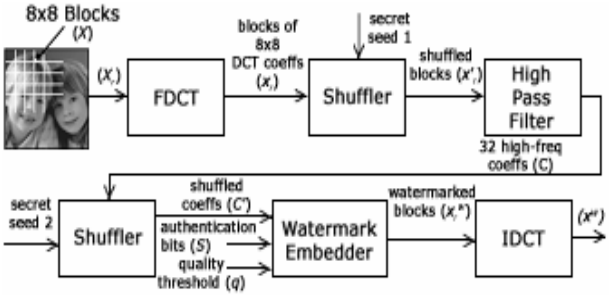


Fig. 2. Scheme 1: The watermark embedding process.

Next, the LSB of the resulting m' integers are XORed to obtain the bit representation b. If b is different from the authentication bit, the LSB of the integer that will cause the minimum distortion will be flipped, otherwise no change is made. These m' integers are then multiplied by the same quantization factor $Q[i]$ and the quality threshold q to obtain the watermarked DCT coefficients. This step is repeated until all authentication bits have been embedded to generate the watermarked blocks, x_r^W. Finally, x_r^W are converted to their spatial domain to form the watermarked image x^W.

2.2. Watermark Extraction Process

The extraction of the watermark is a reverse process of embedding and is shown in Fig. 3. The received watermarked image y^W is partitioned into non-overlapping blocks, y_r^W of size 8 by 8 pixels. The blocks are then converted to their DCT domain to obtain the set of DCT blocks Y'^W_r. Next, Y'^W_r are shuffled with secret seed 1, to obtain a set of shuffled blocks y'_r. For each shuffled block y'_r, 32 high frequency coefficients D, from the watermarking region are selected in the order determined by secret seed 2. From the set of shuffled coefficients D', m coefficients are selected to extract one watermark bit.

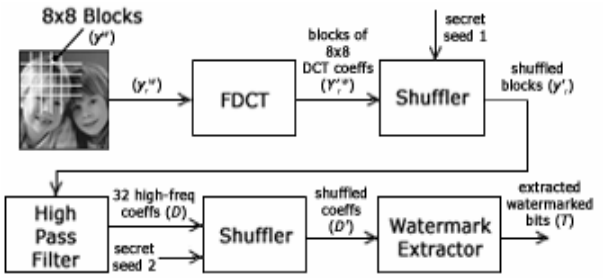


Fig. 3. Scheme 1: The watermark extraction process.

The number of coefficients selected is similar to that of embedding process. Thereafter, each of the m shuffled coefficients is divided by its corresponding quantization factor $Q[i]$ and the quality threshold q, and rounded to the nearest integer. Next, the LSB of the resulting m integers are XORed to obtain the extracted watermark bit T_i. This step is repeated until all watermark bits T have been extracted to form the bi-level images $V^{(i)}$, i = 1, 2.

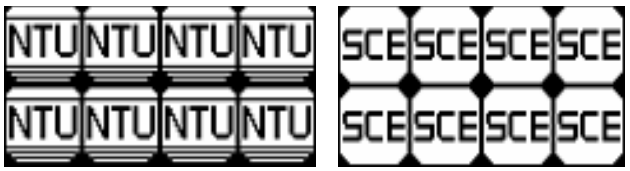
2.3. Image Authentication Process

With the extracted bi-level images $V^{(i)}$, the authenticator is able to identify the RoTs by incorporating consensus seeking on $V^{(i)}$. As $A^{(i)}$ is tiled to form $W^{(i)}$ during embedding, several sets of $A^{(i)}$ must exist in $V^{(i)}$. Through the use of voting, the majority decision can be followed to identify the tampered bits and in turn the tampered blocks from the sets of $A^{(i)}$.

2.4. Simulation and Results

A simulation was carried out for the image authentication of the Two-Girls image of size 256 x 256 pixels, with QF=75 and S_i=16. Two bi-level images $A^{(i)}$, i=1, 2, each of size 32 x 32 pixels were used as watermarks. As S_i=16 was used, $A^{(1)}$ was tiled to form another image $W^{(1)}$ comprising of 8 copies of $A^{(1)}$. Another bi-level image $W^{(2)}$ comprising of 8 copies of $A^{(2)}$ was generated similarly. The two bi-level images $W^{(i)}$, i=1, 2, are shown in Fig. 4.

To generate the authentication bits S_i, the pixel values of $W^{(i)}$ are read in a raster scan manner but in an interleaving fashion between them. Original Two-Girls image and the watermarked image (embedded with Figs. 4(a) and (b)) are shown in Figs. 5(a) and 5(b), respectively. It can be seen that the two images are perceptually indistinguishable.



(a) (b)

Fig. 4. Scheme 1: (a) 1st and (b) 2nd bi-level images.

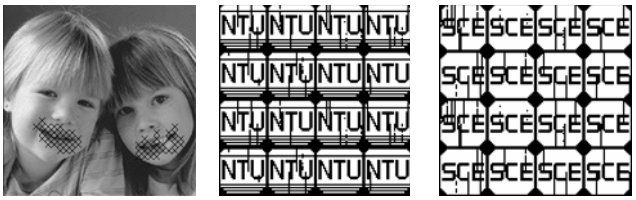
(a) (b) (c)

Fig. 5. Scheme 1: (a) Original Two-Girls image; (b) Watermarked image embedded using QF=75 and S_i=16; (c) Tampered watermarked image.

To evaluate the performance of the watermarks against JPEG lossy compression, the level of QF was varied to observe its effect on the minimum level of JPEG compression that the watermarks can survive. As shown in Table 1, all the watermarks were able to survive lossy compression at the levels that were close to their expected levels. For example, with S_i=16 and QF=80, our scheme can survive JPEG compression of the watermarked image up to QF=86 instead of theoretical value of QF=80.

Table 1. Scheme 1: Minimum QF for the watermarks to survive JPEG lossy compression before distortion occurs.

S_i	Quality Factor (QF) used during embedding								
	95	**90**	**85**	**80**	**75**	**70**	**65**	**60**	**55**
2	99	94	90	86	82	74	70	66	61
4	99	94	90	86	82	75	70	66	61
8	99	94	90	86	82	78	70	66	61
16	99	94	90	86	82	79	70	66	62

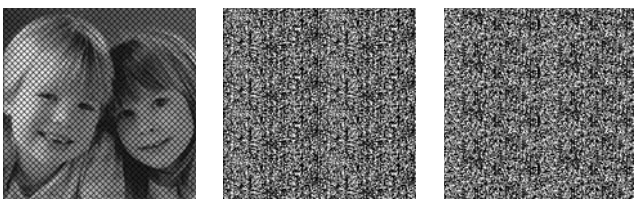


(a) (b) (c)

Fig. 6. Scheme 1: (a) Authenticated image using the correct seeds with QF=75 and S_i=16; (b) 1st and (c) 2nd extracted bi-level image.

The watermarked image was tampered by swapping the mouths of the two girls and is shown in Fig. 5(c). The results of the authentication when S_i=16 are shown in Fig. 6. As shown in Figs. 6(b) and 6(c), the extracted bi-level images clearly indicate that some tampering had been performed although the RoTs could not be identified straight away. The RoTs were identified through consensus seeking on the extracted bi-level images and marked on the watermarked image as shown in Fig 6(a). With the use of the correct secret seeds, the authorized person will be able to determine the authenticity of the images.

Figure 7 shows the authenticated image and the extracted bi-level images when a wrong secret seed 1 and/or 2 was used during extraction. Thus, the received image fails the authentication test. Figure 8 shows the authenticated image after the watermarked image had undergone a level of JPEG lossy compression that was beyond acceptable level.



(a) (b) (c)

Fig. 7. Scheme 1: (a) Authenticated image using the wrong value for secret seed 1 and/or 2 with QF=75 and S_i=16; (b) 1st and (c) 2nd extracted bi-level images.

(a) (b) (c)

Fig. 8. Scheme 1: (a) Authentication of a watermarked image that had been compressed beyond the acceptable limit with QF=70 and S_i=16; (b) 1st and (c) 2nd extracted bi-level images.

3. THE SECOND SCHEME: AUTHENTICATION ON BI-LEVEL IMAGES

In the first scheme, the result of the authentication was reflected on the watermarked image. In this scheme, we propose to insert two bi-level images as watermarks into the host image in such a way that any tampering made to the watermarked image will be reflected on the bi-level images directly. To generate the authentication bits, first a block of pixel values $W_r^{(1)}$, of size k x l is read from $W^{(1)}$, followed by a block of pixel values $W_r^{(2)}$, of size k x l from $W^{(2)}$. The size of $W_r^{(i)}$, i = 1, 2, is equal to S_i. The dimension of $W_r^{(i)}$ should be carefully chosen so that the position of $W_r^{(i)}$ in $W^{(i)}$ will be preserved in the watermarked image when embedded.

3.1. Watermark Embedding and Extraction Process

The embedding process is similar to that of the first scheme except that the order of the DCT blocks x_r is not shuffled in this case. The authentication bits S_i, must be embedded into the host image in the same order in which they are generated in order to reflect the RoTs accurately. The extraction process is also similar to that of the first scheme, with the omission to shuffle the order of y_r^w.

3.2. Image Authentication Process

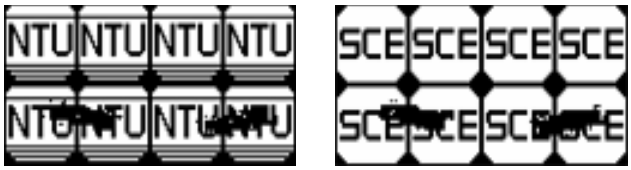
The authenticity of the image can be determined by visually inspecting $V^{(i)}$. If any tampering has been made to the watermarked image, the RoTs will appear on $V^{(i)}$. Similar set of simulation was carried out for the image authentication of the Two-Girls image in Fig. 5(a). The results of the minimum QF acceptable by the watermarks before distortion are similar to that shown in Table 1. The watermarked image and the tampered image are shown in Fig. 9. The results of the authentication for S_i=16 are shown in Fig. 10. The RoTs have been identified to be in the lower half of the watermarked image.

4. CONCLUSIONS

We proposed two novel schemes for image authentication and ownership verification. The proposed schemes are capable of surviving JPEG lossy compression up to a pre-determined level while detecting malicious manipulations. They are able to detect and report any RoTs made on the watermarked image. Furthermore, it is only possible for an authorized person who has possession of the secret seeds to check the ownership of the image.

(a) (b)

Fig. 9. Scheme 2: (a) Watermarked image embedded using QF=75 and S_i=16; (b) Tampered watermarked image.



(a) (b)

Fig. 10. Scheme 2: (a) 1st and (b) 2nd extracted bi-level images.

5. REFERENCES

[1] R. B. Wolfgang and E. J. Delp, "A Watermark for digital images", *IEEE Proc. of ICIP*, Lausanne, Switzerland, September 1996

[2] J. Fridrich, "Image Watermarking for tamper detection," *IEEE International Conf. on Image Processing*, Chicago, October 1998

[3] C-Y. Lin and S.-F. Chang, "Semi-fragile watermarking for authenticating JPEG visual content," *SPIE International Conf. on Security and Watermarking of Multimedia Contents II*, vol. 3971, No. 13, EI'00, San Jose, USA, January 2000

[4] C-Y. Lin and S.-F. Chang, "A robust image authentication method distinguishing JPEG compression from malicious manipulation," *IEEE Trans. Circuits and Systems of Video Tech.,* vol. 11, no. 2, pp. 153-168, Feb. 2001.

[5] C-T. Hsu and J-L. Wu, "Hidden digital watermarks in images," IEEE Trans. Image Processing, vol. 8, no. 1, pp. 58-68, Jan. 1999.

[6] F. Bartolini, A. Tefas, M. Barni, and I. Pita, "Image authentication techniques for surveillance applications," *Proc. IEEE*, vol. 89, pp. 1403-1418, Oct. 2001.

[7] C. Rey and J-L. Dugelay, "A survey of watermarking algorithms for image authentication, *EURASIP Applied Signal Processing*, vol. 6, pp. 613-621, 2002.

[8] M U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," vol. 11, pp. 585-595, June 2002.