Fragile Watermarking Scheme for Accepting Image Compression

Mi-Ae Kim[1], Kil-Sang Yoo[2], Won-Hyung Lee[3]

Department of Image Engineering,
Graduate School of Advanced Imaging Science, Multimedia & Film,
Chung-Ang University
#10112, Art Center, 221 Hukseok-Dong, Dongjak-Gu, Seoul, Korea, 156-756
kimma@dreamwiz.com[1], lucky@ms.cau.ac.kr[2], whlee@cau.ac.kr[3]

Abstract. As images are commonly transmitted or stored in compressed form such as JPEG lossy compression, image authentication demands techniques that can distinguish incidental modifications (e.g., compression) from malicious ones. In this paper, we propose an effective technique for image authentication which can prevent malicious manipulations but allow JPEG compression. An image is divided into blocks in the spatial domain, each block is divided into two parts by randomly selecting pixels, and average gray values for the parts are calculated. The average value is compared with that of the adjoining block to obtain an authentication signature. The extracted authentication information becomes the fragile watermark to be inserted into the image's frequency domain DCT block. The experimental results show that this is an effective technique of image authentication.

1 Introduction

Image authentication plays an extremely important role in the digital age, allowing verification of the originality of an image. This is due to the fact that powerful and easy-to-use image manipulation has made it possible to modify digital images without leaving evidence of modification. In order to better utilize the bandwidth and to minimize the space required for storage, most multimedia content such as images, audio or video are stored or transmitted in compressed formats like JPEG lossy compression. Consequently, an image authentication system must accept compressed images while detecting malicious manipulations such as replacement or removal of original objects.

Image authentication techniques fall into two broad categories: digital signature and digital watermark. A digital signature is based upon the idea of public key encryption. An extracted image digest is encoded using a hashing function and then transmits them to a receiver along with image. A private key is used to encrypt a hashed version of the image. If the hash values correspond, the image is authenticated. This approach does not permit even a single bit change. Therefore, it is not appropriate to apply this method to an image authentication system, as images must often be compressed and/or quality enhanced. Different from digesting of data as described above, there is the digital signature approach, which is based on the features of an image [1-4]. In this approach, which is used frequently for image authentication, the features of an image that are resistant to common image processing (including compression) are extracted and are used as a digital signature. The digital signature is stored (or transmitted) separately from the image. Thus, the original image is not modified; however, it is cumbersome to manage digital signature separately from images. In the watermark-based approach, authentication information is inserted imperceptibly into the original image [5-8]. If the image is manipulated, it should be possible to detect the tampered area through the fragility of the hidden authentication information (watermark). Ideally, the embedded watermark should only be disrupted by malicious modifications; it should survive acceptable modifications such as compression.

Bhattacharjee and Kutter [3] proposed techniques to generate digital signatures based on the locations of feature points. The advantage of this technique is its compact signature length. However, the selection process and relevance of the selected points are unclear [10].

The scheme proposed by Chun-Shin Lu and Hong-Yuan Mark Liao [4] relies on the fact that the interscale relationship is difficult to destroy with incidental modification but is hard to preserve in cases of malicious manipulation. But, the image authentication scheme is verified by having the sender store the digital signature.

Kundur and Hatzinakos [5] designed a wavelet-based quantization process that is sensitive to modification. The disadvantages are that their method cannot resist incidental modifications and the tampering detection results are very unstable.

Zou *et al.* [9] embedded a verification tag into the spatial domain of the image after having extracted it using a DSA (digital signature algorithm) on the DCT (discrete cosine transform) block. However, their image authentication system can only tolerate JPEG-quality factors greater than or equal to 80, and the algorithm requires extensive computation.

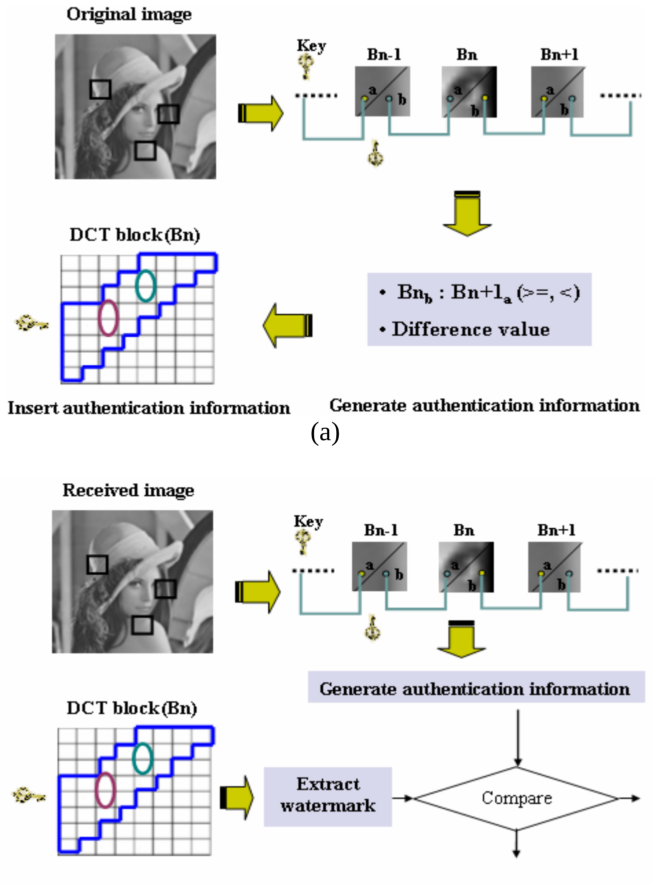
This paper proposes an effective fragile watermarking scheme for image authentication that can detect malicious manipulations while remaining robust towards JPEG lossy compression. An image is divided into blocks in the spatial domain, each block is divided into two parts by randomly selecting pixels, and average gray values for the parts are calculated. The average value is compared with that of the adjoining block to obtain an authentication signature. The extracted information becomes the fragile watermark to be inserted into the image's frequency domain DCT block.

The advantage of the proposed image authentication scheme is that it can easily extract authentication information robust to JPEG lossy compression from the spatial domain and insert it into the host image through a very simple method.

The remainder of this paper is organized as follows. The proposed fragile watermarking scheme is explained in Section 2. Sub-sections of Section 2 describe the generation of the watermark and procedures for its insertion and verification. Experimental results and conclusions are given in Sections 3 and 4, respectively.

2 Proposed Fragile Watermarking Scheme

The proposed fragile watermarking scheme is shown in Fig. 1. The scheme is divided into two parts: (a) generation of authentication signature of the image and (b) subsequent verification of the authentication information of the suspected image with the extracted authentication signature. The two parts are discussed briefly below.



(a)

(b)

Fig. 1. (a) Generating and embedding the watermark, (b) Verification scheme

2.1 Generation of Watermark and Insertion Procedure

First, an authentication signature is extracted from the image's spatial domain as follows. The image is divided into non-overlapping *8* x *8* blocks, which are permuted in the random order generated by the pseudo-random number generator (PRNG) using the seed (key_1) that only authorized users know. Next, pixels are selected randomly from each block; *64* pixels use the seed (key_2). We divide them into *2* parts (Bn_a, Bn_b), with *32* pixels each. An average gray value for the *32* pixels in each of the two parts is obtained. The averages are compared with that of the adjoining block. In other words, of the average pixel values of the two parts, one (Bn_a) is compared with the average pixel value of the one area ($Bn\text{-}1_b$) of the block located in front of the block from the randomly permuted *8* x *8* block, and the average pixel value of the other (Bn_b) is compared with the average pixel value of the one area ($Bn+1_a$) of the block after it. Average pixel values of the two blocks are compared using signs of (in) equality (>=, <), and the difference between the two pixel values can range from *0* to *255*. Here, we have categorized the differential values into 4 levels. The results of the comparison (>=, <) and the differential is used to create a watermark (authentication signature). In order to maintain the authentication signature obtained from the image during the image compression processing, we modify the gray values of each area within a range that does not damage the original image. If the difference between the average pixel values of the two blocks falls around the boundary values categorized into 4 levels, we modify the pixel values so that the difference is no longer in the range of the boundary value. By adjusting the gray values of each block according to an appropriate threshold, visibility of the image is maintained while enabling extraction of the identical authentication signature for later image verification.

The authentication information generated with this procedure is a fragile watermark, which is inserted into the frequency domain as follows. The *8* x *8* block image is translated into DCT. Next, we insert the magnitude relation acquired in the spatial domain between parts of the adjoining blocks and the differential of the average pixel values into the middle band of the DCT block (*Bn*). In other words, coefficients from 5 groups, *4* from each middle band of the DCT block according to seed (Key_3), are randomly selected. At this time, the 4 groups are modified according to the 4 levels of average pixel value. Rather than the comparison value with the adjoining block, the magnitude information of the average pixel value of the two parts of each single block is inserted for the remaining group. According to the differential value of the average pixels for the 4 levels, one group is selected among 4 DCT coefficient groups. From the selected group, two out of four coefficients are chosen, and the sum of the coefficients is obtained. If the magnitude relation of the sum of the coefficients corresponds with the magnitude relation of the average pixel values of the block areas, the coefficient values are not modified. Otherwise, each coefficient value is modified so that the magnitude relations coincide. At this time, the quantity of the modified coefficients should not change the authentication information acquired in the spatial domain, and maintain the inserted information for image compression processing. Considering visibility of the image, the authentication information is inserted into the middle band of the DCT block.

2.2 Verification Procedure

The verification method for an image is similar to the generation and insertion procedures of the authentication signature. From the spatial domain of a suspect image, an authentication signature is obtained using the identical seed (*key_1*) used during the generation of watermark. Next, each block is translated to the DCT, and the magnitude relationships are obtained between the sums of coefficients by locating the DCT coefficients divided into 5 groups according to the average pixel differential values of the adjoining block. The magnitude relationship between blocks acquired from the image's spatial domain and the magnitude relationship of the coefficient sums obtained from the DCT block are compared to determine whether the block has been manipulated. A block (*Bn*) is determined to have been manipulated if adjoining blocks on both sides are compared and any one of them is different. For example, if the magnitude relationship, (Bn_b, $Bn{+}1_a$), of the average pixel values between the rear (right) block extracted from the spatial domain and the magnitude relationship of the coefficient sums obtained from the block's DCT domain are different, or the magnitude relationship, (Bn_a, $Bn{-}1_b$), of the average pixel value between the front (left) block and the magnitude relationship of the coefficient sums obtained from the front (left) block (*Bn-1*) DCT domain are different, the block, (*Bn*), is determined to have been manipulated. However, blocks in the following cases are not considered to have been manipulated. From the verification scheme, if the blocks located at both ends are modified from the randomly permuted block order, the middle block will be designated as manipulated. In order to resolve this problem, the middle block is extracted from what is determined to be *3* consecutive manipulated blocks from the randomly permuted blocks. After returning the block to the original position prior to the random permutation, if the right/left/top/bottom blocks are not manipulated, the block is considered innocent. In an image authentication system, detection results according to malicious manipulations generally focus on a particular area, whereas incidental manipulations, including compression (i.e., filtering and/or sharpening), are indicated throughout the image [11]. Furthermore, the magnitude relationship of the average pixel values in the *2* parts within a single block is compared with the corresponding magnitude relationship of the DCT coefficient sums. If they do not match each other, the block is ultimately determined to have been manipulated.

The proposed image authentication scheme maintains security by using pixel locations divided into the adjoining block and 2 parts within a single block, as well as the seed (*key*) that allows only the authorized users to know the location of coefficients within the DCT block. Moreover, the authentication signature is generated in the image's spatial domain, and the watermark inserted into the image's frequency domain is fragile to malicious modification but robust to JPEG lossy compression. In addition, the watermarked image maintains such visibility that it cannot be distinguished from the original image upon visual comparison. Below, performance analysis is given, and experimental results are presented.

3 Experimental Results

We tested our fragile watermarking scheme with over 100 images. The size of the images used in the experiment was *512* x *512*. The gray value was modified to *5* to maintain the average pixel differential value among each partial block. In other words, if the differential value with the adjoining partial blocks is near the boundary value that is divided into 4 levels, 5 is added to or subtracted from each pixel value. The difference of the sum of coefficient within the DCT block is adjusted to be around 55 as the absolute value of coefficient.

Referring to an experiment on various compression ratios of several images, Table 1 shows the number of error blocks detected for each image's total number of blocks (4096 blocks). In the experiment, images had either less than two or no detected error blocks in JPEG-quality factor 60. Based on the experimental results, we have concluded that our fragile watermarking scheme can be practically applied in image authentication applications.

Table 1. Number of error blocks against JPEG lossy compression

JPEG Compression (QF)	**Image**					
	Lena	**Barbara**	**Baboon**	**Bridge**	**Goldhill**	**Girl**
90%	0	0	0	0	0	0
80%	0	0	0	0	0	0
70%	0	0	0	0	0	0
60%	0	1	0	0	0	2
50%	36	82	25	46	8	155

Fig. 2 shows the detection results for a manipulated image. The original image is shown in Fig. 2(a), and Fig. 2(b) shows the watermarked image, for which the PSNR is 37.8dB. Fig. 2(c) shows the JPEG-compressed image (QF = 60%), and Fig. 2(d) shows the manipulated watermarked image. In Fig. 2(d), the replaced part of the image is the flower attached to the Lena’s hat. Fig. 2(e) shows the detection result when the attack is object placement only, and Fig. 2(f) is the detection result of the JPEG-compressed image

(QF = 60%) after being manipulated. As can be seen from the experiment results, tampered regions are sufficiently identifiable although the tampered shape might not be indicated in detail.

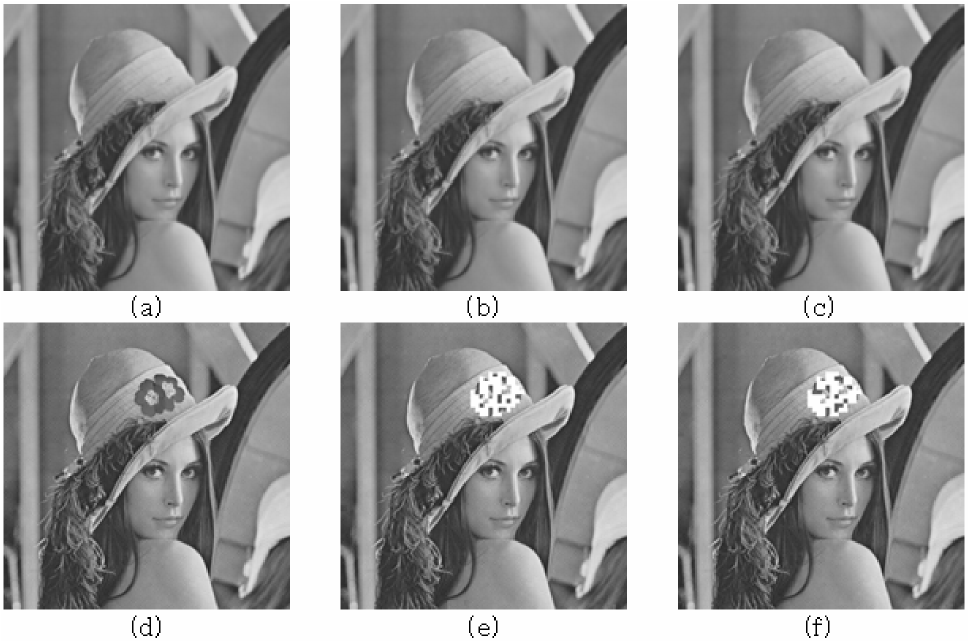


Fig. 2. Detection result of a maliciously manipulated image. (a) is the original image, (b) is the watermarked image (PSNR = 37.8dB), (c) is the JPEG-compressed image (QF = 60%), (d) is the manipulated watermarked image, (e) is the detection result of the manipulated image only, and (f) is the detection result of the JPEG-compressed image (QF = 60%) after manipulation.

4 Conclusions

In this paper, a new image authentication scheme has been proposed, which involves simply extracting an image feature that is robust to JPEG compression from the spatial domain of the image and then inserting it into the frequency domain. The watermark, which is extracted from the image and embedded into the DCT block, is resistant to JPEG compression but is fragile to malicious manipulation.

Future work is needed to ensure that the embedded watermark remains robust at low JPEG compression quality factors and that detection of tampering is indicated more specifically.

References

1. M.Schneider, S.F.Chang: A robust content based digital signature for image authentication. In Proc. IEEE ICIP (1996) 227-230
2. C.Y.Lin, S.F.Chang: A robust image authentication method surviving JPEG lossy compression. In Proc. SPIE Storage and Retrieval of Image/Video Database, San Jose (1998)
3. S.Bhattacharjee, M.Kutter: Compression tolerant image authentication. In Proc. IEEE Int. Conf. on Image Processing (1998) 435-439
4. C.S.Lu, H.M.Liao: Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme. Proc. ACM Multimedia and Security Workshop at the 8th ACM Int. Conf. on Multimedia, Los Angeles, California, USA (2000) 115-118
5. D.Kundur, D.Hatzinakos: Digital watermarking for telltale tamper proofing and authentication. In Proc. IEEE ICIP (1997) 1167-1180
6. M.Yeung, F.Mintzer: An invisible watermarking technique for image verification. In Proc. IEEE Int. Conf. on Image Processing (1997) 680-683
7. M.Wu, B.Liu: Watermarking for image authentication. In Proc. IEEE Int. Conf. on Image Processing (1998) 437-441
8. P.W.Wong: A public key watermark for image verification and authentication. In Proc. IEEE Int. Conf. on Image Processing (1998) 455-459
9. D.Zou, C.W.Wu, G.Xuan, Y.Q.Shi: A content-based image authentication system with lossless data hiding. In Proc. ICME Int. Conf. on Multimedia and Expo (2003) 213-216
10. C.Y.Lin, S.F.Chang: A robust image authentication method distinguishing JPEG compression from malicious manipulation. IEEE Trans. on Circuits and Sys. of Video Tech. (2001) 153-168
11. C.S.Lu: Multimedia Security. IGP, Hershey London (2005)