Halftoning-based Self-embedding Watermarking for Image Authentication and Recovery

Jose Antonio Mendoza-Noriega, Brian M. Kurkoski
Dept. of Information and Communication Engineering
The University of Electro-Communications (UEC)
Chofu-Shi, Tokyo, Japan

Mariko Nakano-Miyatake, Hector Perez-Meana
Mechanical and Engineering School-Culhucan Campus
National Polytechnic Institute of Mexico (IPN)
Mexico City, Mexico
mariko@infinitum.com.mx, hmperezm@ipn.mx

***Abstract*— This paper presents a block-wise semi-fragile watermarking algorithm for image content authentication, with tamper region localization and recovery capability. A halftone image is generated by the error diffusion halftoning method and embedded using the Quantization Index Modulation (QIM) method in the Discrete Cosine Transform (DCT) domain of the original image. The proposed method is robust to JPEG compression, because the halftone image is embedded as a watermark sequence in the middle frequencies of the DCT coefficients using QIM. Also to improve the recovered image quality, Multilayer Perceptron neural network (MLP) is used in inverse halftoning process. Data in the tampered region is estimated with gray-scale data obtained from the MLP, using the embedded halftone as input. The experimental results show desirable performance of the proposed algorithm, such as watermark imperceptibility, robustness to JPEG compression, detection accuracy of tampered region and high quality of recovered image.**

I. INTRODUCTION

With the growth of Internet, digital images play an important role as evidences in the news and reports in digital media. However, using software tools, digital images can be easily modified without any trace. Generally these altered images can cause economic and social damages to the involved persons. Therefore the development of a reliable digital image authentication scheme is an urgent issue. Among several approaches, a watermarking based approach is considered as a possible solution.

The watermarking-based authenticators can be classified into two schemes: fragile watermarking based schemes [1] and semi-fragile watermarking based schemes [2,3]. Fragile watermarking schemes are used for complete authentication, in which only images without any modification are considered as authentic. While the semi-fragile watermarking schemes can be used for content authentication, in which the authenticator distinguishes between images altered intentionally to tamper the image contents, and images suffering content-preserving modification which must be considered as authentic. Therefore content authentication must be robust to content-preserving modification, such as JPEG compression with reasonable compression rate.

Many content authentication methods determine if the image has been tampered or not, and some of them can localize the tampered regions [2]; however, only a few schemes have the capability to recover the tampered region without using the original image [3-6]. In [3] the image is divided into sub-blocks and a mapping list between sub-blocks is generated by a secret key. Here the watermark bits sequence is formed by a compressed version of an image block, which is extracted from the quantized DCT coefficients, and then it is embedded into two LSB's of the corresponding image block. This method is classified as a fragile watermarking scheme, because spatial LSB plane is vulnerable to non-intentional modifications, such as image compression, contamination by noise, etc. In [4], authors used halftone representation of the original image as watermark sequence and embedded it into LSB plane of the image. Because embedding domain is spatial LSB, also this scheme is not robust to JPEG compression. In [5] a hybrid block based watermarking technique, which includes robust watermarking scheme for self-correction and fragile watermarking scheme for sensitive authentication, is proposed. In this scheme all alterations, including the content-preserving modification, are detected and the recovery mechanism is triggered; therefore the quality of the final recovered image can be affected. To increase watermark robustness, [6] and [7] introduced a concept of region of interest (ROI) and region of embedding (ROE), and the original image is segmented into these two regions. Information of ROI is embedded into ROE in DCT domain. In this scheme, the size of ROI is limited for correctly operation, and for some types of images, the segmentation of ROI and ROE can not be done in advance.

In this paper, an image authentication and recovery scheme is presented. It relies on embedding a halftone of the image into the image itself, using QIM, as described in Sec. II-A. Unlike [4], which used LSB embedding method, in this

This research project was collaboration between UEC and IPN. The project was sponsored by JASSO, CONACyT and ICyTDF of Mexico.

paper the watermark sequence is embedded into the DCT domain to increase watermark robustness. At authentication time, the halftoning process is repeated; if this halftone matches the embedded halftone, the images are declared authentic, as described in Sec. II-B. However, if they do not match, then a region of modification can be identified. Recovery inside this region is performed by using a neural network to estimate the original grey-scale data from the embedded halftone. The neural network is trained using the suspicious image. This recovery process is described in Sec. II-C. The recovery effectiveness of this algorithm is demonstrated in Sec. III, and the paper is concluded in Sec. IV.

II. PROPOSED ALGORITHM

The proposed authentication algorithm has three stages: self-embedding, authentication and recovery stage. The block diagrams of these stages are shown in Figs. 1 and 2.

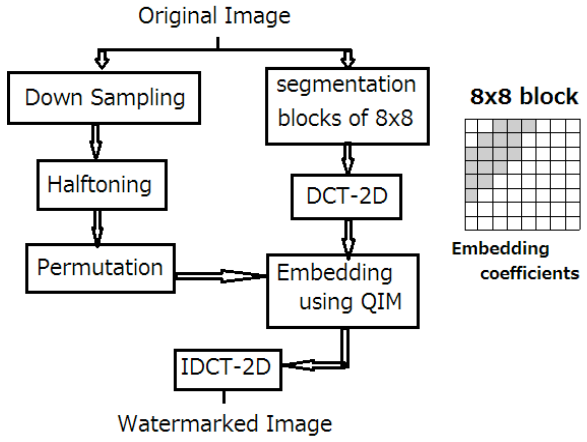

Figure 1. Self-embedding stage of the proposed algorithm

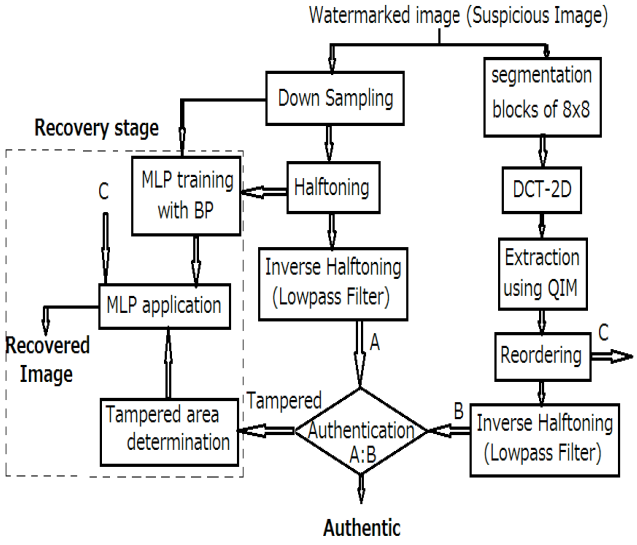

Figure 2. Authentication and recovery stage of the proposed algorithm

A. Self-embedding Stage

In the self-embedding stage, to generate watermark sequence, the original image is down-sampled with half size in height and width. The error diffusion halftoning method proposed by Floyd-Steinberg is applied to the down-sampled image; however, any halftoning method can be used. Using a user's secret key, the halftone image is permuted. There are many methods to realize the permutation, and here we used the chaotic mixing method [8]. On the other hand, the original image is segmented into 8x8 pixels blocks and each of them is transformed by the 2D-DCT. The 16 watermark bits are embedded into middle frequency range (shadowed coefficients in Fig.1) of each block by QIM method [9]. This embedding algorithm is given by :

$$w_k = \begin{cases} 0 & \tilde{c}_{i,j} = 2 * Q * round\left({}^{c_{i,j}}/_{Q}\right) \\ 1 & \tilde{c}_{i,j} = 2 * Q * round\left({}^{(c_{i,j}-1)}/_{Q}\right) + Q \end{cases}, \quad (1)$$

where w_k is k-th watermark bit, $c_{i,j}$ and $\tilde{c}_{i,j}$ are the original and watermarked DCT coefficients, respectively, and Q is the quantization step size. Finally applying inverse 2D-DCT to each watermarked block, a watermarked image is obtained.

B. Authentication Stage

In the authentication stage, firstly the watermark sequence is extracted in DCT domain of the suspicious image, and the extracted bits sequence is reordered using same secret key used in embedding stage. The watermark extraction process is given by :

$$\hat{w}_k = \begin{cases} 0 & \text{if } round\left({}^{\tilde{\tilde{c}}_{i,j}}/_{Q}\right) = \text{even} \\ 1 & \text{if } round\left({}^{\tilde{\tilde{c}}_{i,j}}/_{Q}\right) = \text{odd} \end{cases}, \quad (2)$$

where $\hat{w}_k$ is extracted watermark bit, and $\tilde{\tilde{c}}_{i,j}$ is DCT coefficient of the watermarked and possibly tampered image. Q is the same quantization step size used in embedding stage. The reordered watermark sequence is the halftone version of the original image and then it is converted to gray scale image using a low-pass filter based inverse halftoning.

A halftone image is generated from the suspicious image and also it is reconverted to a gray-scale image using the same low-pass filter. Low-pass filter-based inverse halftoning is the simplest method, although it produces a low quality gray-scale image. However in this stage, accurate detection of the tampered region is important and high quality gray-scale image is not required. Then both gray-scale images are compared each other to detect tampered region. The comparison is carried out in each block and the mean square

error (MSE) of each block given by (3) is compared with the predetermined threshold value *Th*,

$$D = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(A(i,j) - B(i,j)\right)^2 \quad (3)$$

where A and B are blocks of gray-scale image in Fig 2, respectively. If $D \geq Th$ then the block is considered as tampered, otherwise the block is authentic.

C. Recovery Stage

If the authentication stage determines that the suspicious image has been tampered, the recovery stage is triggered receiving the down-sampled suspicious image, its halftone image, the tampered region information and the extracted halftone image (signal C in Fig. 2) as input data. In this stage, firstly using the down-sampled suspicious image and its halftone version, a Multilayer Perceptron neural network (MLP) is trained by using the Backpropagation (BP) algorithm. The neighborhood template, shown by Fig. 3, composed of 16 binary pixels, including the center pixel 'o' is used to get an input pattern of MLP. The desired output data is a corresponding gray-scale value of the center pixel. The extracted halftone image of the tampered region is introduced to the previously-trained MLP to get a high quality gray-scale recovered region.

a	a	a	a
a	a	a	a
a	a	o	a
a	a	a	a

Figure 3 Neighborhood template used to generate input pattern of MLP, here 'o' is the center pixel.

Generally MLP-based inverse halftoning is not useful, because the gray scale image is not available, however in this case, the non-tampered region of the suspicious image is available and then, using this region, a high quality gray scale image can be generated. Fig 4 shows a comparison between a gray-scale image generated by using our MPL method and a Gaussian low-pass filter based method used in [4].

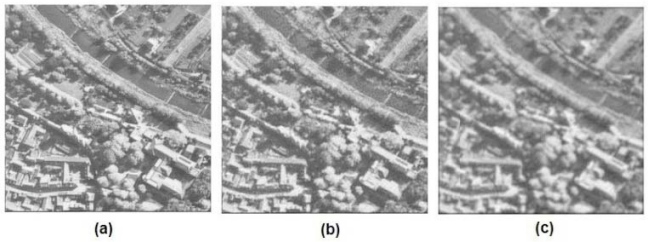

Figure 4 Image quality comparison. (a) original image. (b) gray scale images generated by MLP method. (c) gray-scale image by Gaussian low-pass filter.

The PSNRs of both images respect to the original one are 28dB and 24.5dB, respectively. The gray-scale image generated by MLP conserves the details of the original image.

III. EXPERIMENTAL RESULTS

To evaluate the proposed scheme from the watermark imperceptibility and robustness points of view, computer simulations are carried out. To set adequate value for the step size of the QIM algorithm is very important, because this value has serious effect on previously mentioned requirements: watermark imperceptibility and robustness. Fig. 5 shows the relationship between watermark imperceptibility and the step size, and the relationship between quality factor of the JPEG compression and BER of the extracted watermark bit sequence respect to the original halftone image, in which the performance of different step sizes are compared.

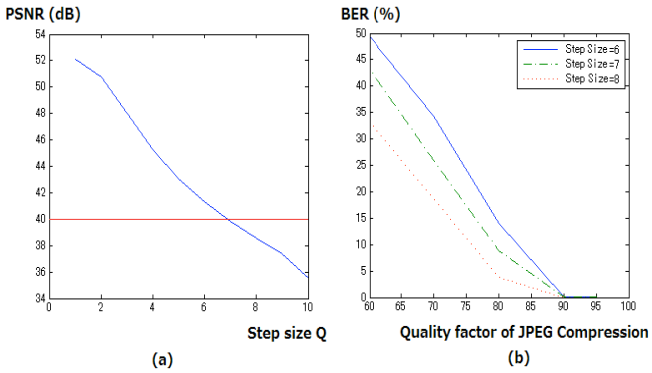

Figure 5 (a) Relationship between step size and PSNR of watermarked image respect to the original one. (b) Relationship between quality factor of JPEG compression and BER of extracted watermark sequence respect to the original halftone image.

Taking account of the watermark imperceptibility and robustness, from Fig. 5, the step size is set equal to 7. Also to obtain good performance of tampered region detection, the selection of an adequate threshold value is very important. Fig. 6 shows the relationship between the false alarm error rate *(Fa)* and threshold value when the watermarked image is compressed by JPEG with quality factor 80. From the figure, the threshold value 0.01 is considered as the best value. Fig. 7 shows the original image and watermarked one generated by the proposed algorithm using step size equal to 7. Here, the average PSNR of the watermarked image respect to the original one is 39.9dB. Figs 8 and 9 show the tampered region detection and recovery capability of the proposed algorithm. From these figures it follows that tampered regions are detected and recovered correctly.

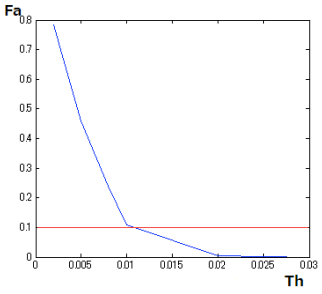

Figure 6 Relationship between threshold value and false alarm error rate (step size=7).

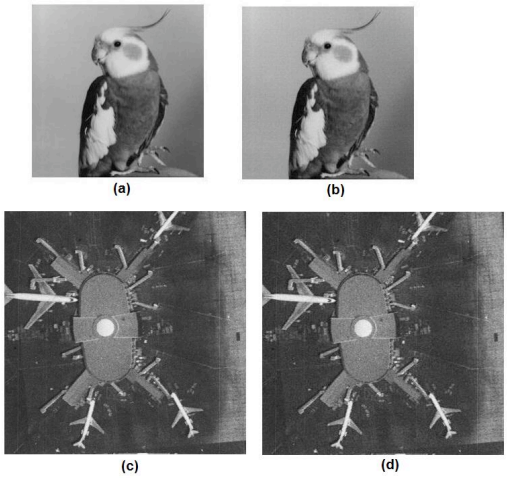

Figure 7 (a), (c) Original images. (b), (d) Watermarked images.

IV. CONCLUSIONS

In this paper, an image authentication algorithm with recovery capability is proposed, in which a halftone version of the original image is embedded as watermark sequence. In embedding stage, QIM method is applied in DCT domain, which increases watermark robustness to JPEG compression. Also in the proposed algorithm to increase the recovered image quality, MLP is trained by BP algorithm to estimate a gray scale image from its halftone version. The simulation results show the correct detection of the tampered regions and good quality image.

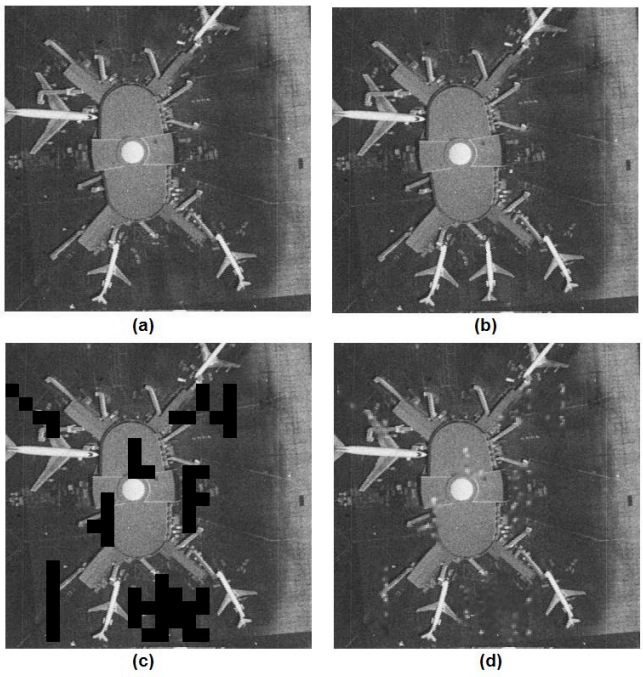

Figure 8 (a) original image, (b) tampered image adding an airplane, (c) tampered region detection and (d) recovered image

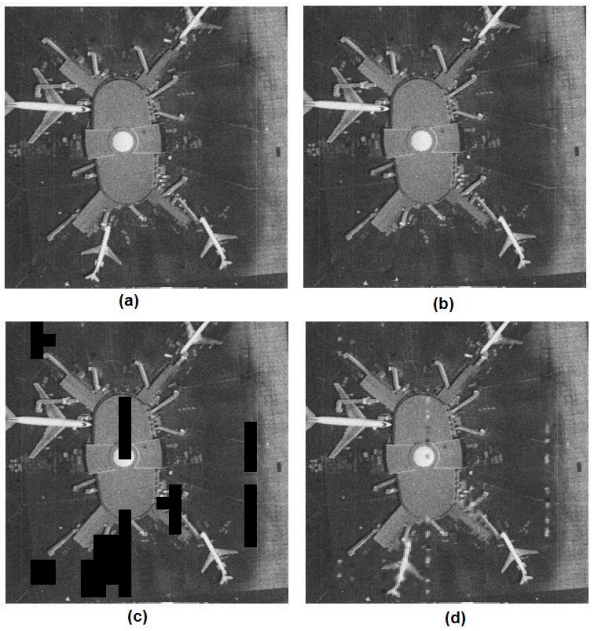

Figure 9 (a) original image, (b) tampered image eliminating an airplane, (c) tampered region detection and (d) recovered image.

REFERENCES

[1] P. Wong, N. Memon, "Secret and Public Key Image Watermarking Schemes for Image Authentication and Ownership Verification", IEEE Trans. Image processing. 10(10), 2001, pp. 1593-1601.

[2] K. Maeno, Q. Sun, S. Chang, M. Suto, "New Semi-Fragile Image Authentication Watermarking Techniques Using Random Bias and Nonuniform Quantization", IEEE Trans. Multimedia, 8(1), 2006, pp. 32-45.

[3] J. Fridrich, M. Goljan, "Image with Self-Correcting Capabilities", 1999 Int. Conf. on Image Processing, vol. 3, pp. 792-796.

[4] H. Luo, S-C Chu, Z-M Lu, "Self Embedding Watermarking Using Halftoning Technique", Circuit Systems and Signal Processing, (2008) 27, 2008, pp. 155-170.

[5] Y. Hassan, A. Hassan, "Tamper Detection with Self Correction on Hybrid Spatial-DCT Domains Image Authentication Technique", Communication systems Software and Middleware Workshops,2008, pp. 608-613.

[6] C. Cruz, J. Mendoza, M. Nakano, H. Perez, B. Kurkoski, "Semi-Fragile Watermarking based Image Authentication with Recovery Capability", ICIECS 2009, pp.269-272.

[7] C. Cruz, R. Reyes, M. Nakano, H. Perez, "Image Authentication Scheme based on Self-embedding Watermarking", LNCS 5856, Springer-Verlag, 2009.

[8] G. Voyatzis, I. Pitas, "Embedding Robust Watermarks by Chaotic Mixing", Int. Conf. on Digital Signal Processing, 1(1), 1997, pp. 213-216.

[9] B. Chen, G. Wornell, "Quantization Index Modulation: A class of provably good methods for digital watermarking and information embedding", IEEE Trans. On Information Theory, 48(4), 2001, pp. 1423-144.