Watermarking-based Image Authentication with Recovery Capability using Halftoning and IWT

Luis Rosales-Roldan, Manuel Cedillo-Hernández, Mariko Nakano-Miyatake, Héctor Pérez-Meana

Postgraduate Section, Mechanical Electrical Engineering School, National Polytechnic Institute of Mexico

Abstract – *In this paper we present a watermarking algorithm for image content authentication with localization and recovery capability of the modified areas. We use a halftone image generated by the Floyd-Steinberg kernel as an approximate version of the host image. We adopt this halftone image as a watermark sequence and embed it using the quantization watermarking method into the sub-band LL of the Integer Wavelet Transform (IWT) of the host image. Due to the fact that the watermark is embedded into the sub-band LL of IWT, the proposed method is robust to JPEG compression. Moreover, we employ a Multilayer Perceptron neural network (MLP) in inverse halftoning process to improve the recovered image quality. Using the extracted halftone image, the gray-scale of the modified area is estimated by the MLP. The experimental results demonstrate the effectiveness of the proposed scheme.*

Keywords: Watermarking, Content Authentication, Recovery Capability, Integer Wavelet Transform, Multilayer Perceptron

1 Introduction

Nowadays the digital age has reached an important development in some technical fields, such as computer and the telecommunications. This development has a strong impact on the people's life, for example it is quite common to take pictures everywhere and every time using his/her cell phones with digital cameras. And also 700,000 pictures per hour are uploaded to any social network to be shared among friends. However these digital pictures can be easily modified using computational drawing tools, such as Photoshop, without causing any distortion. Considering that a scenario where some of these digital pictures could be required as evidence to prove the truth of the statement of a person who is defending his innocence on court, the integrity of these digital images becomes an urgent and important issue.

The Cryptographic Hashing, such as MD5 and SHA-1, have been used to authenticate the digital data, however it cannot be used for digital images in an efficient manner. The main problem is that there are many different formats, such as JPEG, BMP, TIF, PCX and so on, to save a digital image and besides some of them have their different compression mode. For example, an image could be compressed and converted to another format during its distribution. Although image format or compression mode is changed, the content of the image is conserved totally. Taking these aspects under consideration, the Cryptographic Hashing, digital fingerprinting and other techniques, which cannot tolerate the content conserving modifications, are not adequate for image content authentication.

Among several approaches, a watermarking-based approach is considered as a possible solution. Early image authentication methods [1] result in an integrity decision, which indicates only if the image under analysis is authentic or not. The watermarking-based authenticators can be classified into two schemes: fragile watermarking-based scheme [2] and semi-fragile watermarking-based scheme [3, 4]. The fragile watermarking scheme can be used for complete image authentication in which only those images without any modification are considered as authentic. On the other hand, the semi-fragile watermarking scheme can be used for content authentication, in which those images, that are modified no intentionally and conserved its original content, are considered as authentic. Consequently, content authentication scheme must be robust to content-preserving modification, such as JPEG compression.

Many content authentications methods determine if image has been modified or not, and some of them can localize the modified areas [3]; moreover, only a few schemes have the capability to recover the modified area without using original image [4-8]. In [4] they divide the image into sub-blocks and then mapping the sub-block with a secrete key. With this, a watermark bits sequence is formed by a compressed version of an image block, which is extracted for the quantized DCT coefficients, and then it's embedded into two LSB's of the corresponding image block. This method is classified as a vulnerable scheme to non-intentional modifications, such as image compression, contamination by noise, etc. In [5] they used halftone representation of the original image as watermark sequence and embedded it into LSB plane of the image. Due to embedding domain is spatial LSB; also this scheme is not robust to JPEG compression. In [6] the authors proposed a hybrid block-based watermarking technique, which includes robust watermarking scheme for self-correction and fragile watermarking scheme for sensitive authentication. 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, [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 con not be done in advance. Due to the fact that the quality of the image is important for further process, in [8] the authors proposed a new watermarking method consisting of the detection and recovery of the modified areas. They used a halftone image from the original one as a watermark sequence and embedded it into the Discrete Cosine Transformation (DCT) using the Quantization Index Modulation (QIM). The QIM applied in DCT domain makes their method be robust to JPEG compression. Also they used a Multilayer Perceptron neural network (MLP) to obtain a high quality of recovered image.

In this paper, we proposed an image authentication and recovery scheme in which a halftone of the original image is embedded as a watermark into the image using quantization-watermarking algorithm. Unlike [8], which used DCT embedding method, in the proposed scheme the watermark sequence is embedded into the sub-band LL of the Integer Wavelet Transform (IWT) domain to increase watermark robustness. We used a similar process of the embedding process in the authentication stage, if the extracted halftone image matches with the embedded one, the image is declared as authentic, otherwise the altered area can be detected and then the recovery process is started to estimate the original gray-scale image of the altered area from the extracted halftone image using the previously trained MLP.

The rest of this paper is organized as follows. Section 2 describes the proposed algorithm and experimental results are presented in Section 3. Finally Section 4 concludes this work.

2 Proposed Algorithm

The proposed authentication algorithm is composed by three stages: self-embedding, authentication and recovery stage.

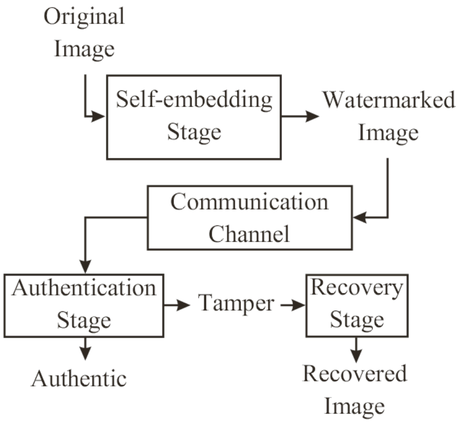


Figure 1. General scheme of the proposed algorithm.

2.1. Self-Embedding Stage

In the self-embedding stage, the original image is down-sampled with half size in height and width to generate the watermark sequence. Then we applied the error diffusion halftoning method proposed by Floyd-Steinberg to the down-sampled image to get halftone image. The halftone image is permuted by the chaotic mixing method [9] using user's secret key. On the other hand, the original image is decomposed using the IWT to obtain four sub-bands: LL, LH, HL and HH. The permuted halftone image is embedded into the sub-band LL using quantization watermarking method [10]. The embedding algorithm is given by:

$$\mathrm{w_k} = \begin{cases} \tilde{c}_{ij} = v_1 & \text{if } \left|c_{ij} - v_1\right| \le \left|c_{ij} - v_2\right| \\ \tilde{c}_{ij} = v_2 & \text{otherwise} \end{cases} \qquad (1)$$

where

$$v_1 = \begin{cases} sign\left(c_{i,j}\right) \times \left\lfloor \frac{\left|c_{ij}\right|}{2S} \right\rfloor \times 2S, & w_k = 0 \\ sign\left(c_{i,j}\right) \times \left(\left\lfloor \frac{\left|c_{ij}\right|}{2S} \right\rfloor \times 2S + S \right), & w_k = 1 \end{cases}$$

$$v_2 = v_1 + sign\left(c_{ij}\right) \times 2S$$

and w_k is the k-th watermark bit, $c_{i,j}$ and $\tilde{c}_{i,j}$ are the original and the watermarked IWT coefficients, respectively, and S is the quantization step size. Finally we obtained the watermarked image applying inverse IWT to the watermarked LL sub-band and the rest of the sub-bands (LH, HL and HH). This stage is shown in Figure 2.

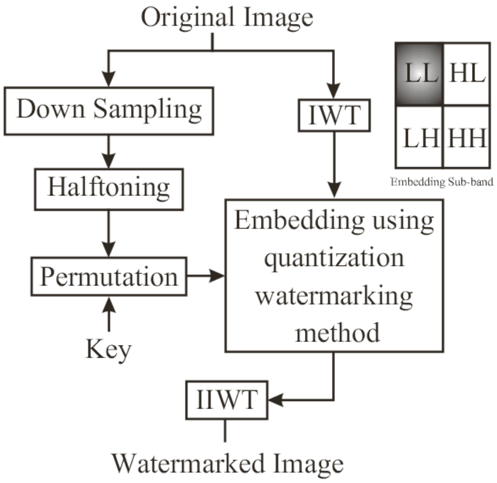


Figure 2. Self-embedding stage.

2.2. Authentication Stage

In the authentication stage (see Figure 3) firstly the watermark is extracted for the sub-band LL of the suspicious image and the extracted bits are reordered using the user's secret key given in the embedding stage. The watermark extraction process is given by:

$$\tilde{w}_k = \begin{cases} 0 & \text{if } round\left(\hat{c}_{ij} \middle/ S \right) = even \\ 1 & \text{if } round\left(\hat{c}_{ij} \middle/ S \right) = odd \end{cases} \tag{2}$$

where $\tilde{w}_k$ is extracted watermark bit, and $\hat{c}_{i,j}$ is IWT coefficient of LL sub-band of the watermarked and possibly modified image. S 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 Gaussian low-pass filter given by (3).

$$F_G = \frac{1}{11.566} \begin{bmatrix} 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.4035 & 0.7970 & 1 & 0.7970 & 0.4035 \\ 0.3215 & 0.6352 & 0.7970 & 0.6352 & 0.3215 \\ 0.1628 & 0.3215 & 0.4035 & 0.3215 & 0.1628 \end{bmatrix} \tag{3}$$

Next we generate a halftone image from the suspicious watermarked image and it is re-converted in a gray-scale image using the same Gaussian low-pass filter. This inverse halftoning is the simplest method, even though it produces low quality gray-scale image. In this stage, an accurate detection of the modified areas is important; therefore high quality of the gray-scale image is not necessary. Then both images (gray-scale image generated from the extracted watermark sequence and gray-scale image generated from suspicious watermarked image) are compared each other to localize the modified areas. To do this we employed a block-wise strategy, in which the comparison is carried out in each block of NxN pixels and the mean square error (MSE) of each block is calculated by (4) and it is compared with a predetermined threshold value Th.

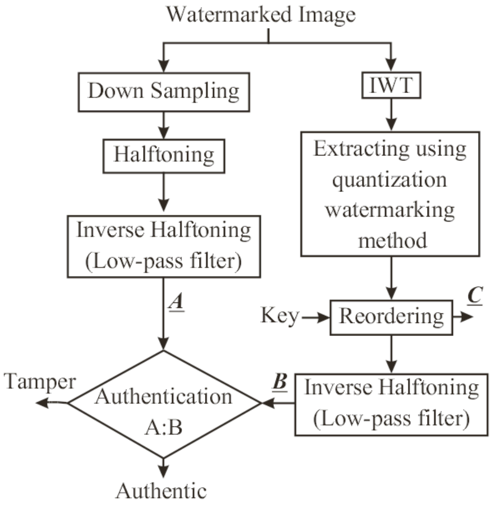

Watermarked Image
Down Sampling
IWT
Halftoning
Extracting using quantization watermarking method
Inverse Halftoning (Low-pass filter)
A
Key
Reordering
C
Tamper
Authentication A:B
B
Inverse Halftoning (Low-pass filter)
Authentic

Figure 3. Authentication stage.

$$D = \frac{1}{N_2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(A(i,j) - B(i,j) \right)^2 \tag{4}$$

where A and B are the blocks of gray-scale image in Figure 3, respectively, and NxN is a block size. If $D \geq Th$ the block is considered as tampered, otherwise the block is authentic.

2.3. Recovery Stage

If the authentication stage shows that some blocks of the suspicious image are tampered, then the recovery stage will be triggered. In this stage we will use as input data, the down-sampled suspicious watermarked image, its halftone version, the information about modified blocks and the extracted halftone image (signal C in Figure 3). In this stage we firstly use the down-sampled suspicious image and its halftone version to train MLP by the Backpropagation (BP) algorithm. This recovery stage is shown in Figure 4 and the MLP used to estimate the gray-scale image is shown in Figure 5.

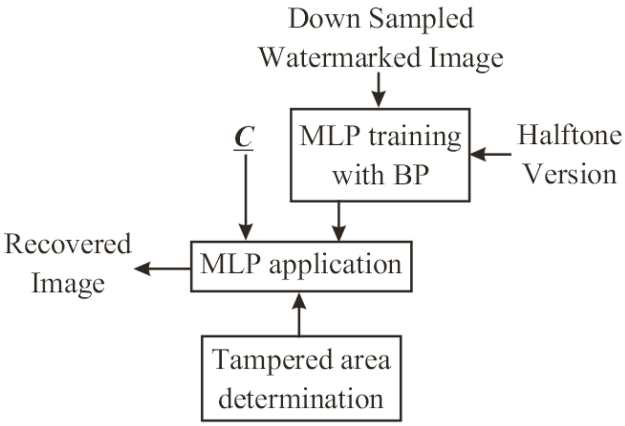
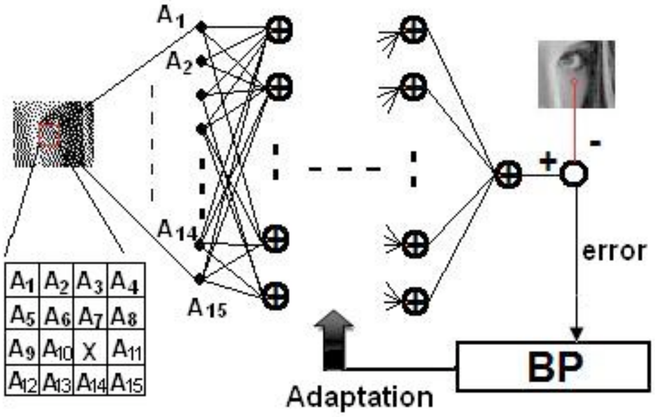


Figure 4. Recovery process.



A1	A2	A3	A4
A5	A6	A7	A8
A9	A10	X	A11
A12	A13	A14	A15

Figure 5. MLP used to estimate the gray-scale image

The 4x4 neighborhood template, show in bottom-left part of the Figure 5, composed of 16 binary pixels including the center pixel "X", is used to get an input pattern of MLP. The output data is a gray-scale estimated value of the corresponded center pixel "X". The extracted halftone image of the modified area is introduced to this MLP to get a better quality of the recovered region.

In the general case of inverse halftoning, the gray-scale image is not available, therefore the MLP-based inverse halftoning in meaningless, however in this case the non-modified area of the suspicious gray-scale image is available. So we can use the halftone and the corresponded gray-scale image of this non-modified area to generate a high quality image using MLP-based inverse halftoning. Figure 6 shows a comparison between images obtained using Gaussian low-pass filter and MLP.

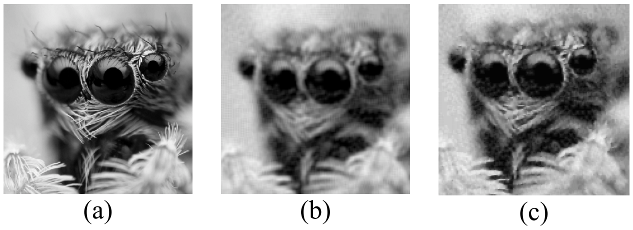


Figure 6. Image quality comparison. (a) Original Image. (b) Gray-scale image by Gaussian low-pass filter (24dB). (c) Gray-scale image by MLP (27dB).

The PSNR of both images respect to the original one are 24 dB and 27 dB, respectively, which indicates that the image generated by MLP can conserve more details of the original image than the gray-scale image generated by a Gaussian low-pass filter.

3 Experimental Results

To evaluate the performance of the proposed watermarking scheme, the watermark imperceptibility and robustness are assessed using several images. It is very important to select an adequate value of the quantization step size used in the embedding algorithm, because this value has serious effects on the watermark imperceptibility and robustness. In Figure 7 we show the relationship between watermark imperceptibility and the quantization step size for each sub-band decomposed by IWT. As we can see in the Figure 7, highest sub-band HH shows better watermark imperceptibility compared with other sub-bands. Furthermore the lowest sub-band LL can be used as watermark embedding domain if the step size is lower than 7 from watermark imperceptibility point of view. Considering the watermark robustness, we select the lowest sub-band LL as watermark embedding domain together with step size value 7.

Also, in Figure 8 we show the relationship between quality factor of JPEG compression and BER of the extracted watermark sequence respect to the embedded one. In which the performance of different step sizes are compared. In all cases, the watermark sequence is embedded in the lowest sub-band LL. From Figures 7 and 8, we select the value 7 for the quantization step size. Also the selection of the threshold value *Th,* to determine if an area is altered or not, is very important. Figure 9 shows the relationship between the false alarm error rate (*Fa*) and the threshold value when the watermarked image is compressed with JPEG compressor using a quality factor 80. From this figure, the threshold value 0.001 is considered as the best one.

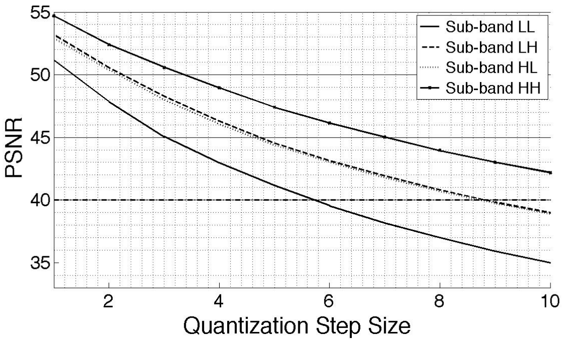


Figure 7. Relationship between quantization step size and PSNR of watermarked image respected to the original one.

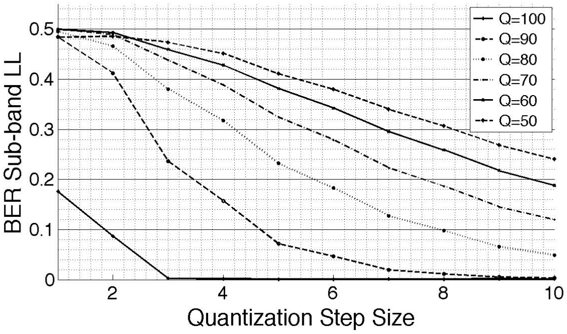


Figure 8. Relationship between quality factor of JPEG compression and BER of extracted watermark respected to the halftone original image.

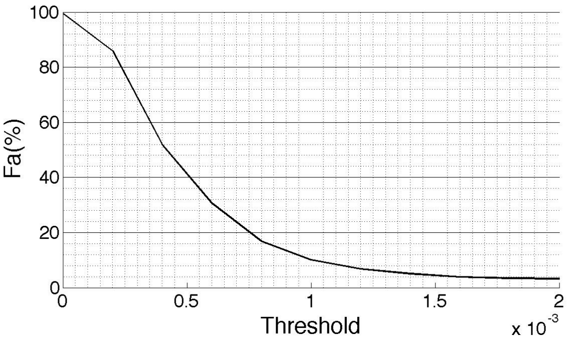
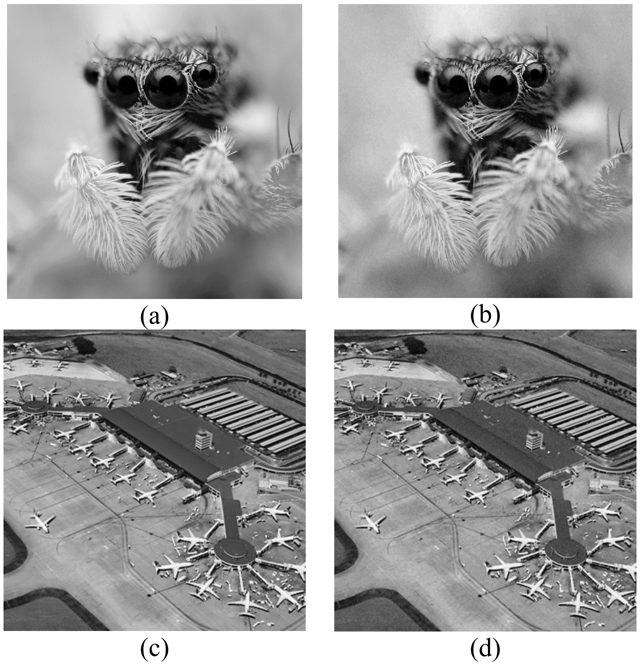


Figure 9. Relationship between threshold value and false alarm error rate with quantization step size equal to 7.

Figure 10 shows the original and the watermarked image generated by the proposed algorithm using step size equal to 7. Here, the average PSNR of the watermarked images respect to their original one is 38.17 dB.


(a) (b)

(c) (d)

Figure 10. (a), (c) Original Images. (b), (d) Watermarked Images.

Now, the Figure 11 shows an example with modified area; in this case we add extra object to the image and the proposed algorithm is applied to detect and recover the modified area. Figure 12 shows another example with different modified area. In this case we erased an object from the image and the proposed algorithm is applied to detect and recover the modified area. From these figures, the modified areas are detected and recovery correctly.

4 Conclusions

In this paper, an image authentication algorithm with recovery capability is proposed, in which a halftone version of the original image is used as a watermark sequence and it is embedded using quantization watermarking method into the LL sub-band decomposed by the IWT.

Important factors, such as the step size value of the embedding algorithm and the threshold value used in the authentication process are estimated taking into account the watermark imperceptibility, robustness and false alarm error rate. The average PSNR of several watermarked image respect to their original versions using an adequate step size value indicates that the embedded watermark is imperceptible by Human Visual System. Also simulation results showed that the embedded watermark is robust to JPEG compression with a quality factor larger than 80%.

The use of the MLP trained by BP algorithm increases the quality of the recovered image and the simulation results showed that the proposed method can detect and recover correctly the modified areas.

5 References

[1] J. Dittmann, “Content-fragile Watermarking for Image Autehntication”, Proceedings of SPIE, vol. 4314, 2001, pp. 175-184.

[2] P. Wong, N. Memon, “Secret and Public Key Imagen Watermarking Scheme for Image Authentication and Ownerchip Verification”, IEEE Trans. Image Processing, vol.10, no.10, 2001, pp. 1593-1601.

[3] K. Maeno, Q. Sun, S. Chang, M. Suto, “New Semi-Fragile Image Authentication Watermarking Techniques Using Random Bias and Nonuniform Quantization”, IEEE Trans. Multimedia, vol. 8, no. 1, 2006, pp. 32-45.

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

[5] H. Luo, S-C Chu, Z-M Lu, “Self Embedding Watermarking Using Halftone Technoque”, Cicuit Systems and Signal Processing, vol. 27, 2008, pp. 155-170.

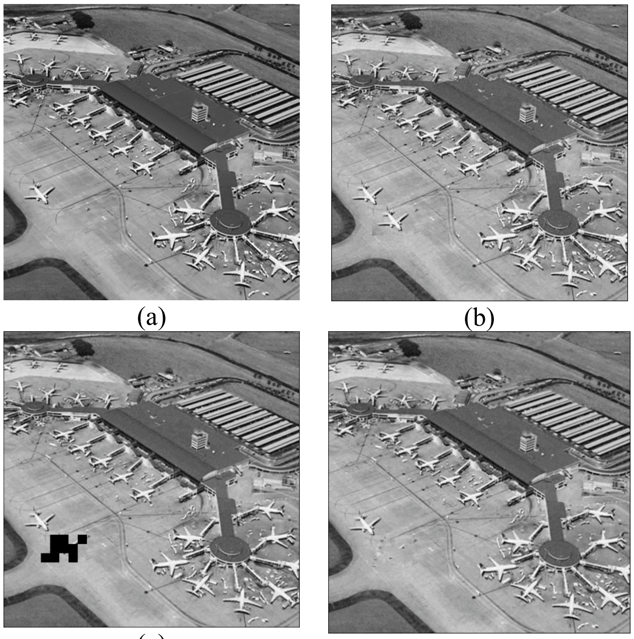
[6] Y. Hassan, A. Hassan, “Tampered Detection with Self Correction on Hybrid Spatial-DCT Domains Image Authentication Technique”, Communication Systems Software and Middleware Workshops, 2008, pp. 608-613.

[7] Clara Cruz, Jose Antonio Mendoza, Mariko Nakano, Hector Perez, Brian Kurkoski, “Semi-Fragile Watermarking based Image Authentication with Recovery Capability”, ICIECS 2009 pp. 269-272.

[8] Jose Antonio Menodza-Noriega, Brian M. Kurkoski, Mariko Nakano-Miyatake, Hector Perez-Meana, “Halftoning- based Self-embedding Watermarking for Image Authentication”, 2010 IEEE Int. 53rd Midwest Symposium on Circuits and Systems, 2010, pp. 612-615.

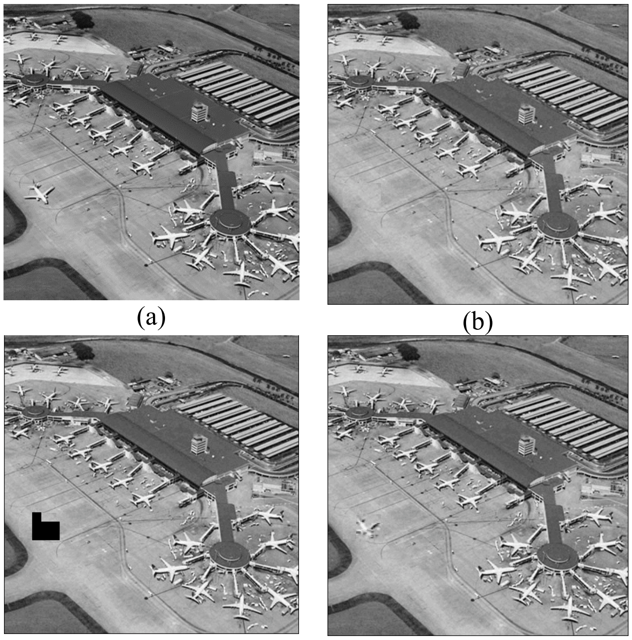
[9] G. Voyatzis, I. Pitas, “Embedding Robust Watermarks by Chaotic Mixing”, Int. Conf. on Digital Signal Processing, vol. 1, no. 1, 1997, pp. 213-216.

[10] B. Chen, G. Wornell, “Quantization Index Modulation: A class of provably good method for digital watermarking and information embedding”, IEEE Trans. On Information Theory, vol. 48, no. 4, 2001, pp. 1423-1444.



(a) (b) (c) (d)

Figure 11. (a) Original Image. (b) Suspicious image, adding extra information. (c) Modified area detection. (d) Recovered image



(a) (b) (c) (d)

Figure 12. (a) Original Image. (b) Suspicious image, extracting some information. (c) Modified area detection. (d) Recovered image.