Semi-Fragile Watermarking based Image Authentication with Recovery Capability

Clara Cruz, Jose Antonio Mendoza, Mariko Nakano Miyatake, Hector Perez Meana
Research and Postgraduate Section, ESIME Culhuacan,
National Polytechnic Institute
Mexico City, Mexico
e-mail mariko@calmecac.esimecu.ipn.mx

Brian Kurkoski
University of Electro-Communications
UEC
1-5-1 Chofugaoka, Chofu-shi
Tokyo, 182-8585, Japan

Abstract—**In this paper, a block wise image authentication scheme using a semi-fragile watermark is proposed. The proposed scheme determines integrity of the image, detecting regions suffered some content modifications and altered region can be recovered without the original image. In the proposed scheme, the original image is divided into two regions: Region of Interest ROI, which is important region that requires protection against malicious modification and Region of Embedding ROE, which is the rest of the image where watermark sequence is embedded. The proposed scheme is evaluated from the several points of view: watermark imperceptibility, accuracy of tamper area detection, quality of recovered regions and robustness to no-intentional manipulations, such as JPEG compression. Experimental results show excellent performance of the proposed scheme.**

Keywords-component; Image authentication; Semi-fragile watermark; recovery capability; self embedding

I. INTRODUCTION

Nowadays the digital images are used as evidence in some situations such like: car crash, politician's scandals, medicals images and others. Under these circumstances, integrity verification has become an important issue in the digital world, because digital images can be tampered easily by third party using computer tools, such as Photoshop®.

Conventionally, the methods used for image authentication can be classified into two classes: digital signature based methods [1], [2], and watermarking based method [3], [4], [5]. A digital signature is a set of features extracted from an image and these are stored in a file. Watermarking, on the other hand, is an image authentication/protection technique that embeds invisible information into an image. Unfortunately, most of the existing watermarking and digital signature based image authentication systems can detect malicious tampering successfully; however there are few system that have recovery capability the tampered region without original image [6],[7]. In [6], a Slant Transform (SLT) based semi-fragile watermarking scheme is proposed. Authors compressed the image using SLT, and the compressed data is replaced by LSB of the image. In [7], both fragile and semi-fragile watermarks are embedded in spatial and frequency domains, respectively. The principal disadvantage of both schemes is fragility of information used for recovery against JPEG compression.

In this paper, an image authentication scheme, with a capability of tampered region localization and recovery, is proposed. In the proposed scheme, an image is segmented by two regions: Regions of Interest (ROI) and Regions of Embedding (ROE). ROI is a region which contains important information and it is required some protection, for example regions of faces of persons involved in some scandal scene, while ROE is rest of the whole image after subtracting region belonged to ROI. ROE can be background of the image. The information of ROI is encoded to generate watermark sequence, and it is embedded into ROE of the same image in an imperceptible manner. In the authentication stage, two watermark sequences, extracted from ROI and ROE respectively, are used. If some blocks of ROI are detected as tampered, the recovery process performs to construct these blocks from the watermark sequence extracted from ROE.

This paper is organized as follows: in section II the proposed authentication method is described, and in Section III the experimental results are provided. Finally conclusions of this paper are described in Section IV.

II. PROPOSED METHOD

A. Watermark sequence generation

Generally in a photo image, some objects or some regions contain information more important than other regions. Therefore we define two regions in the image: region of interest (ROI) and region of embedding (ROE). ROI is important region of the image that requires a protection against malicious modification, while ROE is the rest of the image that no requires any protection. In the proposed algorithm, information of ROI is extracted to generate a watermark sequence and this sequence is embedded into ROE. The proposed watermark generation process can be summarized as follows:

- Subtract 127 from gray levels of the original image to force pixel values to be [-127,128]. It reduces DC-coefficient value after the image is transformed by DCT.

978-1-4244-4994-1/09/$25.00 ©2009 IEEE

- In the original image X, ROI is selected by owner and automatically ROE is determined in order that the following condition is satisfied.

$$ROI \cap ROE = \emptyset \ and \ ROI \cup ROE = X \qquad (1)$$

- ROI region is divided into non-overlapping blocks of 8×8 pixels.
- In each block of ROI, 66 bits watermark sequence is extracted as a following manner.

  a) Compute the 2D-DCT of a ROI block.
  b) The DC-coefficient is rounded and represented by 11 bits (10 bits for absolute value and 1 bit for sign). Because the maximum values of DC for 8x8 block of an image with range [-127,128] is 1016, it can be represented in a binary form using 11 bits, including sign bit.
  c) Taking first lowest 6 AC coefficients in the zig-zag order of the block, encode each one of these coefficients to 8 bits together with 1 sign bit (in total 9 bits).

- The length of watermark sequence of each ROI block is 66 bits, composed by 11 bits of DC-coefficient, 54 bits corresponded to the 6 AC-coefficients of DCT coefficients and finally 1 zero is added in order to divide into 6 segments with 11 bits sequence per segment.

B. Watermark embedding

The proposed watermark embedding process can be summarized as follows:

- Using a user's key K, the mapping list between ROI blocks and ROE blocks is constructed.
- Using this mapping list, each ROI block of 8x8 pixels is mapped into 6 ROE blocks, which are used to embed watermark sequence extracted from the ROI block.
- In each selected 6 ROE blocks, following processes are carried out.
  a) Apply 2D-DCT to 6 ROE blocks.
  b) Quantify by a quantification matrix Q that corresponds to quality factor 70. This value is selected considering tradeoff between watermarked image quality and watermark robustness against JPEG compression. Quantization of DCT coefficients by Q is given by (2)

$$\tilde{C}(u,v) = \lfloor C(u,v)/Q(u,v) \rfloor \qquad (2)$$

  where $C(u,v)$ and $\tilde{C}(u,v)$ are the *(u,v)*-th DCT coefficient and it quantized version, respectively, $\lfloor x \rfloor$ is lower nearest integer value of **x**.
  c) Each 11 bits of watermark sequence is embedded into the LSB of the 11 quantized DCT-coefficients of the middle frequency band of each one of the selected 6 ROE blocks.
  d) The watermarked DCT blocks are multiplied by Q.
  e) It is transformed by the inverse DCT to get watermarked blocks.

- Concatenating all watermarked blocks, the watermarked image is generated.

C. Authentication and Recovery

The authentication procedure verifies if the contents of the received image are authentic or not. To authenticate the image, two watermarks must be extracted and then compared. This authentication and recovery process are described as follows:

- The first watermark W_{ROIext} is generated from the ROI blocks; these operations are same as the watermark generation process before described.
- The second watermark W_{ROEext} is extracted from the ROE blocks. Using the same secret key to construct ROI-ROE mapping lists, the 6 corresponded ROE blocks are determined for each ROI block.
- For selected 6 ROE blocks, the following operations are carried out to get W_{ROEext}
  a) Apply 2D-DCT to each one of 6 ROE blocks.
  b) DCT blocks are quantized by quantification matrix Q.
  c) 11 bits sequence is extracted from LSB of 11 AC coefficients in the middle frequency band of each ROE block.
  d) Concatenated 6 extracted sequences of length 11 bits to generate 66 bits $W_{ROEext.}$

- In the watermark comparison between W_{ROIext} and W_{ROEext}, the tolerant threshold ***Th*** is employed to distinguish a content preserving operation from malicious manipulation. This authenticity check is given by (3).

$$if \sum XOR(W_{ROIext,}W_{ROext}) < Th \ then \ the \ block \ is \ authentic$$
$$if \sum XOR(W_{ROIext,}W_{ROext}) \geq Th \ then \ the \ block \ is \ tampered \qquad (3)$$

Once the authenticity check indicates that a ROI block was tampered, the recovery process of this ROI block is triggered. The recovery process can be summarized as follows:
From the extracted watermark sequence W_{ROEext}, last bit is eliminated to get a watermark sequence with 65 bits.

- Assign the first 11 bits of W_{ROEext} to DC-component and the rest 54 bits are divided into 6 sequences with 9 bits and these are assigned to 6 lowest AC-coefficients of a recovery DCT block.

- Compute the inverse 2D-IDCT of recovered DCT block to get a recovered block.

Replace the tampered ROI block by the recovered one to get recovered image.

III. EXPERIMENTAL RESULTS

We conduct three experiments to evaluate performance of the proposed algorithm, and also the performance of the proposed algorithm is compared with the algorithms Zhao and Hassan [6,7] under same condition. The first experiment is to assess watermark imperceptibility, and in the second one the tamper detection and the recovery capability of the propose algorithm are evaluated. Finally, in the third experiment, the watermark robustness to incidental modification such as JPEG compression is evaluated. Table I shows the values of some factors used during the evaluation. The threshold value th is determined taking false positive error is smaller than 10^{-12}.

TABLE I. PARAMETER'S VALUES USED DURING THE EVALUATION

Number of test images	256-gray level (8 bits/pixel)	100
Length of watermark sequence for each ROI block	W	66 bits
Threshold value	th	18
Number of ROI blocks used	[min, max]	[117,453]

A. Watermarking imperceptibility

The distortion of the watermarked image depends directly on length of the embedded watermark sequence, and the watermark length is related to the number of ROI blocks selected in the image. Since each ROI block requires 6 ROE blocks, the number of ROI blocks is limited by (4).

$$Number(ROI) < NB(I)/7 \quad (4)$$

where *NB(I)* is total number of blocks of the image I. Fig. 1 shows relationship between the percentage of ROI blocks, taking maximum number of ROI blocks as 100%, and average Peek Signal to Noise Ratio (PSNR) of 10 watermarked images. From Fig. 1, the number of ROI blocks can be increased until 30% , keeping watermark imperceptibility.

B. Tamper detection and recovery capability

To evaluate accuracy of tamper detection and recovery capability of the proposed algorithm, the watermarked images were tampered. Fig. 2 shows tamper detection and recovery capability. Fig. 2(a) is the original image, (b) indicates ROI blocks with black areas, (c) is watermarked image, (d) is tampered image, (e) shows detected tampered blocks and (f) is recovered image.

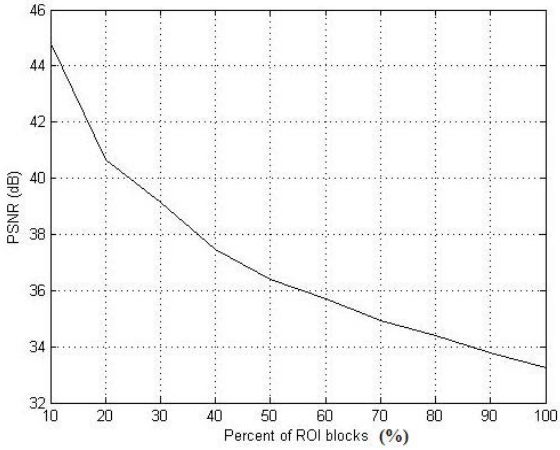

Figure 1. Relationship between number of ROI blocks, taking maximum number of ROI block as 100%, and watermarked image distortion (PSNR).

Figure 2. Tampered blocks detection and recovery results. (a) original image, (b) ROI blocks indicated by black area, (c) watermarked image, (d) tampered image, (e) tampered blocks detection by black area and (f) recovered image.

C. Watermark robustness

Generally any images, including watermarked images, suffer some no-intentional modifications, such as compression and noise contamination; therefore watermark robustness against these incidental modifications must be considered. In the proposed algorithm, information of ROI blocks are embedded into quantized DCT coefficients by a predefined quality factor. This embedding method guarantees that the watermark sequences can be extracted completely, even if the watermarked image is compressed with a better or equal quality factor used in this method. In the experimental process, a quality factor 70 is used.

D. Comparison with other methods

To evaluate performance of the proposed algorithm, comparison with other two algorithms, Zhao's algorithm [6] and Hassan's algorithm [7], is carried out. Fig 3 shows detection accuracy of tampered regions and quality of the recovered images, using the proposed algorithm, Zhao's algorithm and Hassan's algorithm, respectively.

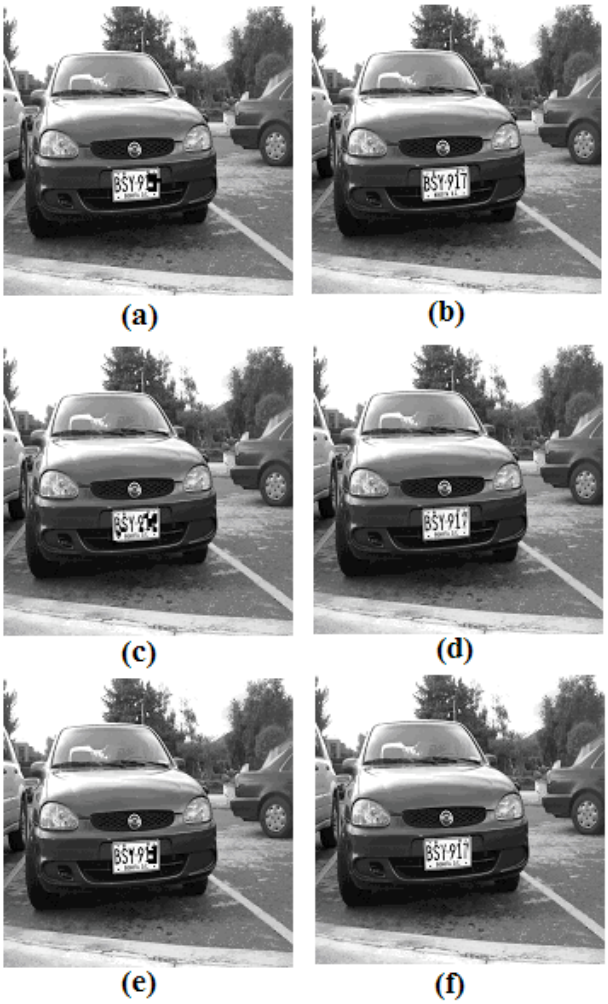

Figure 3. a), b) Tampered blocks detection and recovered image by the proposed algorithm, c), d) Tampered blocks detection and recovered image by Zhao's algorithm and e), f) Tampered blocks detection and recovered image by Hassan's algorithm.

From Fig. 3, the detection accuracy of tampered region provided by the proposed algorithm is better than other two algorithms. Also recovered blocks by the proposed algorithm is clearer than the recovered blocks generated by other two algorithms. The table II shows a summary of the performance comparison among three algorithms. In the table II, the quality of the image recovered by the proposed algorithm is approximately 4 dB better than that of other two algorithms. Also in the proposed algorithm, watermark can survive the JPEG compression with lower quality factor than other two algorithms.

TABLE II. COMPARISON SUMMRY AMONG PROPOSE ALGORTIM, ZHAO ALGORITM AND HASSAN ALGORITHM

	Proposed Algorithm	Zhao Algorithm [6]	Hassan Algorithm [7]
Quality of recovered image	42.9 dB	39.4 dB	37.1 dB
Tampered area Localization accuracy	The trees algorithms were able to detect modified blocks.		
JPEG compression Robustness	QF >80	QF >90	QF >95

VI. CONCLUSIONS

In this paper, a block-wise image authentication with tamper detection and recovery capability is proposed. Firstly image is segmented into two regions: Regions of Interest (ROI) and Regions of Embedding (ROE). The watermark sequence is a compressed version of each ROI block and it is embedded into ROE in DCT domain. Computer simulation results show excellent performance of the proposed scheme, from the following points of view: watermark imperceptibility, tamper detection and recovery capability and watermark robustness against no intentional attacks, such as JPEG compression. In the proposed scheme, recovered image of the tampered ROI blocks are sufficiently clear after watermarked image is compressed by JPEG compression with a reasonable quality factor (QF>80). The comparison results showed that the better quality of recovered image in the proposed algorithm than other previously reported methods, also embedded watermark is more robust to JPEG compression.

REFERENCE

[1] C.-S Lu, H. -Y. Liao, "Structural Digital Signature for Image Authentication: An Incidental Distortion Resistant Scheme", IEEE Trans. Multimedia, vol. 5, no. 2, 2003, pp.161-173.

[2] J. Dittmann, A. Steinmetz, and R. Steinmetz, "Content-based digital signature for motion pictures authentication and content-fragile watermarking," in IEEE Int. Conf. Multimedia Computing and Systems, vol. II, pp. 209–213,1999.

[3] J. Fridrich, "Methods for detecting changes in digital images," in Proc IEEE Int. Workshop on Intelligent Signal Processing and Communication Systems, 1998.

[4] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," Proc. IEEE, vol. 87, pp. 1167–1180, 1999.

[5] C. Lu and H. Liao, "Multipurpose watermarking for image authentication and protection," IEEE Trans. Image Processing, vol. 10, pp. 1579–1592, Oct. 2001.

[6] C.- X. Zhao, A.T.S. Ho, H. Treharne, V. Pankajakshan, C. Culnane, W. Jiang, "A Novel Semi-Fragile Image Watermarking, Authentication and Self-Restoration Technique Using the Slant Transform". IIHMSP 2007. Third International Conference on Volume 1, pp. 283 – 286, 26-28 Nov. 2007.

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