Tamper Localization in Wavelet Domain Using Semi-Fragile Watermarking

Sandipbhai Mangroliya, Ketki Pathak

Departement of Electronics and Communication Engineering
Sarvajaink College of Engineering & Technology, Surat

***Abstract* - This paper proposes a novel image authentication scheme in the DWT domain. The scheme uses a semi-fragile watermark to detect and precisely locate malicious tampering region in images. The wavelet coefficients selected for embedding are randomly permuted with a secret key, achieving high security. A bit of the watermark is embedded in a group of coefficients by means of quantization. The experimental results show, that our algorithm achieves good image quality and high tampering detection resolution at a low watermark payload, compared to block-based authentication schemes. The watermark is protected against local attacks. We have also conducted experiments to demonstrate the robustness of the watermark against mild to moderate JPEG compression.**

***Keyword* - Multimedia security, Digital watermarking, Random permutation, Image authentication**

I. INTRODUCTION

Digital watermarking can be classified into three different categories: robust, fragile and semi-fragile water-marking. A robust watermark should be able to resist intentional or unintentional manipulations, while a fragile watermark is intended to be destroyed even after the smallest unintentional manipulation. Some important applications for robust watermarking are finger printing, data mining and copyright protection [1, 2]. The third category, semi-fragile watermarking [3, 4], uses watermarks that have the ability to resist unintentional manipulations caused by common image processing operations like JPEG compression and are fragile against intentional, malicious manipulations. The main application field of fragile and semi-fragile watermarking is image and video content authentication. This paper focuses on using a semi-fragile watermark for image authentication.

Most of the image authentication techniques use a block-based concept in the spatial, Discrete Cosine Transform (DCT) or wavelet domain [5, 6] for detecting the tampered regions in the image. The original image is partitioned into blocks, and each block is embedded with its own watermark using a secret key. However, most of these schemes are vulnerable to counterfeiting attacks and they cannot discriminate between intentional tampering and unintentional image processing distortions [7, 8]. In such approaches, the tamper detection resolution is limited to the block size. Smaller block sizes and higher watermark payloads are necessary to improve the detection resolution, resulting in a considerable degradation of the image quality. A specific attack against image authentication is the collage attack, introduced by Holliman and Memon in [9]. Using this attack, a counterfeiter can combine independently authenticated blocks to produce forged content. This attack is thus undetectable by conventional block-based watermarking techniques.

The scheme proposed in this paper focuses on eliminating the drawbacks of block-based schemes, using a Discrete Wavelet Transform (DWT) based approach and random permutation of wavelet coefficients. Our algorithm improves the tamper detection localization; unlike conventional block-based methods, the exact tampered region is detected. The watermark payload is significantly lowered, increasing the image quality. The algorithm can also distinguish between intentional tampering and unintentional mild to moderate JPEG compression. Mathematical morphology can be used to improve the detection results [10, 11].

The remainder of this paper is organized as follows. In Section 2 we present the proposed image authentication scheme in the wavelet domain, showing the block diagrams of the watermark encoder and decoder and detailing the steps of the algorithm. Section 3 contains the experimental results and performances of the proposed scheme and conclusions are given in Section 4.

II. THE PROPOSED IMAGE AUTHENTICATION SCHEME IN THE DWT DOMAIN

This section describes the proposed image authentication scheme. The block diagram of the watermark encoder is shown in Fig. 1.

The watermark embedding process is described in the following:

1. The original, gray scale image is transformed into the wavelet domain using a 2D-DWT decomposition on L resolution levels.

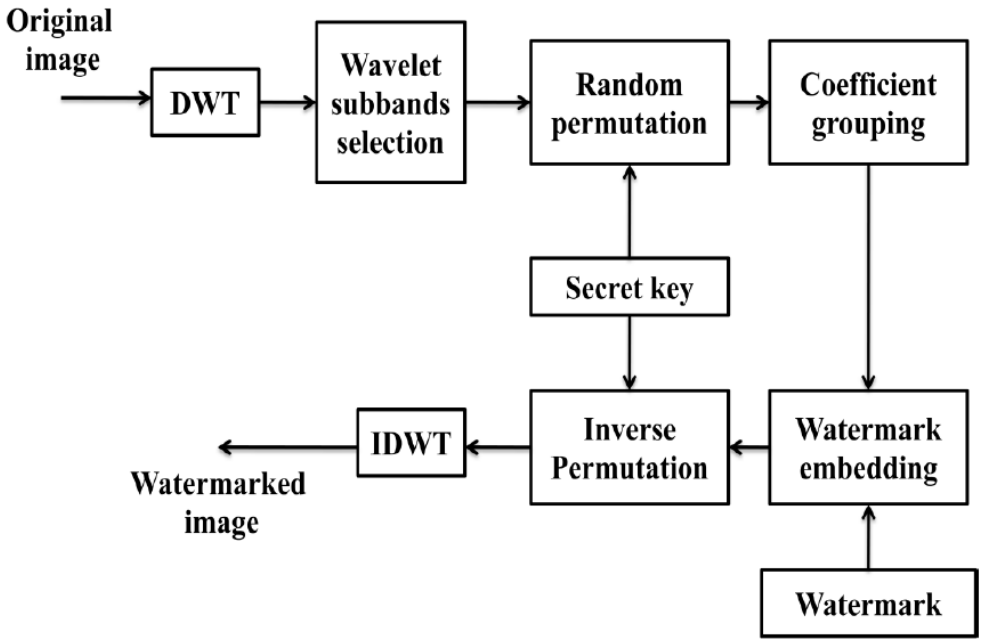

Figure 1. Block diagram of the watermark encoder

2. The detail wavelet coefficients from the LHn, HLn and HHn Wavelet sub-bands of resolution level n ≤ L are selected. For example, in Fig. 2, the coefficients of the second Wavelet decomposition are chosen for further processing. Selecting higher resolution sub-bands for watermark embedding gives a better tampering localization, but decreases the resilience of the algorithm to common, unintentional image processing operations.

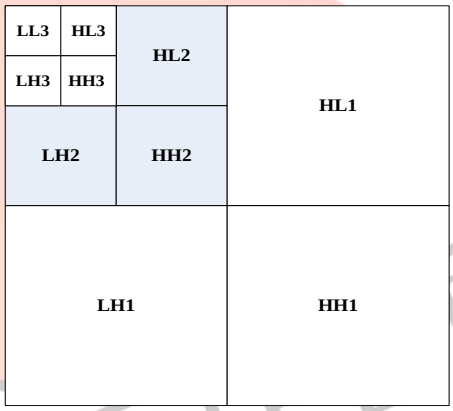

Figure 2. Wavelet decomposition on L = 3 resolution levels

3. The selected wavelet coefficients are concatenated in a 1D vector S and randomly permuted using a secret key K into a new vector S'. This process assures that coefficients corresponding to the same spatial location will be separated in S'.
4. The sequence S' of permuted coefficients is divided into groups of d coefficients. Parameter d controls the watermark payload of the proposed scheme. A watermark bit will be embedded in every group of d coefficients. A smaller group size will result in a bigger watermark payload and thus in a higher degradation of the image quality and a smaller size of the maximal localizable tampered area, but will not decrease the detection resolution of the scheme.
5. The watermark, a binary random sequence w, is generated and serves as the authentication code. w has the same length as the number of wavelet coefficient groups.
6. The weighted mean mi of every group i of permuted wavelet coefficients is obtained using the following equation:

$$m_i = \sum_{j=1}^{d} (-1)^j \left| c_i(j) \right| \qquad (1)$$

Where $c_i(j)$ denotes the j^{th} coefficient of group i. The weight $(-1)^j$ is used to make the scheme more robust to common image processing operations that, in most cases, change the entire image and would not modify the weighted mean.

7. The watermarked mean m_i^w is obtained by quantization of mi to the nearest even or odd quantization level according to the value of the corresponding watermark bit wi, using the following equation:

$$m_i^w = \begin{cases} [m_i / Q] * Q & if \ \bmod 2([m_i / Q]) = w_i \\ [m_i / Q] * Q + Q & if \ \bmod 2([m_i / Q]) \neq w_i \end{cases} \qquad (2)$$

where [ ] is the integer part operator and mod2 is the remainder after division by 2.

8. For every group i of coefficients, the weighted mean mi is changed to the watermarked mean m_i^w by modifying the wavelet coefficient si ,max(j) with the highest absolute value. Because of the random permutation operation every group should have at least one coefficient with high absolute value. The updating of si, max(j) is done as follows:

$$s_{i,max}^{w}(j) = s_{i,max}(j) + (-1)^{j} * sign\left(s_{i,max}(j)\right) * (m_i^w - m_i) \quad (3)$$

$s_{i,max}^{w}(j)$ is the watermarked coefficient and

$$sign(x) = \begin{cases} -1 & if\ x \le 0 \\ 1 & if\ x > 0 \end{cases} \quad (4)$$

9. After updating the coefficients with the highest absolute value from every group, the inverse permutation is applied using the secret key K.
10. Finally, the Inverse 2D-DWT is calculated to obtain the watermarked image.

The block diagram of the watermark decoder and image authentication is shown in Fig. 3. The watermark extraction process and image authentication is described in following:

1. The watermarked, possibly tampered image is transformed into the wavelet domain using 2D-DWT decomposition on L resolution levels.
2. The detail wavelet coefficients from the LHn, HLn and HHn Wavelet sub-bands are selected for watermark extraction.
3. The selected wavelet coefficients are concatenated. Using a secret key K, the same random permutation is performed.
4. The sequence of permuted coefficients is divided into groups of d coefficients.
5. The weighted mean $m_i^{'}$ of every group of coefficients is calculated using Eq. (1).
6. A watermark bit $w_i^{'}$ is extracted from the weighted mean of every group of d coefficients using Eq. (5).

$$w_i' = mod2([m_i'/Q]) \quad (5)$$

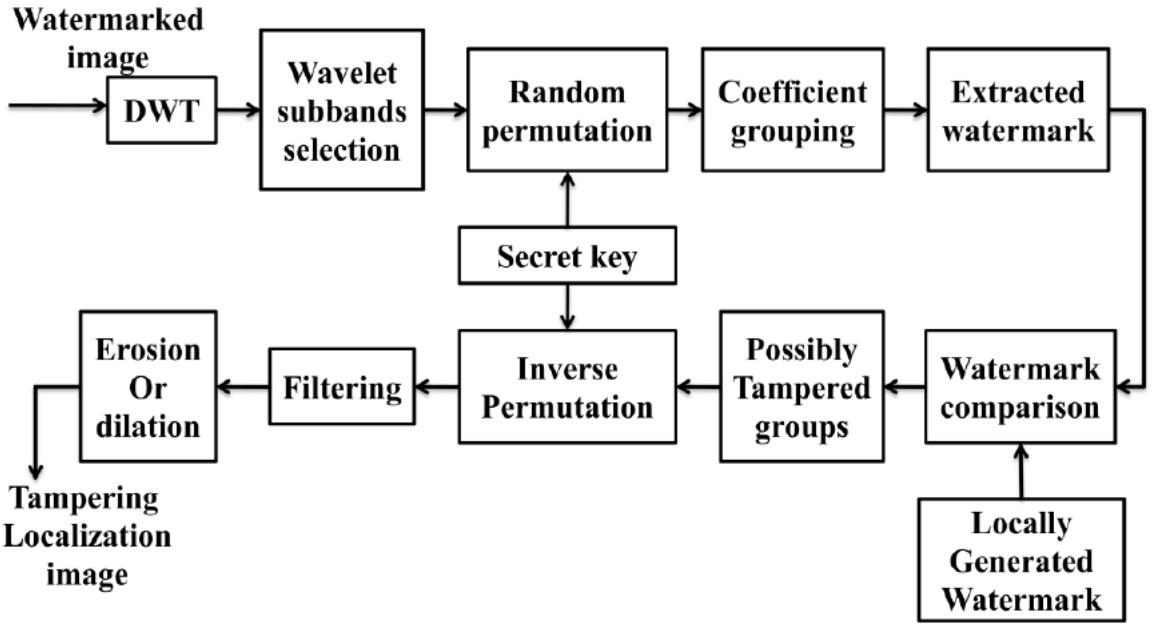

Figure 3. Block diagram of the watermark decoder and image authentication

To authenticate the image the following step need to be done:

1. First, the original watermark w is locally generated. If the extracted watermark w' matches the original one then the image is authentic. If not, the following steps will determine if the image was tampered.
2. If a bit of the extracted watermark $w_i^{'}$ does not match the original one wi, all coefficients belonging to group i are declared as potentially tampered.
3. All coefficients are permuted back to their original position using the inverse permutation with the secret key. The coefficients declared as potentially tampered should be now spread all over the sub-bands. The actual tampered regions

should have a high density of flagged coefficients. The other flagged coefficients should be isolated and distributed like random noise. They are false positives and can be claimed as authentic.

4. For a selected resolution level n we will have three sub-bands, LHn, HLn and HHn, with flagged and non-flagged coefficients. We will construct a binary authentication matrix A of the same size as the selected sub-bands. A(x,y) = 1 if there is a flagged coefficient at the same position (x,y) in any of the HL, LH, or HH selected sub-bands.
5. The isolated ''1'' bits in the authentication matrix A will be removed through filtering.
6. Locations incorrectly flagged as tampered are further eliminated by successive erosion and dilation, using a disk of radius R pixels and a square of size S x S as structural elements. The ''1'' bits in the filtered and eroded authentication matrix should now correctly indicate the tampered locations.
7. The flagged positions in A are then mapped back to the spatial domain to indicate the actual tampered locations.

The sensitivity of the tampering detection can be adapted by selecting the quantization step size Q, choosing the filter size in step 5 and the size of the structural element used for erosion and dilation in step 6 of the authentication process. A larger Q will increase the sensitivity, but will decrease the perceptual quality of the watermarked image. A larger filter or a bigger structural element for erosion and dilation will reduce the sensitivity, but will also reduce the probability of false alarms as a result of common image processing operations. A trade-off has to be found for specific applications.

III. RESULTS AND ANALYSIS

To determine the performances of the proposed image authentication scheme for several images of different sizes were used. We have evaluated our algorithms in terms of image quality, localization capability of the tampering. Table 1 shows the image quality and decoding results of the proposed method for different choices of the wavelet resolution level n for watermark embedding, quantization step size Q and group size d, where PSNR is the mean Peak Signal to Noise Ratio. The PSNR values are above 40 dB, except Q=12 (d = 4), Q =16 (d = 4 and d = 8) and Q =20 (d =4, d = 8 and d = 12), where small distortion are also visible in the watermarked images.

Table 1 Mean PSNR values for the watermarked images

Quantization Level (Q)	Coefficient group size (d)	Image of lena.jpg		Image of cameraman.tif	
		N = 1 (PSNR)	N = 2 (PSNR)	N = 1 (PSNR)	N = 2 (PSNR)
4	4	48.0747	54.1246	47.9594	54.0978
	8	51.0739	57.1024	51.0331	57.0931
	12	52.8387	58.8619	52.8165	58.7949
	16	54.0869	60.0585	54.0899	60.0423
8	4	42.0568	48.1173	41.9420	48.0993
	8	45.0601	51.1148	45.0185	51.1123
	12	46.8278	52.8856	46.8057	52.8543
	16	48.0758	54.1088	48.0822	54.1124
12	4	38.5355	44.5974	38.4208	44.5815
	8	41.5397	47.5991	41.4979	47.5972
	12	43.3080	49.3711	43.2859	49.3495
	16	44.5556	50.6010	44.5630	50.6078
16	4	36.0369	42.0992	35.9223	42.0841
	8	39.0416	45.1025	38.9995	45.1005
	12	40.8100	46.8746	40.7879	46.8574
	16	42.0573	48.1075	42.0653	48.1151
20	4	34.0988	40.1611	33.9842	40.1465
	8	37.1037	43.1654	37.0615	43.1631
	12	38.8723	44.9373	38.8501	44.9227
	16	40.1194	46.1718	40.1276	46.1797

In Fig. 4 we compare our technique with wavelet decomposition level n = 1 and d = 4, d = 8, d = 12 and d = 16. We also measure the PSNR value on different quantization level. We can see that our approach achieves very high image quality with respect to increasing group size (d) and decreasing quantization level (Q).

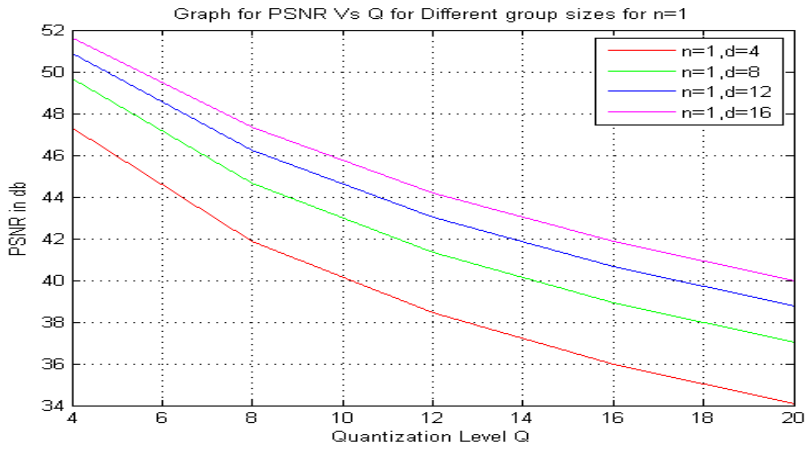

Figure III. Graph for quality of the watermarked images for different group sizes.

Next, we test the capacity of our approach to detect the tampering. For this purpose, in the image in Fig.5a shows the original input image and Fig.5b shows the watermarked image. First we embed the watermark in the first Wavelet decomposition level (n = 1), using groups of d = 8 wavelet coefficients and the quantization step size Q =8. Fig.5c shows the potentially tampered locations in the image before refining the authentication result through filtering.

(a)

(b)

(c)

(d)

(e)

(f)

Figure 5. (a) original image, (b) watermarked image after doing random permutation and adding watermark bit (c) authenticated image before filtering for n =1, d = 8 and Q = 8, (d) authenticated image after filtering and mathematical morphology operations for n =1, d = 8 and Q = 8, (e) authenticated image before filtering for n =2, d = 8 and Q = 8, (f) authenticated image after filtering and mathematical morphology operations for n = 2, d = 8 and Q = 8.

Then, we filter the authentication matrix with a ‘‘cross’’ shaped binary filter of size F = 5. The filtered authentication matrix is then successively eroded using a disk of radius R = 1 and a square of size 3 x 3 as structural elements. Fig. 5d shows the refined authentication result after filtering and erosion operations. In Fig. 5e and f we depict the authentication results before and after refinement using the following parameters: n =2, d = 8 and Q = 8.

IV. CONCLUSION

In this paper we have proposed a new image authentication scheme using a semi-fragile watermark that can detect and locate malicious tampering in images. The embedding and extraction of the authentication watermark is done in the DWT domain. Our technique achieves high image quality and high tampering detection resolution at a low watermark payload using random permutation of the wavelet coefficients before embedding, the watermark is also protected against local attacks.

REFERENCES

[1] R.O. Preda, D.N. Vizireanu, “Robust wavelet-based video watermarking scheme for copyright protection using the human visual system,” *Journal of Electronic Imaging*, vol. 20, pp. 374-379, 2011.

[2] R.O. Preda, D.N. Vizireanu, “A robust digital watermarking scheme for video copyright protection in the wavelet domain, Measurement”, *Measurement*, vol. 43, pp. 1720-1726, 2010.

[3] Preda, R.O.; Vizireanu, D.N., "Blind Watermarking Capacity Analysis of MPEG2 Coded Video," *8th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, pp. 465,468, 26-28 Sept. 2007

[4] A. Piva, F. Bartolini, R. Caldelliy, “Self recovery authentication of images in DWT domain,” *International Journal of Image and Graphics*, vol. 5, pp. 149-166, 2011.

[5] C.-T. Li, H. Si, “Wavelet-based fragile watermarking scheme for image authentication,” *Journal of Electronic Imaging*, vol. 16, pp. 1-9, 2007.

[6] A.H. Ouda, M.R. El-Sakka, “Localization and security enhancement of blockbased image authentication,” *IEEE International Conference on Image Processing* , vol. 1, pp. 673–676, September 2005

[7] C. Fei, D. Kundur, R.H. Kwong, “Analysis and design of secure watermark-based authentication systems,” *IEEE Transactions on Information Forensics and Security*, vol. 1, pp. 43-55, 2006

[8] L.P. Freire, P. Comesana, J.R.T. Pastoriza, F.P. Gonzalez, “Watermarking security: a survey,” *LNCS Transactions on Data Hiding and Multimedia Security*, vol. 1, pp. 41-72, 2006.

[9] M. Holliman, N. Memon, “Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes,” *IEEE Transactions Image Processing*, vol. 9, pp. 432-441, 2000.

[10] D. N Vizireanu, S. Halunga, G. Marghescu, “Morphological skeleton decomposition interframe interpolation method,” *Journal of Electronic Imaging*, vol. 19, pp. 1-3, 2008.

[11] R. M Udrea, D.N. Vizireanu, “Iterative generalization of morphological skeleton,” *Journal of Electronic Imaging*, vol. 16, pp. 1-3, 2007.