IOP Conf. Series: Materials Science and Engineering **288** (2018) 012093 doi:10.1088/1757-899X/288/1/012093

Review of Some Existing Work for Self-Recovery Fragile Watermarking Algorithms

L Rakhmawati[1*] and N Rochmawati [2]

[1]Department of Electrical Engineering, Universitas Negeri Surabaya
[2]Department of Informatics, Universitas Negeri Surabaya

[*]lusiarakhmawati@unesa.ac.id

Abstract. This paper covers brief study about some of the latest research trends in self-recovery fragile watermarking algorithms. Fragile watermarking techniques aim to detect modifications to images that are sensitive to any changes, which they are intolerant of even a one-bit alteration. It is desirable to be able to locate the modified areas when a portion of the original content is replaced with other information. Therefore, fragile watermarks are intended for checking the integrity and authenticity of digital contents. Self-recovery means that some significant feature obtained from a selected image itself as a watermark by modifying the pixel values of the original image (host). Once the picture is modified by other users, which has been inserted watermark can be used for tamper detection and recovery images. In this research, we have studied that generally, embedding algorithm for tamper recovery includes three main steps: preprocess the original image and extract important information from it, reorder and embed important information into spatial or frequency domain of host image, and extract the important information from stego-image and use it to recover damaged parts as tamper occurs.

1. Introduction

The rapid development of internet makes the distribution of multimedia data becomes very easy. Multimedia data such as digital images, could be used in legal processes [1], are vulnerable to be modified and manipulated using digital image processing tools that can be found on almost every PC and mobile devices [2]. The integrity and authenticity of digital images can be guaranteed by using digital watermarking which is a process to embed special information (text or image), known as a watermark, to digital source [3].

There has been a large amount of digital watermarking algorithms devoted to tamper detection. Fragile watermarking is one algorithm that serves as a detection tool that is not tolerant of damage to a single bit change. On the other hand, semi-fragile watermarking is designed to update the content. They are capable to distinguish between categories of innocent and malicious attacks [4].

Recently, there is a self-recovery fragile watermarking method in which we can select the important information from an image and embed to the original image to be used on the receiver to recover the damaged image. It works to authenticate digital content, localize the damaged parts, and recover the tamper images [5]. The algorithm begins with dividing the image into a number of non-overlapping blocks in a certain size; each block is inserted with watermark that can be used to restore the damaged image section, and to authenticate the block.

In this paper, we are focusing on a survey of previous self-recovery fragile watermarking techniques. Self-recovery fragile watermarking techniques which are part of digital watermarking schemes can be broadly divided into two types; based on the two basic image domains which the cover images belong to (1) techniques in the spatial domain (SD) and (2) techniques in the frequency domain (FD). In each domain, we will explain three main steps: preprocess the original image and extract important information from it, reorder and embed important information into spatial or frequency domain of host image, and extract the important information from stego-image and use it to recover damaged parts as tamper occurs.

2. Basic of the self-recovery fragile watermarking algorithms

Digital watermarking terminology can generally be as follows. The image to be protected is called the original image. Images with inserted watermark are called watermarked images. There are two stages of the watermarking algorithm for image tamper detection and recovery, the embedding watermark phase and authentication phase which then followed by tamper detection and recovery image [9].

Based on some characteristics, there are several categories used to estimate the effectiveness of a self-recovery fragile watermarking technique [2,9,15]:

- *Perceptibility:* The embedded watermark must be invisible. It will be hard to identify it with human vision; only authorized user will be able to recover it. The original cover image and the watermarked image are indistinguishable and the recovered images should be in a high quality.
- *Tamper Detection*: A self-recovery fragile watermarking technique must be able to detect with or without original image or some information to derive it for the detection process.
- *Tamper recovery*: The scheme should be able to detect unauthorized modification on images and recover them from the tampered ones
- *Resistant to known attacks*: The scheme should be as robust as it could to the well-known attacks, such as the counterfeit attack, the disturbing attack and the leakage of the secret key.

Besides, some parameters used to measure the performance of the watermarking algorithm are Peak Signal-to-Noise Ratio (PSNR) and bits-per-pixel (bpp). PNSR value can measure the perceptual transparency of the watermarked image with respect to the original image [15]. Meanwhile, the bpp is used to calculate the amount of watermark data which embedded in the original image. The idle value of bpp is 1 but there are certain watermarking algorithms available to improve the bpp value more than 1.

3. Self-recovery fragile watermarking algorithms in the spatial domain

Recently, several watermarking schemes designed for tamper detection and recovery has been proposed as described in Table 1. In the spatial domain, at the encoder site, watermarks are composed by detection bits and authentication bits. At the decoder side for the authentication phase [9, 28] is done by taking the watermark component in order to check whether there is any malicious modification or not. If there is damage, watermark components can be used to recover it, all the existing works in spatial domain was using block-neighbourhood based [8]. General description of the algorithm is shown in Figure 1.

The selection of watermark components for recovery bit determines the success of the recovery process. In addition, the longer the watermark, the more information about the verification and recovery can be maintained. Therefore, a longer watermark generally resulted in more accurate tamper detection and a better quality of recovered image. However, the length of the watermark should be definite for keeping the watermarked image from serious distortion. Hence, to simultaneously enhance the precision of tamper detection and the quality of the recovered image, while preserving the quality of the watermarked image should be taken into consideration in the future research [9].

Sreenivas [6] proposed a self-embedding watermarking scheme that improves the quality of the recovered image. It compares the use of average intensity of the block with the watermarking bits in a variable length. Kiatpapan [14] used dual watermarking approach to ensure a robust performance in image tamper detection and its recovery. The arrangement of the bit-planes of the watermark is also an important factor. In the Kiatpapan's proposed method, two sets of 8 bit-planes of the watermark are

arranged in a centre-point-symmetric fashion so that important image information is spread uniformly. Consequently, his method can recover an image tamper perfectly even if it takes place on the left, right, upper, or lower half of the original image.

Table 1. Summarization of Some Self-Recovery Fragile Watermarking Methods

Method/ Approach	Basic Domain	Segmented Block Size (pixels)	Watermark Payload (bpp)	Watermarked Image PSNR	Tamper Detection& recovery Method	Recovered Image PSNR
[1,30]	FD (DWT)	8x8	3	>35 dB	Daubechies DWT 4 level	>32 dB
[7]	SD	mxn	2	42-45 dB	Shamir algorithm	22-34 dB
[6,8,13]	SD	2x2	2	30- 44.15 dB	block-neighbourhood	32.04-62 dB
[9,12,16, 17, 24]	SD	8x8	3	37.9 -42 dB	block-neighbourhood	24–42 dB
[10, 11,20, 26,27,28]	FD (DCT)	8x8	2	34-38 dB	IDCT	36 – 44 dB
[14,19,25]	SD	4x4	2	36-44 dB	block-neighbourhood	30-76 dB
[21,23]	SD	8X8	2	44.26 dB	block-wise dependent	44-50 dB
[22]	SD	nxn	1	51 dB	block-neighbourhood	41 – 48 dB

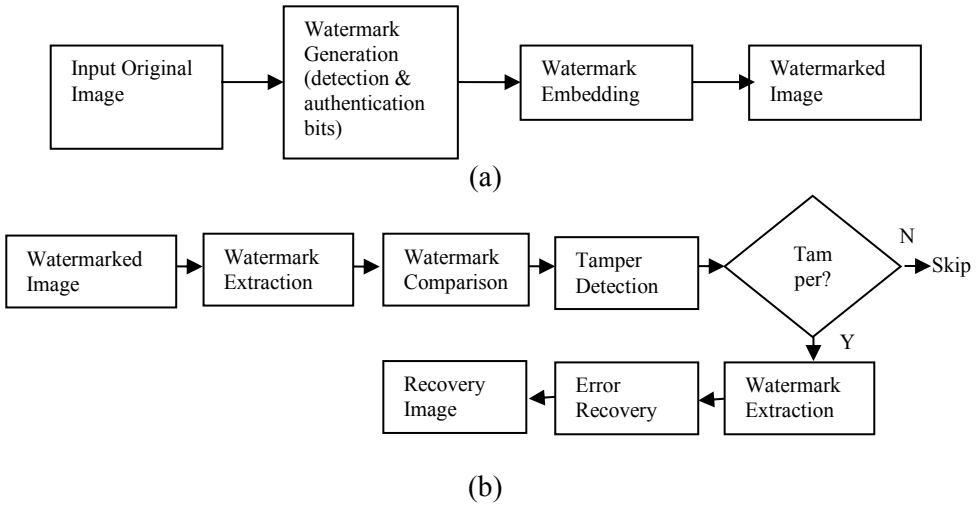

Figure 1. Block diagram of self-recovery fragile watermarking techniques in a spatial domain: (a) encoder side; (b) decoder side

In embedding phase, the method [8, 12] used a pseudorandom sequence to generate the nonlinear block-mapping and employed an optimized neighbourhood characterization method to detect the tampering. His method [8] also investigates three optimization strategies that will further improve the

quality of tamper localization and recovery. They take all adjacent blocks of the test block and its mapping block into account and then utilize a statistic-based rule to determine the validity of image blocks. Based on the analytical analysis of its false acceptance and false rejection probabilities, the post-processing operation is presented to further improvement of the performance from the proposed SDM. It includes three steps: (1) mark dubious blocks, (2) distinguish tampered blocks from dubious blocks by the adjacent blocks, and (3) improve the detection performance by the post-processing [12].

In addition, the needs of authenticity inspection and correctness of the content of an image have been developed fragile watermarking scheme [16, 17]. The method developed by Wong et al. [18] by dividing the image into blocks of a certain size, then the watermark is inserted into each block. The weakness of this method lies in the resistance to various attacks. Furthermore, some fragile watermarking methods have been used with the addition of better restoration capabilities and can be applied in different types of images [19-23]. Research conducted by Lin et al. [19] facilitates the ability to detect hierarchical damage, where this process can locate areas damaged up to three levels. If it is not found in the first phase, it will continue to the second and third phases. However, in this scheme, it is not possible to recover the damaged block when the watermark, which inserted into another block, is also damaged and having no second chance to recover block [23].

The problem of other opportunities for restoration has been proposed by [20, 21], for the recovery of the block assigned to solve the accidental problem of interference by installing two copies of the restoration bits into the image. Increasing the watermarking capacity leads to a quality decrease of watermarked images as seen in Lee et al. [20] which uses 3 bits of the least significant bits (LSB) to store bit recovery. To improve the quality of watermarked images and restoration of damaged images, Qin et al. [22] proposed fragile watermarking methods using adaptive bit allocation mechanisms and image improvements. This method inserts a watermark into one LSB with the ability to change the length of the block image encoding results based on the smoothness of the block. On the side of the decoder, if the extraction of the watermark length is not suitable, it cannot show where the image damaged is, some addition of authentication bit components can degrade the image quality itself. Another method [21] ignores the image content, removes the concealment space, and uses the watermark component of the encoding result of eleven first quantization coefficients. Huo et al. [23] proposed method that divided the image into eight sections according to its roughness level. The watermark component consists of authentication and recovery bits with a length that can be changed. This method can result in the precision of the destructed location.

It is inferred from the survey that the existing methods for performing tamper detection and recovery required a lot of authentication and recovery data to be embedded. Generally, in the self-recovery fragile watermarking, the watermark payload ranges from 1 to 3 bpp (bit per pixel). And with the increase of watermark payload, the PSNR value of the watermarked image decreases gradually [8]. This significantly reduces the perceptual quality of the watermarked image. Furthermore, the localization in tamper detection and recovery is not efficiently dealt in the existing works. The existing methods also do not validate tamper detection and recovery against multiple insertions, deletion and updating attacks. In paper [13], a fragile watermarking scheme is proposed for efficiently detecting tamper information and recovery of the tampered information. The proposed scheme focuses on achieving a higher quality of the recovery tampered regions. In addition, the improvement in the authentication process with higher accuracy in tamper detection is also focused on this method.

From the description of several methods above, the problem of the fragile watermarking method lies in the ability of restoration, which in the case of watermark insertion process, tamper localization, and recovery. There is a trade-off between the watermark image quality and the insertion capacity. The next problem is the accuracy of the localization tamper that makes the fragile watermarking method for self-recovery categorized successful or not. By utilizing more bits to detect the damaged area, the watermarked image quality is reduced. Another method in Ref. [8] proposed the use of restoration bits to locate damaged areas, but did not need to be embedded in watermarked images. The next problem is self-recovery capability. This process will restore the damaged area by extracting effective information using more bits to represent the restoration bit [25]. From the description above, on the sender side, the

selection of watermark components is very instrumental in the process of damage detection and image recovery. At the receiving end, the watermark can be extracted and can be used to cover the damaged part using the restoration bit with the appropriate content section.

.

4. Self-recovery fragile watermarking algorithms in the frequency domain

Unlike spatial-domain watermarking [6-9,12-14], frequency domain watermarking schemes which are summarized in Table I, based on frequencies or moments, image transformations including the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), are usually compatible with popular image compression standards. In many of these schemes, a given secret digital image is usually converted to its frequency domain moments. These frequency moments are embedded into those of the host image. The inverse Fourier transform then applied to the latter to form a watermarked image that is ready to be transmitted to the receiving end as shown in Figure 2. Transform-domain methods provide more information embedding and more robustness against many common attacks, but the computational cost is higher than the spatial-domain watermarking techniques.

A self-recovery of modified regions in digital images is almost had a drawback dealing with JPEG compression attack. Therefore, paper [1] proposes a digital image watermarking algorithm, which consists of two stages; the first one is to protect the digital image that is performed by Daubechies DWT [30, 31], half toning, and QIM methods; the second one is used for authentication, detection, and self-recovery of tampered regions, through IDWT, inverse half toning and median filtering.

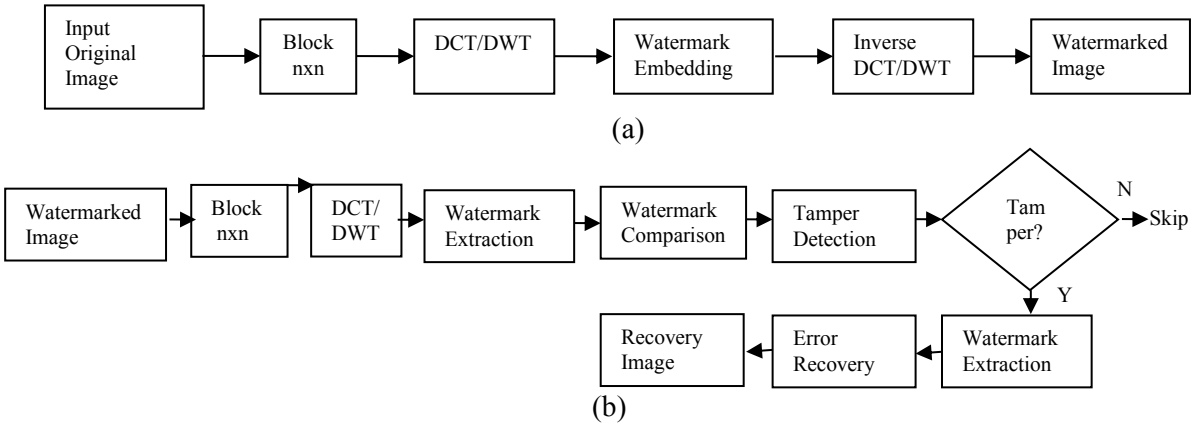

Figure 2. Block diagram of self-recovery fragile watermarking techniques in a frequency domain: (a) encoder side; (b) decoder side

While using another method, the embedded watermark data for content recovery is calculated from the original discrete cosine transform (DCT) coefficients of host image and do not contain any additional redundancy [10, 26, 27, 29]. When a part of a watermarked image is tampered, the watermark data in the area without any modification can still be extracted. If the amount of extracted data is larger, they can reconstruct the original coefficients in the tampered area according to the constraints given by the extracted data. Otherwise, they may employ a compressive sensing technique to retrieve the coefficients by exploiting the sparseness in the DCT domain. This way, all the extracted watermark data contribute to the content recovery. The smaller the tampered area, the more available watermark data will result in a better quality of recovered content. It is also shown that the proposed scheme outperforms previous techniques in general.

For assurance and exactness of recovered tampered region, Dhole [11] is using DCT transfer for obtaining recovery information. Then they insert a watermark into the original image to get first watermark image. In the next step, they shuffle the block of original image and merge this information block in reverse order from the previous block chaining into the original image to obtain next shuffled

image. They do EXOR operation of these two obtained images to get final resultant watermarked image. If the image is getting tampered, the modification in watermark image can be identified as well in the self- embedded image. They can locate modified image by performing reverse DCT.

The better level of security and complexity is lower according to Patra [26] by using Chinese residual theorem (CRT). This scheme has additional security because there is a random selection of blocks, either during watermark insertion or when selecting blocks to be used for recovery. This method has compared its performance with the SVD-based CRT watermarking scheme and spatial, and shows a scheme that is superior to the other two under some major attacks, such as bright and sharp. Furthermore, this scheme demonstrates a strong resistance to JPEG compression.

The characteristic comparison of the considered self-recovery watermarking schemes is summarized in Table 1. The spatial domain, the replacement may introduce some amounts of distortion. Therefore, the quality of the watermarked image fully depends on the number of the LSBs which are replaced with the watermark in a pixel, as shown in the table by increasing the value of bpp, it will decrease PSNR value. The PSNR value of the watermarked image in spatial domain is higher than in frequency domain due to the process of spatial domain more straight-forward without any transformation.

5. Conclusions

Based on the insertion region, there are two main categories of self-recovery fragile watermarking algorithms: spatial domains and frequency domains. Spatial domain technique is a widely used method for replacing the bit value in the LSB planes (1-3 LSB) while keeping the MSB planes of the original image intact with the watermark pixel value. Frequency domain techniques transform and process watermarks in the frequency domain, then inverse-transform into spatial domains. It is clear that in spatial domains, watermarks can be successful and easily restored if the images are in attack. On the other hand, frequency domains provide more security, it has proven to be very powerful for all types of attacks except rotation and cropping and it also has a very small effect on image quality. However, it is difficult to recover the watermark in the frequency domain at the receiving end due to its complexity.

References

[1] Javier Molina-Garcia et al 2016 Watermarking Algorithm for Authentication and Self-Recovery of Tampered Images Using DWT *Proc. Int. Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter, and Submillimeter Waves (MSMW)* 1-4

[2] Liu X.et al 2016 A Survey of Fragile Watermarking-based Image Authentication Techniques *J. Inf. Hid. and Multi. Sig.Proc*. **7** 6 1282–1292

[3] Bhargava N, Sharma M M, Garhwal AS and Mathuria M 2012 Digital Image Authentication System Based on Digital Watermarking *Proc.IEEE Int. Conf. on Radar, Communication, and Computing (ICRCC)* pp185–9.

[4] Ekici O, Sankur B, Coşkun B, Naci U and Akcay M 2004 Comparative evaluation of semi-fragile watermarking algorithms. *J. Elect. Imag.* **13** 1 209

[5] Zhang X and Wang S 2008 Fragile Watermarking with Error Free Restoration Capability *IEEE Trans Multimed* **10** 8 1490–9

[6] Sreenivas K and Prasad V K 2016 Improved Block Encoding Method For An Image Self-Recovery Approach *Proc. IEEE Int. Conf.* on Inf. Comm.and Emb.Syst. (ICICES) 3–7

[7] Sudha M S and Thanuja T C 2016 Randomly Tampered Image Detection and Self- Recovery for a Text Document Using Shamir Secret Sharing *Proc. IEEE Int. Conf. on Rec.Tren.in Elect., Inform.& Comm. Techn. (RTEICT*) 688–691

[8] He H, Chen F, Tai H, Member S, Kalker T and Zhang J 2012 Performance Analysis of a Block-Neighborhood-Based Self-Recovery Fragile Watermarking Scheme *J.IEEE Trans Inf Forensics Secur* **7** 1 185–196.

[9] Chen T, Hwang M and Jan J 2012 A secure image authentication scheme for tamper detection and recovery *J. The Imaging Science Journal* **60** 219–233

[10] Zhang X, Qian Z, Ren Y and Feng G 2011 Watermarking With Flexible Self-Recovery Quality

Based on Compressive Sensing and Compositive Reconstruction *J. IEEE Trans on Inform.Forenc.and Sec.* **6** 4 1223–1232
[11] Dhole V S and Nitin N Patil 2015 Self Embedding Fragile Watermarking for Image Tampering Detection and Image Recovery using Self Recovery Blocks *Proc. IEEE Int. Conf. on Comp. Comm. Cont. and Auto* 752–757
[12] He H, Zhang J and Chen F 2009 Adjacent-block based statistical detection method for self-embedding watermarking techniques *J. Sig. Proc.* **89** 1557–1566
[13] Shivananda N 2014 A new fragile watermarking approach for tamper detection and recovery of document images *Proc. IEEE Int. Conf. on Adv.in Comp.Comm.and Inform. (ICACCI)* 1494–1498
[14] Kiatpapan S and Kondo T 2015 An Image Tamper Detection and Recovery Method Based on Self-Embedding Dual Watermarking *Proc. IEEE Int. Conf. on Elec. Eng./Elect., Comp., Tel. and Inform.Tech.(ECTI-CON)* 1-6
[15] Vyas Chinmay and Munindra Lunagaria 2014 A review on Methods for Image Authentication and Visual Cryptography in Digital image Watermarking *Proc. IEEE Int. Conf. On Comp. Intell. and Comp. Res.Dig.*
[16] Mei Yu et al 2014 New fragile watermarking method for stereo image authentication with localization and recovery *J.Electron. Commun* 2-10
[17] Teng L and Wang X 2013 Cryptanalysis and improvement of a chaotic system based fragile watermarking scheme AEU-International *J Electron Commun* **67** 6 540–547
[18] Wong P W and Memon N 2001 Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Trans Image Process* **10** 10 1593–1601
[19] Lin P L, Hsieh C K and Huang P W 2005 A hierarchical digital watermarking method for image tamper detection and recovery *J.Pattern Recogn* **38** 12 2519–2529
[20] Lee T Y, Lin S D 2008 Dual watermark for image tamper detection and recovery. *J.Pattern Recogn* **41** 113497–3506
[21] Li C, Wang Y, Ma B and Zhang Z 2011 A novel self-recovery fragile watermarking scheme based on dual-redundant-ring structure *J.Comp Electr Eng* **37** 6 927–940
[22] Qin C, Chang C C and Chen P Y 2012 Self-embedding fragile watermarking with restoration capability based on adaptive bit allocation mechanism *J. Signal Process* **92** 4 1137–1150
[23] Huo Y, He H and Chen F 2012 Alterable-capacity fragile watermarking scheme with restoration capability *J.Optics Commun* **285** 7 1759–1766
[24] Zhang X, Wang S and Qian Z 2011 Reference sharing mechanism for watermark self-embedding *J.IEEE Trans Image Process* **20** 2 485–495
[25] Xiao D and Shih F Y 2012 An improved hierarchical fragile watermarking scheme using chaotic sequence sorting and subblock post-processing *J. Optics Commun* **285** 10 2596–2606
[26] J C Patra, Jiliang E Phua and Cedric Bornand 2010 A novel DCT domain CRT-based watermarking scheme for image authentication surviving JPEG compression *J.Digital Signal Process* **20** 1597–1611
[27] Simying Ong, Shujun Li, KokSheikWong and KuanYew Tan 2017 Fast recovery of unknown coefficients in DCT-transformed images *J. Signal Process: Image Comm* 1-17
[28] J C Patra, A Karthik, C Bornand 2010 A novel CRT-based watermarking technique for authentication of multimedia contents *J.Digital Signal Process* **20** 442–453
[29] S. Li, A. Karrenbauer, D Saupe, C.-C. J.Kuo 2011 Recovering missing coefficients in DCT-transformed images *Proc IEEE Inter Conf on Image Processing* 1537–1540
[30] Wei Wang, Aidong Men and Bo Yang 2010 A Feature-based Semi-Fragile Watermarking Scheme In DWT Domain *Proc IEEE Inter Conf on NIDC*
[31] Nidhi Divecha and N N Jani 2013 Implementation and performance analysis of DCT-DWT-SVD based watermarking algorithms for color images *Proc IEEE Intern Conf on Intelligent Syst and Signal Process (ISSP)*