J Inf Process Syst, Vol.14, No.2, pp.510~522, April 2018
https://doi.org/10.3745/JIPS.02.0082

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Review on Self-embedding Fragile Watermarking for Image Authentication and Self-recovery

Chengyou Wang*, Heng Zhang*, and Xiao Zhou*

Abstract

As the major source of information, digital images play an indispensable role in our lives. However, with the development of image processing techniques, people can optionally retouch or even forge an image by using image processing software. Therefore, the authenticity and integrity of digital images are facing severe challenge. To resolve this issue, the fragile watermarking schemes for image authentication have been proposed. According to different purposes, the fragile watermarking can be divided into two categories: fragile watermarking for tamper localization and fragile watermarking with recovery ability. The fragile watermarking for image tamper localization can only identify and locate the tampered regions, but it cannot further restore the modified regions. In some cases, image recovery for tampered regions is very essential. Generally, the fragile watermarking for image authentication and recovery includes three procedures: watermark generation and embedding, tamper localization, and image self-recovery. In this article, we make a review on self-embedding fragile watermarking methods. The basic model and the evaluation indexes of this watermarking scheme are presented in this paper. Some related works proposed in recent years and their advantages and disadvantages are described in detail to help the future research in this field. Based on the analysis, we give the future research prospects and suggestions in the end.

Keywords

Image Authentication and Self-recovery, Least Significant Bit (LSB), Peak Signal-to-Noise Ratio (PSNR), Self-embedding Fragile Watermarking

1. Introduction

In recent years, digital multimedia technology has been widely used in our daily lives. Thanks to this, the image processing and transmission have become more and more convenient than before. However, at the same time, people can modify and duplicate the images casually with the help of image processing tools. Therefore, how to guarantee the authenticity and integrity of digital images becomes an urgent question. To resolve this problem, many image authentication algorithms have been proposed, which mainly include digital signature [1,2] and digital watermarking [3]. The digital signature schemes attach a signature to the host image on the sending side and compare the extracted signature with original signature on the receiving side to achieve image authentication. However, it cannot locate the modified areas. The digital watermarking for image authentication overcomes this shortcoming, which can locate

Manuscript received February 17, 2017; first revision April 24, 2017; accepted May 31, 2017.
Corresponding Author: Xiao Zhou (zhouxiao@sdu.edu.cn)
* School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China (wangchengyou@sdu.edu.cn, sdwhzh@mail.sdu.edu.cn, zhouxiao@sdu.edu.cn)

the tampered regions and even recover the tampered image. Therefore, it gradually becomes a research hotspot in this field.

Depending on their characteristics, digital watermarking schemes can be divided into three categories: robust watermarking, semi-fragile watermarking, and fragile watermarking. The robust watermarking algorithm can resist almost all the attacks including malicious attacks and unintentional modifications like JPEG compression. In view of this, it is generally used in copyright protection [4]. On the contrary, the fragile watermarking algorithm is susceptive to any modification. The semi-fragile watermarking algorithm combines the advantages of the above two watermarking methods. It has certain robustness against some common signal processing operations, and it is also sensitive to malicious attacks. Based on the embedding domain, the digital watermarking schemes can be further split into spatial domain watermarking and transform domain watermarking. In spatial domain, the watermarking message is embedded into host image by directly altering its pixel values. The most typical method in spatial domain is the watermarking scheme based on the least significant bit (LSB). In this scheme, one or few LSBs of each pixel value are selected and substituted by watermark information. This embedding method has good fragility for every possible modification on host image. Due to its simplicity, it has been widely used for image authentication and recovery. In transform domain, the watermark is hidden in host image via modulating transform domain coefficients. The common used transforms mainly include discrete cosine transform (DCT), discrete wavelet transform (DWT), and singular value decomposition (SVD). To achieve better robustness against modifications, the robust and semi-fragile watermarking algorithms are usually performed in transform domain. According to different purposes, the fragile watermarking schemes are divided into two categories [5]: fragile watermarking for image tamper localization and fragile watermarking with recovery ability. The fragile watermarking used for tamper localization can identify and locate the modified areas, but it cannot recover the distorted image. In some fields, especially in court, image recovery is necessary to trace the attacker's intention. In addition, a good recovered image can be reused for further image processing without retransmission. Therefore, the fragile watermarking for image authentication and self-recovery has received much more attention lately. Its main idea is to embed the watermark derived from original image to the image itself. Many relevant algorithms in this field have been put forward recently and they have achieved great success in image tamper detection and recovery. However, to the best of our knowledge, there is few relevant review in this area. In this paper, we make a study on self-embedding fragile watermarking schemes for image tamper detection and recovery. The basic model and common attacks as well as the evaluation indexes are described objectively. To help the future research, some related works and their merits and demerits are summarized. Based on the analysis, we give the future challenges and improvement directions in the end.

The rest of this study is organized as follows. Section 2 introduces the basic model of self-embedding fragile watermarking for image tamper detection and recovery. Section 3 presents some common attacks performed on watermarked images. Section 4 describes the evaluation indexes of this fragile watermarking scheme. The literature review is expatiated in Section 5. Conclusions and future work are discussed in Section 6.

2. Basic Model

In this paper, we focus on the self-embedding fragile watermarking for image tamper detection and

self-recovery. It includes three main stages: watermark generation and embedding, tamper localization, and image restoration. The detailed steps are described as follows.

Currently, most fragile watermarking schemes with recovery ability are accomplished at block-level. The host image is first decomposed into many non-overlapping blocks. For each block, the authentication watermark is derived, which is designed to identify and locate the modified regions. To achieve this goal, the authentication watermark is usually the feature or hash code of image block. The recovery data is usually a highly compressed version of original image block, which has different forms like average pixel value and quantized DCT coefficients. Fig. 1 shows the basic block diagram of general fragile watermarking scheme for image authentication and self-recovery. On the sending side, the authentication watermark and recovery watermark are firstly derived from the host image, respectively. To improve the security of algorithm, the watermark is often encrypted by some encryption methods before it is embedded into the host image. These two kinds of watermarks are then compressed and encoded into binary bits. To achieve tamper detection and recovery, the authentication watermark is embedded into the image block itself, while the recovery watermark is embedded into another block determined by the block mapping sequence. After watermark embedding, the watermarked image is obtained. Fig. 1(a) illustrates the watermark generation and embedding processes, which can be formulated as:

$$\boldsymbol{W} = f(\boldsymbol{I}, K), \ \boldsymbol{I}_{\mathrm{w}} = F(\boldsymbol{I}, \boldsymbol{W}), \tag{1}$$

where $\boldsymbol{I}$ and $\boldsymbol{I}_{\mathrm{w}}$ are host image and watermarked image, respectively; K is the secret key used for encryption; and $\boldsymbol{W}$ is the watermarking message produced by the rule f. With the embedding rule F, the watermark information including authentication data and recovery data are inserted into host image $\boldsymbol{I}$.

Fig. 1(b) shows the tamper detection and image recovery processes. On the receiving end, the authentication data of each block is first extracted from the block itself and compared with the original authentication data. Since the authentication watermark is closely related to the feature of image block, the tampering operation will lead to mismatch between the extracted watermark and regenerated authentication information. In this way, the tampered region can be identified. If the image block is determined as falsified block, the corresponding recovery data is then extracted from its mapped block, which will be used to restore the tampered region approximatively.

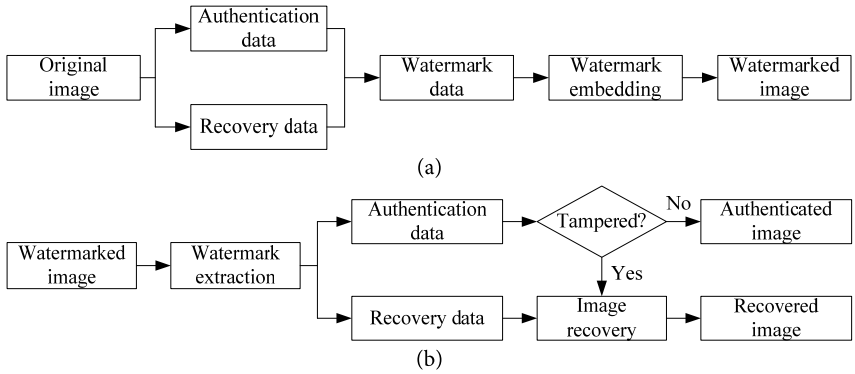


Fig. 1. General fragile watermarking scheme for image authentication and self-recovery: (a) watermark generation and embedding processes and (b) tamper detection and image recovery.

3. Common Attacks

In the process of transmission, digital images might be manipulated illegally by attackers. The good fragile watermarking scheme should be sensitive to any possible attack. In addition to some general attacks like delete operation and copy-move operation, some special attacks on specific algorithms have been put forward in recent years. Compared with general attacks, these special attacks are more difficult to be detected, because they are proposed based on the algorithm loophole. In this section, four special attacks that are commonly used in fragile watermarking schemes are presented.

3.1 Collage Attack

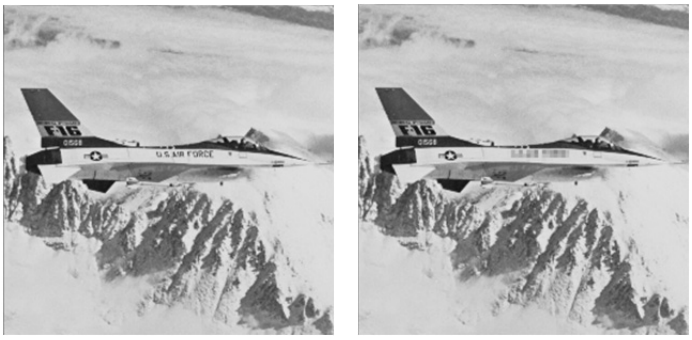
Collage attack proposed in [6] is a manipulation on block-wise independent watermarking schemes. Unlike copy-move operation, the collage attack can create a new watermarked image from multiple authenticated watermarked images by combining parts of different images and keeping their relative spatial locations unchanged. Besides, all the watermarked images used in collage attack are generated by the same watermarking method with identical secret keys. Fig. 2 gives an example of collage attack. As shown in Fig. 2(c), the airplane in watermarked image Airplane is inserted into watermarked image Boat without changing its relative position.

(a) (b) (c)

Fig. 2. Collage attack: (a) watermarked image Airplane, (b) watermarked image Boat, and (c) splicing image.

3.2 Constant Average Attack

In some self-embedding fragile watermarking techniques, the average pixel value of image block is usually served as recovery data to achieve image restoration. Despite its simplicity, different blocks might have the same average intensity. Based on security analysis, Chang et al. [7] proposed a constant average attack. To conduct this attack, the host image is first split into a lot of non-overlapping blocks according to the embedding rules. For each block, the watermarking bits are kept untouched, and a forged block with the same average pixel values is utilized to replace original watermarked block. In this way, the watermark information will not be changed, but the original image content has been tampered already. Fig. 3 shows an example of constant average attack. The tampered image is shown in Fig. 3(b), where the logo on the airplane is removed by constant average attack.



(a) (b)

Fig. 3. Constant average attack: (a) watermarked image Airplane and (b) tampered image.

3.3 Content-only Attack

The content-only attack is similar to constant average attack. It forges the image content while keeping the watermarking bits unchanged. For example, if the watermark information is inserted into 2 LSBs of image pixels, we can extract the 2 LSBs in each pixel first and then manipulate the rest image arbitrarily. After tampering, we insert the watermark information back into the image. At last, we get the tampered image.

3.4 Vector Quantization Attack

Vector quantization (VQ) attack [8] is another counterfeiting attack that is directed against block-wise independent watermarking schemes. In block-wise independent watermarking, each watermarked block depends only on its original block. Given a set of watermarked image blocks, a VQ codebook can be generated. Based on this codebook, a forged image containing the fake watermark is constructed.

4. Evaluation Indexes

Watermark invisibility, tamper detection performance, and the capability of image recovery are three main evaluation criterions in fragile watermarking schemes for image authentication and recovery. Like most watermarking schemes, the embedded watermark should be perceptually invisible. It should not affect the normal use of host image. Generally, the better the imperceptibility is, the higher the security of watermark will be. However, the original image would be distorted more or less, no matter how many pixels are changed by watermark. The tamper detection performance determines the effect of image recovery. The higher tamper detection accuracy will help to achieve the better image recovery result. In addition, the watermark payload and the security of algorithm are another two aspects that should be considered in watermarking schemes.

4.1 Image Quality Evaluation

At present, the watermark invisibility is evaluated by the quality of watermarked image. The better the quality of watermarked image is, the better the invisibility of watermark will be. The capability of

image recovery is generally measured by the quality of restored image. If the recovered image has better quality, it indicates that the self-embedding fragile watermarking has good capability for image recovery. The peak signal-to-noise ratio (PSNR) [9] and structural similarity (SSIM) index [10] are two quality evaluation indexes adopted in most watermarking schemes. For an image with size of $M \times N$, the PSNR can be defined as:

$$\mathrm{PSNR} = 10\log_{10}\frac{255^2}{\mathrm{MSE}}(\mathrm{dB}), \tag{2}$$

$$\mathrm{MSE} = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\boldsymbol{I}(i,j) - \boldsymbol{I}'(i,j)\right]^2, \tag{3}$$

where MSE is the mean square error between the original image $\boldsymbol{I}$ and the processed image $\boldsymbol{I}'$.

Although PSNR is the most widely used index in image quality evaluation, it does not consider the human visual system (HVS). Many experimental results show that the value of PSNR is not completely consistent with the perceptual result of human eyes. In other words, an image with higher PSNR may look like worse than the one with lower PSNR. This is because the HVS is not absolutely susceptive to the visual error. In addition, it is affected by many other external factors. SSIM overcomes the defect of PSNR and becomes a powerful index for similarity measurement. The definition of SSIM is given by:

$$\mathrm{SSIM} = l(\boldsymbol{I},\boldsymbol{I}')c(\boldsymbol{I},\boldsymbol{I}')s(\boldsymbol{I},\boldsymbol{I}'), \tag{4}$$

where $l(\boldsymbol{I},\boldsymbol{I}')$, $c(\boldsymbol{I},\boldsymbol{I}')$, and $s(\boldsymbol{I},\boldsymbol{I}')$ are the luminance, contrast, and structure comparison functions between two images, respectively, which can be expressed as:

$$l(\boldsymbol{I},\boldsymbol{I}') = \frac{2\mu_I\mu_{I'} + C_1}{\mu_I^2 + \mu_{I'}^2 + C_1},\ c(\boldsymbol{I},\boldsymbol{I}') = \frac{2\sigma_I\sigma_{I'} + C_2}{\sigma_I^2 + \sigma_{I'}^2 + C_2},\ s(\boldsymbol{I},\boldsymbol{I}') = \frac{\sigma_{II'} + C_3}{\sigma_I\sigma_{I'} + C_3}, \tag{5}$$

where μ_I and $\mu_{I'}$ are the average values of the two images $\boldsymbol{I}$ and $\boldsymbol{I}'$, respectively; σ_I and $\sigma_{I'}$ are their variances, respectively; and $\sigma_{II'}$ is the covariance between $\boldsymbol{I}$ and $\boldsymbol{I}'$. C_1, C_2, and C_3 are positive parameters. The value of SSIM is between 0 and 1. When SSIM is equal to 1, it means that the two images $\boldsymbol{I}$ and $\boldsymbol{I}'$ are totally the same.

4.2 Tamper Detection Performance

It is generally known that effective evaluation measurements for tamper detection play a vital role for image restoration. False positive rate (FPR) R_{FP} and false negative rate (FNR) R_{FN} [11] are two common indexes that are usually applied in fragile watermarking schemes for image authentication. The FPR reflects the ratio that authentic pixels are falsely detected as tampered pixels, and the FNR refers to the ratio that tampered pixels are incorrectly detected as authentic pixels. The definition of R_{FP} and R_{FN} are expressed as follows, respectively:

$$R_{\mathrm{FP}} = \frac{N_{\mathrm{FP}}}{N_{\mathrm{FP}} + N_{\mathrm{TN}}},\ R_{\mathrm{FN}} = \frac{N_{\mathrm{FN}}}{N_{\mathrm{FN}} + N_{\mathrm{TP}}}, \tag{6}$$

where N_{FP} denotes the number of authentic pixels that are falsely judged as tampered pixels; N_{TN} represents the number of authentic pixels that are correctly determined as authentic pixels; N_{FN} is the number of distorted pixels that are falsely judged as authentic pixels; and N_{TP} refers to the number of tampered pixels that are determined as falsified pixels properly. Generally, the lower FPR and FNR values indicate the better localization accuracy.

4.3 Watermark Payload

To get recovered image with better image quality, the recovery data has to contain enough information about image block. However, at the same time, too much watermark information will affect the quality of watermarked image and reduce watermark invisibility severely. Therefore, when we embed watermark into host image, we have to make a compromise between watermark capacity and watermark imperceptibility. In other words, we should reduce the watermark capacity as much as possible and guarantee the quality of recovered image at the same time. In fragile watermarking for image tamper detection and recovery, the watermark capacity is measured by the average watermarking bits in each pixel. Its unit is bit per pixel (bpp).

4.4 Watermark Security

Watermark security is a major problem in all digital watermarking algorithms. During the process of transmission, attackers can acquire the watermarked image and forge its watermark information by exploiting the security hole. To ensure the safety of embedded watermark, the watermarking message is usually encrypted before it is embedded into host image. On the receiving end, the real watermark is obtained after the corresponding decryption process. The general encryption methods include hash function, chaotic map, cat map [12], and so on.

5. Literature Review

In last decades, many fragile watermarking schemes have been put forward. According to different functions, these watermarking schemes can be divided into two types: fragile watermarking for tamper detection and fragile watermarking with recovery ability. The former can just locate the modified regions when host image is attacked. The fragile watermarking with recovery ability not only can locate the forged regions, but also can further restore the tampered regions roughly. The main idea of this watermarking scheme is that it embeds a highly compressed version of the image into the image itself, which is also called as self-embedding watermarking scheme. By image recovery, people can reuse the recovered images without retransmission. Therefore, this watermarking scheme is more appropriate to practical applications compared with the fragile watermarking only for image authentication. In this section, we give the thorough review about self-embedding fragile watermarking for image tamper detection and self-recovery.

In this field, Fridrich and Goljan [13] made an earlier attempt. They proposed a fragile watermarking algorithm with self-correcting capability. The DCT coefficients of each image block are

encoded and embedded into the LSBs of another block. When tampered block is identified, the recovery data is extracted and used to restore the tampered block. To improve the detection accuracy, Lin et al. [14] proposed a hierarchical fragile watermarking scheme. In their work, the host image is segmented into non-overlapping blocks, and 6 most significant bits (MSBs) of the average intensity in each block are adopted as the recovery data. The biggest strength of this method is that it adopts three-level hierarchical structure in tamper detection, which can highly improve the tamper detection rate. However, different blocks might have the same average intensity, which makes it vulnerable to constant average attack. To resolve this issue, Chang and Tai [15] presented a block-based digital watermarking. To resist constant average attack and collage attack, the parity check and intensity-relation check are used in watermarking generation. The experimental results show that it can resist VQ attack, constant average attack, and collage attack effectively. To improve the security of fragile watermarking scheme, many fragile watermarking schemes based on chaotic system have been proposed. Tong et al. [16] suggested a chaos system based fragile watermark algorithm for image authentication and recovery in which a new chaotic map called two-dimensional cross chaotic map was applied as encryption method. The method achieves superior tamper detection rate, especially when the forged regions are relatively large. However, it leaves some traces on the recovered areas, which is intolerable for human eyes. Chen et al. [17] proposed a chaos-based self-embedding watermarking with flexible watermark payloads. In their scheme, the 2×2 blocks are first classified into smooth blocks and rough blocks. For different blocks, different compression codes are used to generate recovery data. To further improve the quality of recovered image, many fragile watermarking schemes use transform coefficients to recover the tampered image. D. Singh and S. K. Singh [18] proposed a DCT based fragile watermarking method. In their scheme, two authentication bits and ten restoration bits are produced from each block with the size of 2×2. The first two largest DCT coefficients in each block are selected as recovery data and embedded into 3 LSBs of its mapped block. The detection accuracy is improved by utilizing blocks with small sizes and two-level detection method. Dadkhah et al. [19] presented a SVD based watermarking algorithm for image tamper detection and recovery. A mixed block-partitioning method for image blocks with the sizes of 4×4 and 2×2 is performed to enhance the detection precision of watermarking algorithm.

From the above analysis, we can see that the basic idea of the above literatures is to embed the watermark derived from a particular block into its mapped block. Though these methods have achieved great success in tamper detection and image recovery, a potential security problem still exists in these methods. If a block and its mapped block are destroyed at the same time, the tampered block would not be restored as we expect, which is called as tamper coincidence problem [20]. To overcome this problem, Lee and Lin [21] presented a dual fragile watermarking scheme with recovery capability. In their method, each block contains the watermark of other two blocks. Once the image block and its mapped block are both tampered, the tampered block can still be restored by another copy. In other words, it offers the second chance for image recovery. Based on [21], Som et al. [22] proposed dual self-embedding watermarking using DWT. The DWT approximation coefficients contain the most information of host image. Therefore, the 5 MSBs of approximation coefficients in each block are used as the recovery bits to restore the tampered block. Unfortunately, these methods cannot resolve the tamper coincidence problem radically. When the two copies of watermark for tampered block are both

destroyed, the tampered block would not be retrieved. In [20], Zhang et al. firstly introduced reference sharing mechanism into fragile watermarking scheme for image recovery. A reference derived from 5 MSBs planes of host image is served as the watermark and shared by the whole image for image recovery. Experimental results indicate that this method can avoid the tamper coincidence problem effectively, when the tampered regions are not too large. Since then, a great deal of work based on reference sharing mechanism has sprung up [23,24]. Qian et al. [25] proposed an image self-embedding watermarking with multi-level encoding. According to the degree of smoothness, image blocks are first divided into different types. For different kinds of blocks, different numbers of DCT coefficients are selected and encoded into variable lengths, which are utilized to obtain reference bits for image recovery. This method takes advantage of energy concentration characteristic in DCT transform and achieves high-quality restoration for tampered image. Nevertheless, the later research found that this method cannot resist collage attack. In addition, the watermark capacity is fixed for different images, regardless that the image is smooth or rough. To resolve watermark wasting problem, Huo et al. [26] presented a fragile watermarking with alterable watermark capacity. For each block, the alterable-length codes are derived based on the roughness of image block, and then they are divided into three parts. By a user key, the watermark is embedded into another three blocks. On the receiving end, two copies of the significant-codes are used to restore the image. However, this method is poor at detecting random block missing attack. Qin et al. [27] proposed a fragile watermarking based on reference-data interleaving mechanism and adaptive selection of embedding mode. Different from the earlier methods, the reference bits in this algorithm are generated from the interleaved and scrambled MSBs planes. Two embedding modes including overlapping-free embedding and overlapping embedding are adaptively adopted to achieve satisfactory performance for different tampering ratio.

Recently, fragile watermarking technique for image tamper detection and self-recovery has got great development in biology and medicine. In addition to the common metrics mentioned above, the fragile watermarking for medical images or biometric images also has to preserve the recognition rate of host images. If medical image is tampered in the process of transmission, it will lead to serious misdiagnosis for the state of an illness. Eswaraiah and Reddy [28] proposed a medical image watermarking for tamper detection and recovery. In their algorithm, the medical image is partitioned into region of interest (ROI) and region of background (ROB). Generally, doctors and patients pay more attention to ROI. To reduce the effect of watermark embedding on ROI, the authentication data is inserted into ROI while the recovery data is inserted into ROB. Experimental results indicate that it achieves high quality restoration for ROI. However, the ROB of medical images cannot be recovered in their method. Liew et al. [29] proposed a lossless recovery fragile watermarking based on ROI segmentation and multi-level authentication. The image information in ROI is all embedded into ROB, which achieves lossless recovery for ROI. Li et al. [30] presented a salient region-based fragile watermarking for biometric images. The multi-level authentication scheme is designed to verify the integrity of biometric images. The Eigen-face coefficients of biometric image are utilized as recovery watermark to restore the tampered image. In addition, the salient regions which contain more information of images are embedded by more watermark information. Compared with other methods, this scheme can resist VQ attack and collage attack effectively. To make better comparisons, Table 1 summarizes the self-embedding fragile watermarking algorithms mentioned above.

Table 1. Different fragile watermarking schemes for image authentication and self-recovery

Algorithm	Block size	Payload (bpp)	Recovery data	Application
Fridrich and Goljan [13]	8×8	2	DCT coefficients	No specific applications suggested
Lin et al. [14]	4×4, 2×2	2	Average pixel values	
Chang and Tai [15]	2×2	2	Average pixel values	
Tong et al. [16]	2×2	3	Average pixel values	
Chen et al. [17]	2×2	1.62–2.19	Average pixel values	
D. Singh and S. K. Singh [18]	2×2	3	DCT coefficients	
Dadkhah et al. [19]	4×4, 2×2	2	Average pixel values	
Zhang et al. [20]	8×8	3	5 MSBs planes	
Lee and Lin [21]	2×2	3	Average pixel values	
Som et al. [22]	2×2	3	Approximate wavelet coefficients	
Qian et al. [25]	8×8	3	DCT coefficients	
Huo et al. [26]	8×8	1.18–1.84	DCT coefficients	
Eswaraiah and Reddy [28]	4×4 (ROI) 3×3 (ROB)	-	Average pixel values	Medicine
Liew et al. [29]	40×40 (ROI) 2×2 (ROB)	-	All the pixels in ROI	
Li et al. [30]	4×4	3 (ROI) 1 (ROB)	Eigen-face coefficients	Biology

6. Conclusions

Currently, digital images have been broadly used in our daily lives. How to protect the authenticity of images and recover the tampered regions when they are tampered becomes the focus of study. The fragile watermarking scheme for image authentication and recovery provides an effective solution to resolve this issue. The watermarks in this scheme include two forms: authentication watermark for tamper detection and recovery watermark for image recovery. Generally, the recovery watermark is usually a highly compressed version of the host image. Therefore, this watermarking scheme is also known as self-embedding fragile watermarking. In this paper, we make a review on self-embedding fragile watermarking schemes for image tamper detection and self-recovery. The basic model and common attacks as well as the evaluation indexes of this scheme are described. In addition, we have discussed the existing fragile watermarking methods with self-recovery capability, which were proposed in recent years. From the review and comparisons among different methods, we can see that there are four main problems in existing self-embedding fragile watermarking schemes:

(1) The current methods in this field cannot resist all the attacks, especially some special attacks. Therefore, the security of the algorithm is still a serious problem that needs to be improved.

(2) Most self-embedding watermarking methods adopt fixed embedding mode for different images or blocks, regardless they are smooth or rough. In other words, the watermark capacity is fixed for different images, which leads to serious watermark wasting problem.

(3) At present, most fragile watermarking schemes with recovery ability are based on image blocks. In other words, the tamper localization accuracy is limited to block-level. It cannot give sufficient support for the subsequent image recovery process.

(4) The average pixel values or DCT coefficients of image blocks are commonly used as recovery data to restore the tampered image. However, the restoration effect is not always satisfactory. The recovered images usually have serious block artifacts, especially when the tampered regions are relative large.

Therefore, a fragile watermarking method with adaptive embedding rules and high-quality recovery capability is urgently needed. The recovery data and watermark capacity should be different for different blocks or images. In other words, the new fragile watermarking should be adaptive to different images. In the future work, we will further research the self-embedding methods and propose a novel fragile watermarking method for image authentication and recovery based on the above analysis.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61702303, No. 61201371); the Natural Science Foundation of Shandong Province, China (No. ZR2017MF020, No. ZR2015PF004); and the Research Award Fund for Outstanding Young and Middle-aged Scientists of Shandong Province, China (No. BS2013DX022).

References

[1] D. C. Lou and J. L. Liu, "Fault resilient and compression tolerant digital signature for image authentication," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 1, pp. 31-39, 2000.

[2] C. S. Lu and H. Y. M. Liao, "Structural digital signature for image authentication: an incidental distortion resistant scheme," *IEEE Transactions on Multimedia*, vol. 5, no. 2, pp. 161-173, 2003.

[3] H. Nyeem, W. Boles, and C. Boyd, "Digital image watermarking: its formal model, fundamental properties and possible attacks," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1-22, 2014.

[4] S. Bekkouch and K. M. Faraoun, "Robust and reversible image watermarking scheme using combined DCT-DWT-SVD transforms," *Journal of Information Processing Systems*, vol. 11, no. 3, pp. 406-420, 2015.

[5] K. Sreenivas and V. K. Prasad, "Fragile watermarking schemes for image authentication: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 8, pp. 1-26, 2017.

[6] J. Fridrich, M. Goljan, and N. Memon, "Cryptanalysis of the Yeung-Mintzer fragile watermarking technique," *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 262-274, 2002.

[7] C. C. Chang, Y. H. Fan, and W. L. Tai, "Four-scanning attack on hierarchical digital watermarking method for image tamper detection and recovery," *Pattern Recognition*, vol. 41, no. 2, pp. 654-661, 2008.

[8] M. Holliman and N. Memon, "Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes," *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 432-441, 2000.

[9] H. Zhang, C. Y. Wang, and X. Zhou, "Fragile watermarking based on LBP for blind tamper detection in images," *Journal of Information Processing Systems*, vol. 13, no. 2, pp. 385-399, 2017.

[10] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[11] H. Zhang, C. Y. Wang, and X. Zhou, "Fragile watermarking for image authentication using the characteristic of SVD," *Algorithms*, vol. 10, no. 1, article no. 27, 2017.

[12] O. Benrhouma, H. Hermassi, and S. Belghith, "Tamper detection and self-recovery scheme by DWT watermarking," *Nonlinear Dynamics*, vol. 79, no. 3, pp. 1817-1833, 2015.

[13] J. Fridrich and M. Goljan, "Images with self-correcting capabilities," in *Proceedings of the International Conference on Image Processing*, Kobe, Japan, 1999, pp. 792-796.

[14] P. L. Lin, C. K. Hsieh, and P. W. Huang, "A hierarchical digital watermarking method for image tamper detection and recovery," *Pattern Recognition*, vol. 38, no. 12, pp. 2519-2529, 2005.

[15] Y. F. Chang and W. L. Tai, "A block-based watermarking scheme for image tamper detection and self-recovery," *Opto-Electronics Review*, vol. 21, no. 2, pp. 182-190, 2013.

[16] X. J. Tong, Y. Liu, M. Zhang, and Y. Chen, "A novel chaos-based fragile watermarking for image tampering detection and self-recovery," *Signal Processing: Image Communication*, vol. 28, no. 3, pp. 301-308, 2013.

[17] F. Chen, H. H. He, H. M. Tai, and H. X. Wang, "Chaos-based self-embedding fragile watermarking with flexible watermark payload," *Multimedia Tools and Applications*, vol. 72, no. 1, pp. 41-56, 2014.

[18] D. Singh and S. K. Singh, "Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 775-789, 2016.

[19] S. Dadkhah, A. A. Manaf, Y. Hori, A. E. Hassanien, and S. Sadeghi, "An effective SVD-based image tampering detection and self-recovery using active watermarking," *Signal Processing: Image Communication*, vol. 29, no. 10, pp. 1197-1210, 2014.

[20] X. P. Zhang, S. Z. Wang, Z. X. Qian, and G. R. Feng, "Reference sharing mechanism for watermark self-embedding," *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 485-495, 2011.

[21] T. Y. Lee and S. D. Lin, "Dual watermark for image tamper detection and recovery," *Pattern Recognition*, vol. 41, no. 11, pp. 3497-3506, 2008.

[22] S. Som, S. Palit, K. Dey, D. Sarkar, J. Sarkar, and K. Sarkar, "A DWT-based digital watermarking scheme for image tamper detection, localization, and restoration," in *Applied Computation and Security Systems*. New Delhi, India: Springer, 2015, pp. 17-37.

[23] F. Cao, B. W. An, J. W. Wang, D. P. Ye, and H. L. Wang, "Hierarchical recovery for tampered images based on watermark self-embedding," *Displays*, vol. 46, pp. 52-60, 2017.

[24] C. Qin, P. Ji, X. P. Zhang, J. Dong, and J. W. Wang, "Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy," *Signal Processing*, vol. 138, pp. 280-293, 2017.

[25] Z. X. Qian, G. R. Feng, X. P. Zhang, and S. Z. Wang, "Image self-embedding with high-quality restoration capability," *Digital Signal Processing*, vol. 21, no. 2, pp. 278-286, 2011.

[26] Y. R. Huo, H. J. He, and F. Chen, "Alterable-capacity fragile watermarking scheme with restoration capability," *Optics Communications*, vol. 285, no. 7, pp. 1759-1766, 2012.

[27] C. Qin, H. L. Wang, X. P. Zhang, and X. M. Sun, "Self-embedding fragile watermarking based on reference-data interleaving and adaptive selection of embedding mode," *Information Sciences*, vol. 373, pp. 233-250, 2016.

[28] R. Eswaraiah and E. S. Reddy, "A fragile ROI-based medical image watermarking technique with tamper detection and recovery," in *Proceedings of the 4th International Conference on Communication Systems and Network Technologies*, Bhopal, India, 2014, pp. 896-899.

[29] S. C. Liew, S. W. Liew, and J. M. Zain, "Tamper localization and lossless recovery watermarking scheme with ROI segmentation and multilevel authentication," *Journal of Digital Imaging*, vol. 26, no. 2, pp. 316-325, 2013.

[30] C. L. Li, Y. H. Wang, B. Ma, and Z. X. Zhang, "Tamper detection and self-recovery of biometric images using salient region-based authentication watermarking scheme," *Computer Standards and Interfaces*, vol. 34, no. 4, pp. 367-379, 2012.

Chengyou Wang https://orcid.org/0000-0002-0901-2492

He received his M.E. and Ph.D. degrees in signal and information processing from Tianjin University, China, in 2007 and 2010, respectively. He is currently an associate professor and supervisor of postgraduate students at Shandong University, Weihai, China. His current research interests include image/video coding, digital watermarking, and tamper detection.

Heng Zhang https://orcid.org/0000-0003-1864-5432

He received his B.E. degree in communication engineering from Shandong University of Technology, China, in 2015. He is currently pursuing his M.E. degree in electronics and communication engineering at Shandong University, China. His current research interests include watermarking-based image authentication and tamper detection, and computer vision.

Xiao Zhou https://orcid.org/0000-0002-1331-7379

She received her M.E. degree in information and communication engineering from Inha University, Korea, in 2005; and her Ph.D. degree in information and communication engineering from Tsinghua University, China, in 2013. She is currently a lecturer and supervisor of postgraduate students at Shandong University, Weihai, China. Her current research interests include channel estimation, image communication, and image watermarking.