EURASIP Journal on Image and Video Processing

REVIEW Open Access

A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability

Lusia Rakhmawati[1,2*], Wirawan Wirawan[1] and Suwadi Suwadi[1]

Abstract

The rapid developments of storage technology and information exchange encourage the development of research in the field of information security. In the process of sending information, tamper and issues about data ownership may occur. The fragile watermarking is one technique that can overcome the problem, in which this technique exploits sensitivity to tamper with the inserted watermark components. Therefore, it is not tolerable to change even only one bit. Self-embedding can be defined as some important features obtained from a selected image as a watermark by modifying the pixel value of the original image. Once a picture which has been inserted with a watermark is modified by other users, it can be used for tamper detection and recovery images. Due to the importance of this fragile watermarking scheme, this paper discusses the principles and characteristics of a fragile watermarking algorithm. The main contribution of this paper survey is that it summarises the current mechanisms of selection, generation, method of watermark insertion, detection and tamper localisation and recovery procedures. Comparison of several watermarking techniques was analysed and presented in tabular form, as well as experimental evaluation of four watermarking schemes in many graphs to show the performance of the self-embedding fragile watermarking scheme.

Keywords: Fragile watermarking, Self-embedding, Image authentication, Image recovery

1 Introduction

Currently, the availability of image processing technology is highly diverse, thus an image content becomes easily manipulated by irresponsible parties. Replacement of content is very harmful to the owner especially when it comes to legal cases [1]. Therefore, research on detection and localisation of tampered images is an important issue. One approach that is widely used is the technique of digital watermarking. This is a technique of inserting secret information or image's information into an original image prior to sending. The original image is the image where the confidential information is inserted, while the secret information is a watermark [2–4]. The watermark component selection is done by considering the purpose of the watermarking technique. In general, the watermark used for image authentication and recovery uses two watermarks, which are bit authentication and bit recovery. For image recovery, bit recovery is selected from the image feature, usually from the original image compression form. When the watermark used is generated from the original image, it is known as the self-embedding watermarking scheme [2, 5–8].

In general, digital watermarking techniques can be categorised as robust watermarking, fragile watermarking and semi-fragile watermarking. Watermark is specifically selected to be able to withstand deliberate or intentional manipulation, in this case belonging to a strong watermarking category, commonly used for copyright and verification of the image [9, 10]. In the fragile watermarking technique that is being used for the authentication process, the embedded watermark vulnerability indicates that there are modifications either intentionally or unintentionally [2, 3, 11–24]. Meanwhile, in semi-fragile watermarking techniques, there are common image processing methods, such as JPEG compression, scaling, sharpening; and tidy against deliberate manipulations, such as content modified by false

* Correspondence: lusiarakhmawati@unesa.ac.id; lusiarakhmawati17071@mhs.its.ac.id
[1]Department of Electrical Engineering, Faculty of Electrical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

information, then the watermark is designed to be able to resist a variety of non-heretical attacks [4, 21, 25–29].

For the purposes of authentication, fragile watermarking schemes can be categorised into two main categories as fragile watermarking pixel-wise schemes [12, 18–20, 30] and fragile watermarking block-wise schemes [1, 3, 4, 10, 11, 17, 21–23]. In the pixel-wise fragile watermarking scheme, the watermark information is generated from the grey pixel value of the host and inserted into the host pixel. Meanwhile, in the fragile watermarking scheme of block-wise, the original image is subdivided into blocks. Each sub-block has watermark information and is authenticated with a watermark retrieval successfully embedded in it. If the watermarked image is modified, the watermark of a sub-block is not successfully retrieved, and then the sub-block is identified as a damaged or incorrect block. The fragile nature for authentication applications is also used in reversible data hiding techniques. According to [31], most reversible watermarking techniques are fragile and allow for extracting watermarks along with the complete restoration of the original image. Reversible data hiding methods for uncompressed images have been discussed in [32], where there are two criteria to show their performances, i.e. imperceptibility and planting capacity. Imperceptibility is a measure of similarity between watermarked and covers images. Meanwhile, the capacity of embedding is a measure of the maximum number of bits of information that can be pinned on the cover image. The least significant bit (LSB) method in data hiding has simplicity, easy detection and high embedding capabilities [24, 25]. The LSB method, therefore, is generally preferred for embedding watermarks on fragile image watermarking.

To facilitate the understanding of fragile watermarking, in addition to the discussion of conceptual understanding, this article also explains the comparison of methods ranging from insertion, extraction and recovery results. At the end, the analysis of the results of the evaluation is also described by selecting four methods that represent spatial domains to experiment with many tampered images. A block diagram in Fig. 1 provides a brief overview of the self-embedding fragile watermarking scheme for image authentication.

The organisation of this paper survey is described in Section 2, which is broken into the following subsections. Section 2.1 of this paper explains an overview of the basic principles of self-embedding fragile watermarking. Section 2.2 shows the mechanisms of selection, generation and method of watermark insertion. Section 2.3

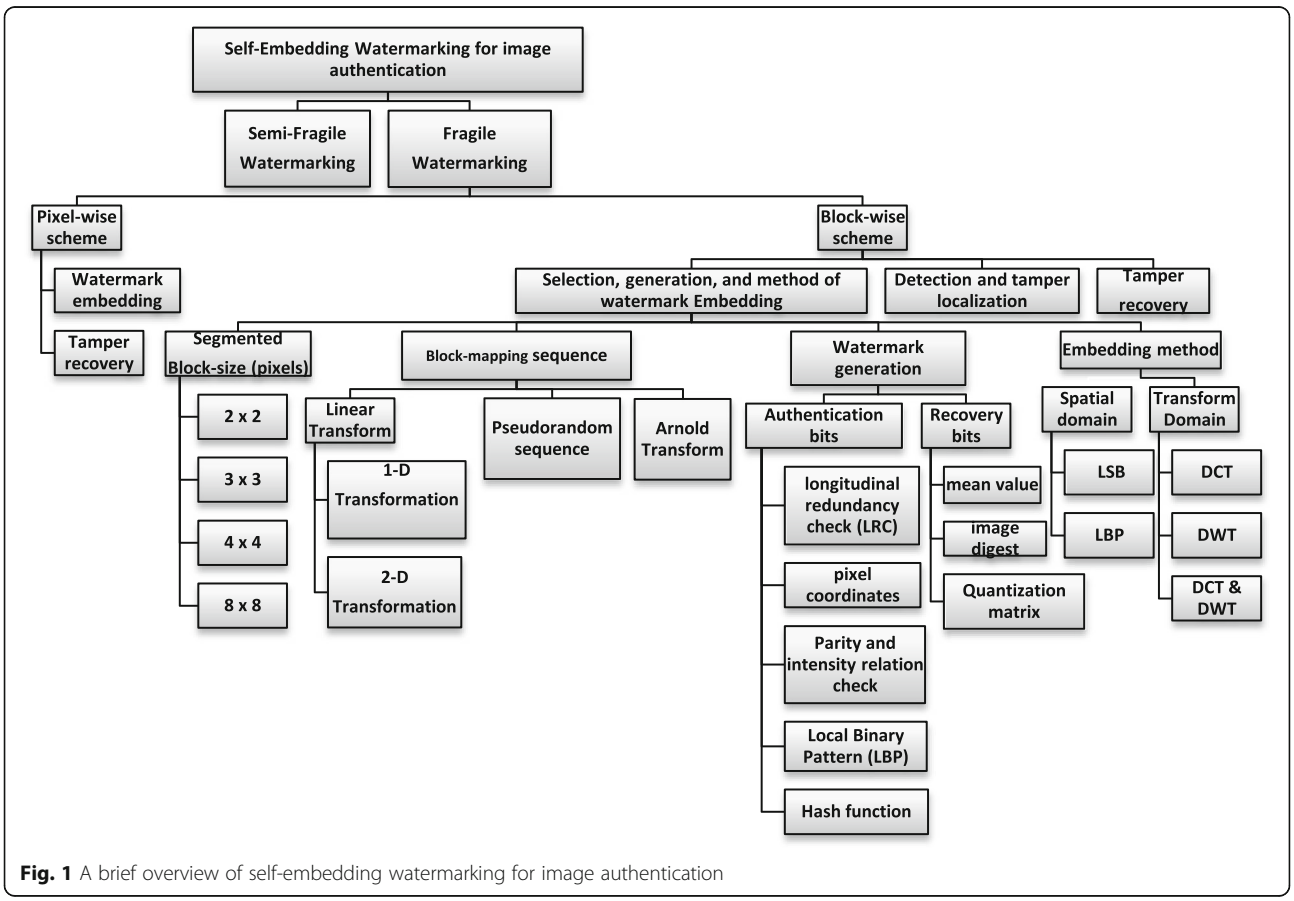

Fig. 1 A brief overview of self-embedding watermarking for image authentication

describes the comparison and discussion based on two domains. Section 2.4 discusses the experimental results of four watermarking techniques. Finally, the conclusion is given in Section 3.

2 Review

2.1 Basic principles of self-embedding fragile watermarking scheme

In recent years, more research in the field of fragile watermarking is aimed at the process of damaged image authentication and recovery. This method can be derived under distinct purposes. According to the mechanism of image capture feature, it can be divided into block-wise and pixel-wise mechanisms as shown in Fig. 1 from Cheng et al. [30]; the block-wise algorithm [1–6] divides the original image into small blocks to embed a watermark, which may be a hash of the main content taken from each block. The block-wise algorithm can locate the damaged block, but the blocks may contain some original pixels, so that the accuracy of the localisation placement is decreasing. This is the weakness of the block-wise algorithm. The pixel-based algorithm [7–9] generates data from the original pixel grey value and embeds the mark into the pixel. By calculating the consistency of the extracted watermark and its derivative signal, we can detect the damaged pixels. However, due to the limited precision of biplanes, it is possible that the data derived from the damaged pixels equal the watermark even though it has been modified. Therefore, modifications to these pixels are not directly detected.

There are two stages of the watermarking algorithm for image tamper detection and recovery; the embedding watermark phase and authentication phase which are then followed by tamper detection and recovery image [9], as shown in the Figs. 2 and 3.

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 authorised user will be able to recover it. The original cover image and the watermarked image are indistinguishable, and the recovered images should be of a high quality.
Tamper detection: a self-recovery fragile watermarking technique must be able to detect with or without an original image or some information to derive it for the detection process.
Tamper recovery: the scheme should be able to detect unauthorised 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 general attack, the collage attack and the disturbing attack.

In addition, some parameters used to measure the performance of watermarking algorithms are peak signal-to-noise ratio (PSNR) and bit-per-pixel (bpp). The value of PNSR can measure the perceptual comparison of the watermarked image to the original image [15, 33]. Meanwhile, bpp is used to calculate the amount of watermark data inserted in the original image. The idle bpp value is 1, but there is a certain watermarking algorithm using more than 1 bpp.

2.2 Mechanisms of selection, generation and method of watermark insertion

In the self-embedding watermarking system, in addition to the authentication process, the watermark component is selected based on the need for digital image recovery at the receiving end, thus the original image needs to be divided into the appropriate blocks, mapping the blocks before insertion by a specific method to be discussed in the following.

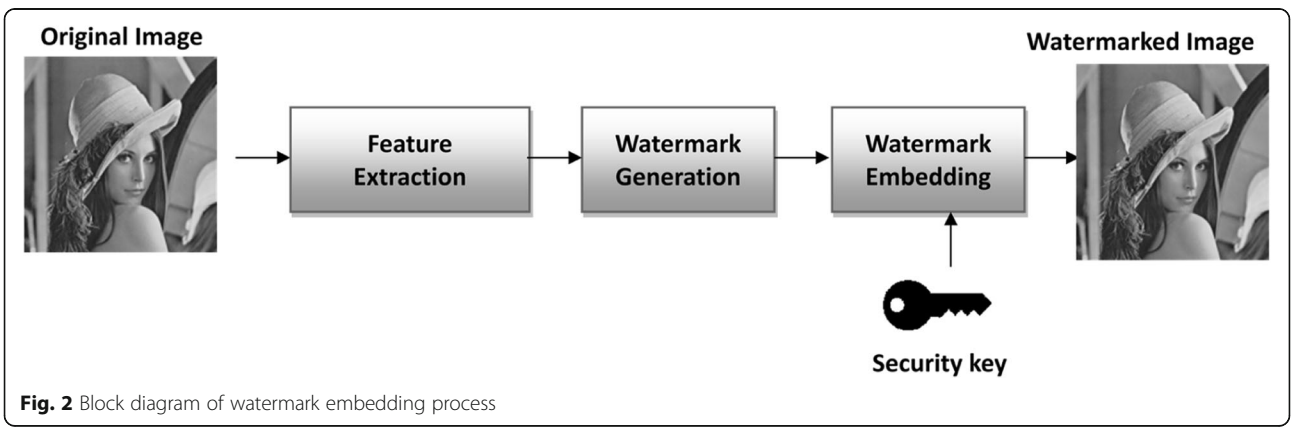

Fig. 2 Block diagram of watermark embedding process

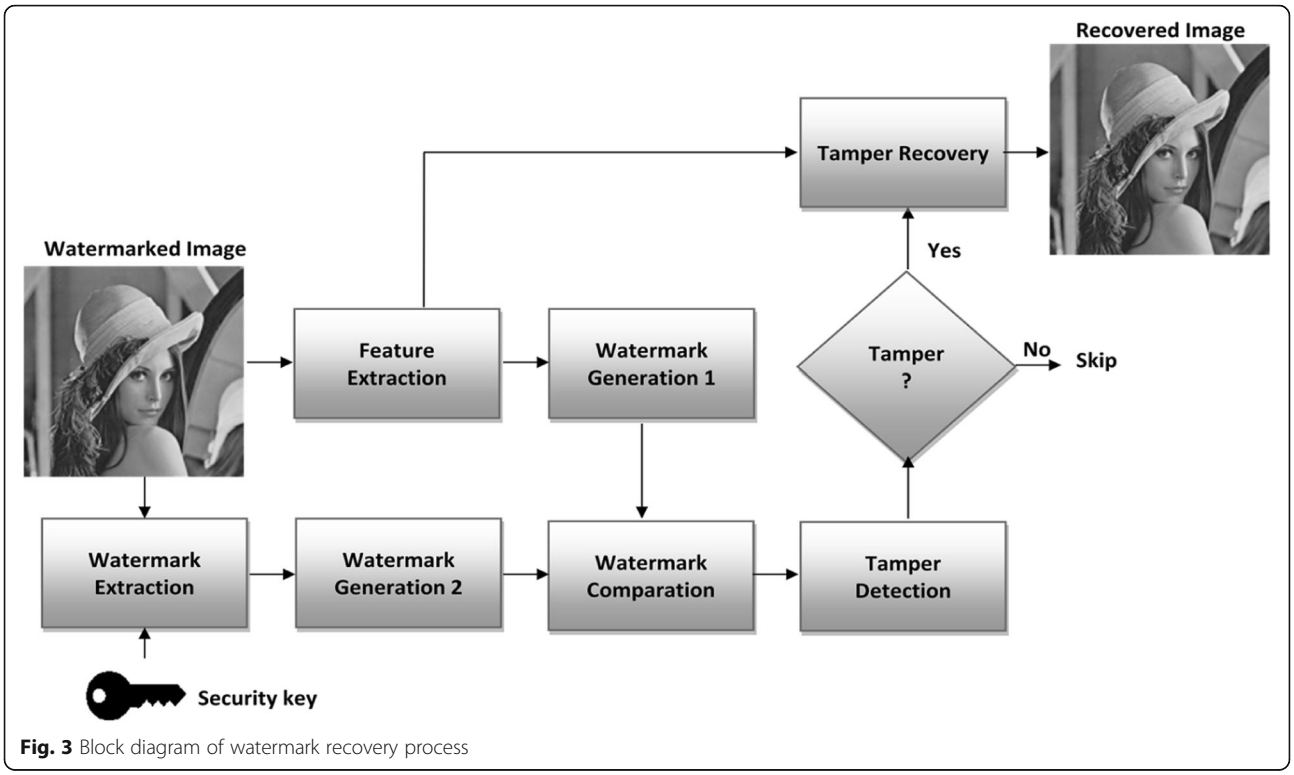

Fig. 3 Block diagram of watermark recovery process

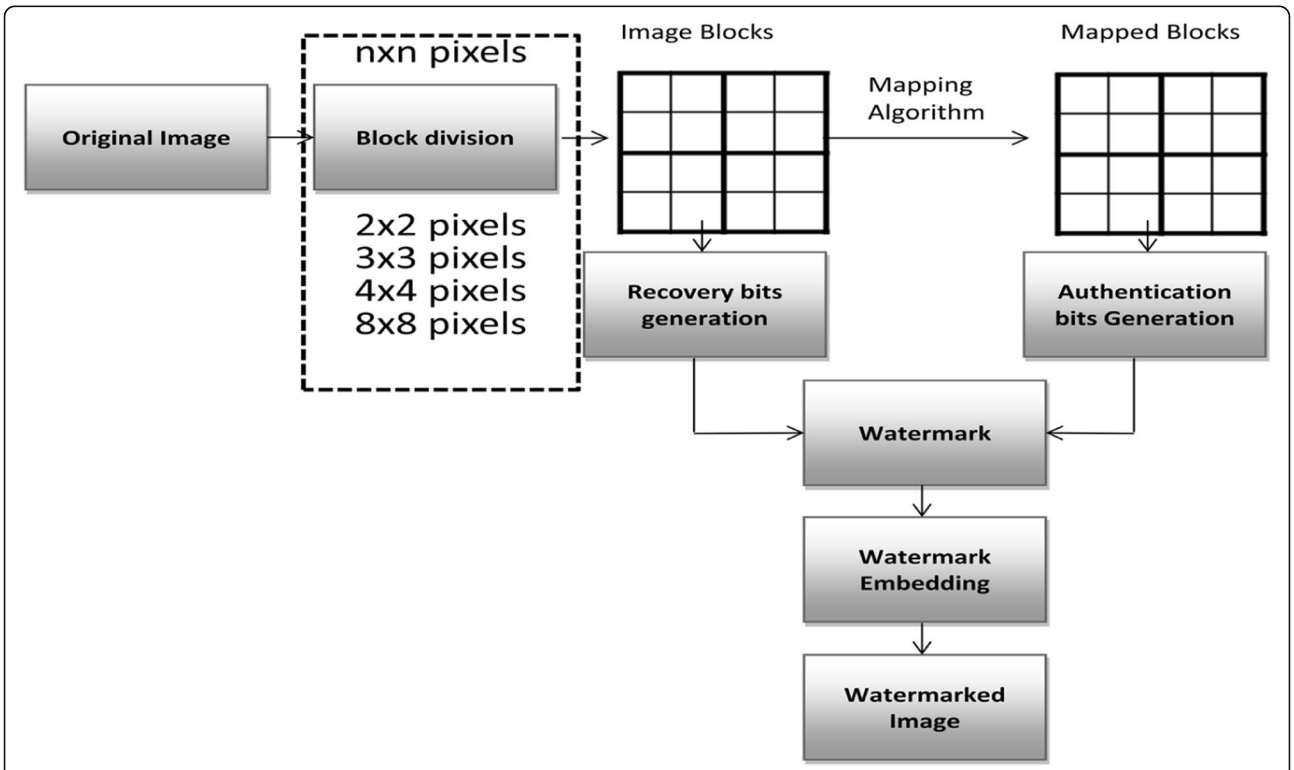

Fig. 4 Block diagram of the embedding procedure of Singh et al. [37], segmented block size modification-based fragile watermarking

2.2.1 Segmented blok size

Many researchers have conducted research in the modification of block size based on fragile watermarking. The watermark embedding process is shown in Fig. 2. In this procedure, most of the researchers divided the original image into blocks of uniform size ($n \times n$ pixel). So far, they have investigated experiments with blocks of different size, i.e. 2×2 pixels [34], 3×3 pixels[35], 4×4 pixels[36] and 8×8 pixels[23]. These blocks were used to generate more watermarks, as authentication and restoration bits.

In 2016, Singh et al. [37] developed a scheme using a small non-overlapping block sized 2×2 to improve the accuracy of localisation and effectively remove the blocking artefacts as illustrated at Fig. 4. The original image X of size $m \times n$ is divided into non-overlapping blocks of size 2×2 [34]. This argument was also expressed by Sreenivas et al. [38], stating that small block sizes generally allow for better tamper detection. This also allows better block encoding when textured blocks are smooth. Meanwhile, to localise tampering, the original image is divided into blocks of size 3×3 [35].

Xiao's method [36] divides an image into 4×4 non-overlapping blocks, generates authentication watermarks for the blocks by comparison and parity check among average intensities and embeds them into corresponding blocks. The recovery information of another block is embedded into its mapping block. Huo et al.'s scheme [15] proposes the recovery of discrete cosine transform (DCT-based) alternate self-recovery fragile watermarking scheme. The image was divided into non-overlapping sub-blocks with 8×8 sizes and sub-blocks classified into different types according to block variance. Experiments in [23] also show that the scheme can detect and localise damage by 8×8 pixels and can recover 40% tampered image. Chetan et al. [34] calculated the tamper detection accuracy and values were recorded for blocks of some various sizes. The smaller the block size is, the more accurate the tamper localisation is.

2.2.2 Block mapping sequence

In the self-embedding watermark scheme, block mapping is done before the watermark insertion process. In this case, a certain block feature will be inserted as a watermark payload for another block. In [16], it is mentioned that block mapping procedures are available but are generally grouped into linear transformations: 2-D transformations [17, 19] as shown in Eq. (1), and 1-D linear transformations [18, 20, 23] as shown in Eq. (2).

$$A = \begin{pmatrix} 1 & 1 \\ k & k+1 \end{pmatrix}, \begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix} = A\, x \begin{pmatrix} x_i \\ y_i \end{pmatrix} \ mod\ N, \tag{1}$$

Where a point (x_{i+1}, y_{i+1}) can be transformed from another point (x_i, y_i) and $(x_i, y_i) \in [0, N-1]$ x $[0, N-1]$ and $k \in [0, N-1]$, N is the total number of blocks in the image.

For 1-D transformation, a one-to-one mapping sequence was obtained as shown in Eq. 2

$$X' = [f(X) = (k * X)\ mod\ N] + 1 \tag{2}$$

Where X, $X' \in [0,\ N-1]$, k is a secret key and N is the total number of blocks in the image.

However, due to the limited number of degrees of freedom, linear transformations can be easily reconstructed by only experimenting with some sample images, and other weaknesses are low security [25, 26]. To solve this problem, He et al. [15] have proposed a nonlinear block mapping construction using a pseudorandom sequence. A block mapping sequence B is computed from a key based pseudorandom permutation $[B(1) \ldots B(N)]$ of the integer interval $[1 \ldots N]$. The differences between pseudorandom and 1-D transformation can be seen in Fig. 5. Where for an even number key, on 1-D transformation there is a repetition, so it does not produce one-to-one mapping.

Another method for pixel scrambling has been adopted in Chow et al.'s Scheme [39], which is the Arnold transform. By using the Arnold transformation shown in Eq. 3, high pixel correlation can be dissociated.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix}\ mod\ N = \begin{bmatrix} 1 & p \\ q & pq+1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix}\ mod\ N \tag{3}$$

Where (x, y) is the initial pixel positions, (x', y') is the positions after transformation and p and q are positive integers, which are used to determine the period of a given matrix. Det (A) = 1, N is the total number of blocks in the image. Correlations between adjacent pixels can be solved by applying multiple iterations.

Shruthy et al. [8] and Yu et al.'s [12] methods use pseudorandom sequences to generate nonlinear block mapping and use an optimal environmental characterisation method to detect the disturbance. Method [8] also investigates three optimisation strategies which will further improve the quality of localisation and tamper recovery. They take all the blocks of adjacent test blocks and their mapping blocks and then take advantage of statistical rules to determine the validity of the image blocks. Based on the false acceptance analysis and the probability of a false rejection, post-processing operations are presented to improve the further performance of the proposed human resources. This includes three steps: (1) marking the dubious block, (2) distinguishing blocks that are damaged from dubious blocks with adjacent blocks and (3) improving detection performance by post-detection processing [12].

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56
57	58	59	60	61	62	63	64

(a)

61	46	54	11	7	56	26	41
16	42	32	50	58	44	23	21
48	34	4	52	39	15	2	22
6	53	5	35	29	13	43	37
1	57	20	14	47	8	28	31
38	64	51	25	59	55	19	24
9	33	62	30	36	60	63	17
49	3	10	40	45	27	12	18

(b)

64	5	10	15	20	25	30	35
40	45	50	55	60	1	6	11
16	21	26	31	36	41	46	51
56	61	2	7	12	17	22	27
32	37	42	47	52	57	62	3
8	13	18	23	28	33	38	43
9	33	62	30	36	60	63	17
24	29	34	39	44	49	54	59

(c)

4	21	48	5	27	64	57	53
30	39	33	52	58	2	55	9
38	46	49	40	29	43	18	8
11	45	54	13	7	26	14	6
35	60	10	15	34	25	47	31
36	22	42	56	12	24	41	51
50	23	17	3	20	63	59	19
61	1	37	28	62	44	16	32

(d)

7	13	19	25	31	37	43	49
55	61	3	9	15	21	27	33
39	45	51	57	63	5	11	17
23	29	35	41	47	53	59	1
7	13	19	25	31	37	43	49
55	61	3	9	15	21	27	33
39	45	51	57	63	5	11	17
23	29	35	41	47	53	59	1

(e)

Fig. 5 Illustration of 8 × 8 mapping block. **a** Original image block mapping. **b** Mapping block using a pseudorandom sequence with key = 77. **c** Mapping block using 1-D linear transformations with key = 77. **d** Mapping block using a pseudorandom sequence with key = 6. **e** Mapping block using 1-D linear transformations with key = 6

INPUT: nth block pixels of cover image and the pixel coordinates: $P_n(i,j), P_n^r(i,j)$ and $P_n^c(i,j)$

procedure A_{b1} BIT GENERATION

$b_n(1:8) \longleftarrow dec2Bin(P_n(i,j),8), 1 \leqslant n \leqslant M/4, 1 \leqslant i \leqslant 2, 1 \leqslant j \leqslant 2$ ▷ 8 bits binary conversion of each pixel

$b_n(1:3) \longleftarrow 0$ ▷ Set 3LSBs of each pixel to zeros

$b_n^r(1:8) \longleftarrow dec2Bin(P_n^r(i,j),8)$ ▷ 8 bits binary conversion of row value of each pixel

$b_n^c(1:8) \longleftarrow dec2Bin(P_n(i,j),8)$ ▷ 8 bits binary conversion of column value each pixel

$pr_n(1:8) \longleftarrow b_n(1:8) \oplus b_n^r(1:8)$ ▷ bind pixel value with row value

$pc_n(1:8) \longleftarrow b_n(1:8) \oplus b_n^c(1:8)$ ▷ bind pixel value with column value

Matrix $M_{4\times4}$ is filled by $pr_n(1:8)$ and $pc_n(1:8)$

Generate a random number Key_2 using a seed value

$Key_1 = mod(Key_2, 16)$

$K_2 = dec2bin(Key_2, 4)$ ▷ dec2Bin(), used to convert decimal value into binary

$k_{4\times1} = M \times K_2'$

$a^1 = \sum_{m=1}^{4}(k(m)) mod \quad 2$

Similar to a^1, a^2, a^3 and a^4 calculated for remaining 3 pixels of the block

$A_{b1} = \sum_{m=1}^{4}(a^m) mod \quad 2$

end procedure

OUTPUT: A_{b1}

Fig. 6 First authentication bit generation of Singh's scheme [1]

2.2.3 Watermark generation

The watermark embedding phase occurs at the sender side. As mentioned earlier, the first step is the original image divided into non-overlapping blocks of size $n \times n$. Then a block mapping method is used to generate a chaotic map. Next step is generating authentication data and the feature information as recovery data, both of them form watermark component.

Singh [1] generated two authentication bits, which the first authentication bit algorithm is shown in Fig. 6 and the second algorithm is shown in Fig. 7. Generation of the first authentication bit (A_{b1}) was using most significant bit (MSB) value; in this method, it was using 5 bit MSB of 8 binary bits, while 3 bit LSB values for the insertion process was used. For the second authentication bit generation (A_{b2}), this method used check bit as a part of bit authentication; it implemented a longitudinal redundancy check (LRC) and pixel mean values for each block. Figures 6 and 7 represent the block diagram of the A_{b1} and the A_{b2} generation process in a spatial domain, where block pixel nth, $P_n(i,j)$, n indicates the position of the pixel block, where $1 \le n \le M/4$, $1 \le i \le 2$, $1 \le j \le 2$. Meanwhile, row and column values of $P_n(i,j)$ are symbolised by $P_n^r(i,j)$ and $P_n^c(i,j)$ respectively.

Other authentication bit generation schemes are given by Chang [3] where authentication bit information is generated through the local binary pattern (LBP) operator of each block as shown in Fig. 8.

$$S_x = \begin{cases} 1, if\ P_x \ge P_c \\ 0, if\ P_x < P_c \end{cases} \quad (4)$$

Where P_c is the centre pixel value, P_x is each neighbouring pixel value and S_x is the sign of each neighbouring pixel.

In addition, a hash function used by Kunhu [40] is considered as a content authentication bit information. In this scheme, they divide an image into B blocks non-overlapping. Therefore, 64 hexadecimal unique hash keys for each block by using SHA-256 were acquired and converted each hexadecimal digit into 4-bit binaries and converted them into (1×256) up to the information limit. Finally, all the shared blocks will generate B * 256-bit hash key information.

The embedded watermark data for image recovery are calculated from the original DCT of the host image and contain no additional redundancy [10, 26, 27, 29]. If the part of the watermarked image is damaged, the watermark data in the area can still be retrieved. If the amount of extracted data is greater, it will be able to reconstruct the original coefficient in the tampered area according to the given constraint. Otherwise, they may use press-sensing techniques to extract coefficients by utilising smoothness in the DCT domain. In this way, all extracted watermark data contribute to content recovery

INPUT: nth block pixels of cover image and the pixel coordinates: $P_n(i,j), P_n^r(i,j)$ and $P_n^c(i,j)$

procedure A_{b2} BIT GENERATION

$b_n(i,j) \longleftarrow \lfloor \frac{P_n(i,j)}{8} \rfloor, 1 \leqslant n \leqslant M/4, 1 \leqslant i \leqslant 2, 1 \leqslant j \leqslant 2.$ ▷ Delete the 3LSBs from each pixels

$m \longleftarrow \frac{\sum_{i=1}^{2}\sum_{j=1}^{2}(b_n(i,j))}{4}$ ▷ mean value of nth block pixels

for $i \longleftarrow 1$ to 2 **do** ▷ Start for loop for thresholding

  for $j \longleftarrow 1$ to 2 **do**

    if $b_n(i,j) \geqslant m$ **then** ▷ Apply threshold m on each pixel and generate binary matrix M_b

      $M_b(i,j) \longleftarrow 1$

    else

      $M_b(i,j) \longleftarrow 0$

    end if

  end for

end for ▷ Thresholding done

$a_1 \longleftarrow \sum_{i=1}^{2}\sum_{j=1}^{2}(M_b(i,j)) mod\ 2$ ▷ Convert into a single bit

Organize $b_n(i,j), \leqslant i \leqslant 2, 1 \leqslant j \leqslant 2$ in a matrix M of 4 rows and 5 columns

$a(1:5) = LRC(M)$ ▷ Apply LRC function

$a_2 \longleftarrow \sum_{i=1}^{5}(a(i)) mod\ 2$ ▷ Convert into a single bit

$A_{b2} = a_1 \oplus a_2$

end procedure

OUTPUT: A_{b2}

Fig. 7 Second authentication bit generation of Singh's scheme [1]

Fig. 8 The (8,1) local binary pattern used in Chang's [3] scheme [3]

[41]. The smaller the damaged area, the available watermark data will result in better quality recovered content. It also shows that in general, the proposed scheme is superior to the previous technique.

In Wu's scheme [35], the watermark component consists of watermark parity and two copies of the restoration watermark section. All watermarks are used for tamper detection. Therefore, with the same watermark load size, the tamper detection performance of the proposed scheme is better. The PSNR of the watermarked image is about 44 dB since only two LSB of each pixel are used for watermark insertion. Under general disruption, content disruption only and attack averages are constant, the probability of false acceptance (PFA), the probability of false rejection (PFR) and the probability of false detection (PFD) are close to zero for different tamper ratios from 0 to 80%. For collage attacks, the proposed new scheme is superior to Lee's scheme and He's method [23].

To improve the detection process of damaged image areas, Dhole [42] uses DCT transfers to obtain the information data used for the recovery process. Then, the watermark is inserted into the original image to get the first watermark image. In the next step, the original image blocks are scrambled and combined in reversed order from the previous block to the original image to get the watermarked image. Then, they perform an Ex-OR operation to get a watermarked image. If the image starts to be damaged, modifications to the watermarked image can also be identified in the self-embedded image. They can find a modified image by doing inverse DCT.

It is concluded from the survey that the existing methods for tamper detection and recovery require a lot of authentication and recovery data to be embedded. Generally, in self-embedding fragile watermarking, payload watermarks range from 1 to 3 bpp. With an increase in watermark payload, the PSNR value of the watermarked image will rise gradually [7] as described in Eq. (6) and (7) and Table 1. This significantly reduces the quality of watermarked image perceptions. Furthermore, localisation in tamper detection and recovery is not handled efficiently in existing works. The existing method also does not validate detection and tamper recovery against some insertion, deletion and attack updates. In the study [13], a fragile watermarking scheme is proposed to efficiently detect tamper information and restore damaged information. The scheme in [13] focuses on achieving higher quality regional recovery. In addition, improved authentication processes with higher accuracy on tamper detection are also focused on this method.

For the evaluation of self-embedding watermarking techniques, performance metrics such as imperceptibility (PSNR) and embedding capacity (bpp) are used. To investigate the performance of tamper detection algorithms, as mentioned in Section 5, we use PFR, PFA and PFD as the quantitative performance measures [16]. In terms of PSNR [43], this measures the quality of the watermarked images and the recovered images. PSNR is computed as Eq. (8).

$$\mathrm{PSNR} = 20 * \log_{10}(Max_I) - 10 * \log_{10}(\mathrm{MSE}) \qquad (5)$$

Where $MSE = \frac{1}{m*n}\sum_{i=0}^{m=1}\sum_{j=0}^{n=1}(X(i,j)-Y(i,j))^2$.
Here, X and Y represent the watermarked image and recovered image respectively and Max_I is the maximum possible pixel value of the image.

Generally, in the self-embedding fragile watermarking, the watermark payload ranges from 1 to 3 bpp. The quality of the watermarked image is indicated by the increase of the PSNR value, to reduce distortion caused by watermark insertion can be done by replacing the LSB value and keeping the remaining MSB value. In [16], supposed the original data in the LSB field is a uniform distribution, the average energy distortion emitted by the embedding watermark bit, α, is Eq. (6)

Table 1 Approximate PSNR value of the watermarked image under different number of the LSB

Number of LSB	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$	$\alpha = 4$	$\alpha = 5$
E_d	0.5	2.5	10.5	42.5	170.5
PSNR	51.14	44.15	37.92	31.85	25.81

$$E_d = \frac{1}{2^{2\alpha}} \sum_{i=0}^{(2^{\alpha}-1)} \sum_{j=0}^{(2^{\alpha}-1)} (i-j)^2 \quad (6)$$

Then, the approximate PSNR value of the watermarked image with respect to the original one is Eq. (7)

$$\text{PSNR} \approx 10 \ log_{10}\left(\frac{255^2}{E_D}\right) \quad (7)$$

In terms of PFA, PFR and PFD defined, respectively, as Eq. (8), (9) and (10)

$$\text{PFA} = 1-N_{td}/N_t \quad (8)$$

$$\text{PFR} = 1-N_{ab}/(N-N_t) \quad (9)$$

$$\text{PFD} = \frac{N_t}{N} x\ PFA + \left(1-\frac{N_t}{N}\right) x\ \text{PFR} \quad (10)$$

Where N is the blocks number, N_t is the number of actual blocks, N_{td} is the number of tampered blocks which are correctly detected and N_{ab} is the number of authentic blocks which are falsely detected.

2.2.4 Embedding method: spatial domain

Several watermarking schemes designed for the detection and recovery of image tampers have been proposed as shown in Table 2. In spatial domains, two watermarks are generated from the sender's side, i.e. detection bits and recovery bits. On the decoder side for the authentication phase [9, 28], the watermark component is extracted to check whether there are any malicious modifications or not. If there is a tampered image, the watermark can be used to determine the tampered part, most papers using spatial domains with block-neighbourhood based [8].

The important thing that needs to be underlined is the selection of watermark components for recovery bit will determine the success of the recovery process. In addition, the more watermarks are inserted, the more information about verification and recovery can be maintained. Therefore, more watermarks generally result in more accurate tamper detection and improved image recovery quality. However, the number of watermarks should be selected while still protecting the image from serious distortion. Therefore, we must choose simultaneously between improving the accuracy of tamper detection and the quality of the recovered image, while preserving the image quality that watermarks should be considered in future studies [9].

Sreenivas [38] uses a self-embedding watermarking scheme that improved the quality of the restored image. This scheme compares the use of the average intensity of blocks with watermarking bits in varying lengths. Kiatpapan [5] uses a dual watermarking approach to ensure better image tamper detection and recovery performance. In this case, setting the bit-plane watermark is also worth noting. In the proposed Kiatpapan's method, two sets of 8-bit watermark components are arranged in symmetric centre mode so that the image information with higher information values is evenly distributed. This method, therefore, can recover the image tamper perfectly even on the left, right, up or down of the original image.

In addition, the needs of authenticity inspection and correctness of the content of an image have developed fragile watermarking scheme [16, 17]. The method was developed by Wang et al. [21] by dividing the image into blocks of a certain size, and 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. [50] facilitates the ability to detect tamper hierarchically, where this process can locate areas damaged up to three levels. If it is not found in the first level, it will continue to the second and third level as illustrated in Fig. 9. However, in this scheme, it is not possible to recover the damaged block when the watermark that is inserted into another block is also damaged and there is no second chance to recover the block.

The problem of other opportunities for restoration has been proposed by [20, 21], for the recovery of the block was 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 Li et al. [19], which uses 3 bits of the LSB to store bit recovery. To improve the quality of watermarked images and restoration of damaged images, Qin et al. [24] use 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 is damaged; some additions 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 11 first quantisation coefficients. Huo et al. [25] propose a method that divides 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.

Table 2 Summary of some self-embedding fragile watermarking methods in the spatial domain

Watermarking approach	Watermark generation	Segmented block size (pixels)	Embedding technique	Watermark payload (bpp)	PSNR water-marked image (dB)	Attacks	PSNR of restored images (dB)
[2] Qin 2012	Authentication bit: Hash calculation and folding operation Recovery bit: non-sub sampled contourlet transform (NSCT)	Authentication bit: 8×8 Restoration bit:32×32	LSB Adaptive bit Allocation	1	51	Content-preserving manipulations.	41–48
[38] Sreenivas 2016	Authentication bit: Ex- OR operation Recovery bit: 7 different ways including block average intensity	2×2	LSB a chaotic map	3	37	Collage attack	26–38
[5] Kiatpapan 2015	Two identical down-sampled images embedded to upper and lower sections of LSB	4×4	LSB	2	NA	Collage attack	19–26
[1] Singh 2016	Authentication bit: longitudinal redundancy check (LRC) and the mean value of the pixels of the block Restoration bit: DCT and quantisation matrix	2×2	LSB	3	37–39	Object addition attacks, Object removal attacks, and Cropping	30–43
[40] Dhole 2015	Authentication bit: XOR operation Recovery bit: DCT	8×8	LSB	2	34–38	Vector quantisation (VQ), collage and quantisation attacks	36–40
[6] Zhang 2011	Authentication bit: Hash data Recovery n bit: average intensity	8×8	LSB	3	37	Content-tampering attack	22–40
[7]Cao2017	Authentication bit: Hash data Recovery bit: average intensity	2×2	LSB	2	44	VQ attack and collage attack. In	46
[3] Chang 2013	Authentication bit: LBP, Recovery bit: mean value of the pixels of the block	3×3	LSB	2	44	Collage attack VQ attack, constant–average attack	44–68
[13] Doyoddorj 2017	Authentication bit: OR operation Recovery bit: average intensity	NA	LSB	3	44–45	Malicious and incidental attacks	29–32
[16] He 2012	Authentication and Recovery bit: average intensity	2×2	LSB	2	44	Collage attack, content-preserving manipulations	27–32
[24] Qin 2017	Authentication bit: hash data Authentication and Recovery bit: average intensity	3×3	LSB overlapping embedding strategy	2	42–44	Collage attack	33–38
[19] Li 2011	Recovery bit: DCT, Quantisation	8×8	LSB	2	44	ES attack, collage attack and only-content-tampering attack	45–51
[44] Zhang 2011	Authentication bit: hash data Recovery bit: DCT	8×8	LSB	3	37	Tampered-block	30
[45] Saeed 2015	Authentication bit: MD5 hash algorithm Recovery bit: SPHIT	3×3	LSB	3	37	Malicious attacks	35
[14] Lin, 2006	Authentication bit: X-OR Operation Recovery bit: Average intensity of block	2×2	LSB	3	37	Cropped image	28–50
[46] Haghighi 2018	Authentication bit: X-OR operation Two recovery bit: lifting wavelet transform (LWT) and Stucki Kernel	2×2	LSB	2	46	Cropped image	44–45
[47] Qin 2018	Authentication bit: hash function	8×8	LSB	2	44	Cropped image	31

Table 2 Summary of some self-embedding fragile watermarking methods in the spatial domain *(Continued)*

Watermarking approach	Watermark generation	Segmented block size (pixels)	Embedding technique	Watermark payload (bpp)	PSNR water-marked image (dB)	Attacks	PSNR of restored images (dB)
	Recovery bit: optimal iterative block truncation coding						
[48] Qin 2016	Authentication bit: hash function Recovery Bit: Average intensity of block	2 × 2	LSB	2	44	Cropped image	46
[49] Wu 2017	Authentication bit: vector quantisation (VQ) Recovery bit: average intensity of block	4 × 4	LSB	1	51	Cropped image	40

From the description of some of the methods above, the problem of fragile watermarking methods lies in the ability of restoration, i.e. in the case of watermark insertion, tamper localisation and recovery. There is a trade-off between image watermark quality and insertion capacity. The success of the fragile watermarking method is seen from the accuracy of localisation tamper. By utilising more bits to detect the damaged area, the quality of the watermarked image is reduced. Another method in [8] proposes the use of restoration bits to locate damaged areas but does not need to be embedded in watermarked images. The next problem is the ability of self-recovery. This process will restore the damaged area by extracting the effective information by using more bits to represent the restoration bit [25]. From the description above, on the sender side, the selection of watermark components is 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 by using the restoration bit with the appropriate contents.

2.2.5 Embedding method: transform domain

Unlike spatial domain watermarking [6–9, 12, 14], the popular image compression standards are compatible with frequency domains, image transformations including DCT, discrete wavelet transform (DWT) and a combination of both DCT-DWT. In many of these schemes, the watermark is taken from the conversion process to the frequency domain. Then this frequency is inserted in the original image. The inverted Fourier transformation is then applied in the second phase to form a ready-to-send watermarked image to the receiving point. Another possible explanation is that the domain transformation method provides the possibility of insertion of more information and stronger resistance to many common attacks; a prominent weakness is the higher computational cost of spatial-domain watermarking techniques.

The self-embedding fragile watermarking on the spatial domain has a weakness when the attack was in the form of JPEG compression. Therefore, DCT domain was used by paper [51] because of its lower computational complexity and used in JPEG compression algorithm.

The DCT is one of the most popular frequency transformations in image and video processing due to its simplicity and high energy compaction [52]. For an image block x of size N x N, the type II 2-D DCT is defined as Eq. (12).

$$X(k_1,k_2)=\frac{2}{N}u(k_1)u(k_2)\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}x(n_1,n_2)\cos\frac{\pi(2n_1+1)k_1}{2N}\cos\frac{\pi(2n_2+1)k_2}{2N} \tag{11}$$

Where $k_1,k_2=0,1,2,\cdots,N-1, u(0)=\frac{1}{\sqrt{2}}$ and $u(k)=1$ *for* $n\neq 0$.

The 2-D IDCT can be expressed as

$$x(n_1,n_2)=\frac{2}{N}\sum_{n_1=0}^{N-1}\sum_{n_2=0}^{N-1}u(k_1)u(k_2)x(n_1,n_2)\cos\frac{\pi(2n_1+1)k_1}{2N}\cos\frac{\pi(2n_2+1)k_2}{2N} \tag{12}$$

Each DCT coefficient represents a certain spatial frequency. Those patterns are sometimes referred to as basis functions. Figure 10 shows the spatial frequency patterns of the 8 × 8 DCT which is the most common block size used in image processing and the size we will be using throughout this thesis. In the DCT domain, an image block is represented as a combination of these basis functions with different magnitudes and/or signs. To put it in different words, if the image blocks are thought of as visual words, then the basic functions of the DCT are the visual alphabet that composes these words.

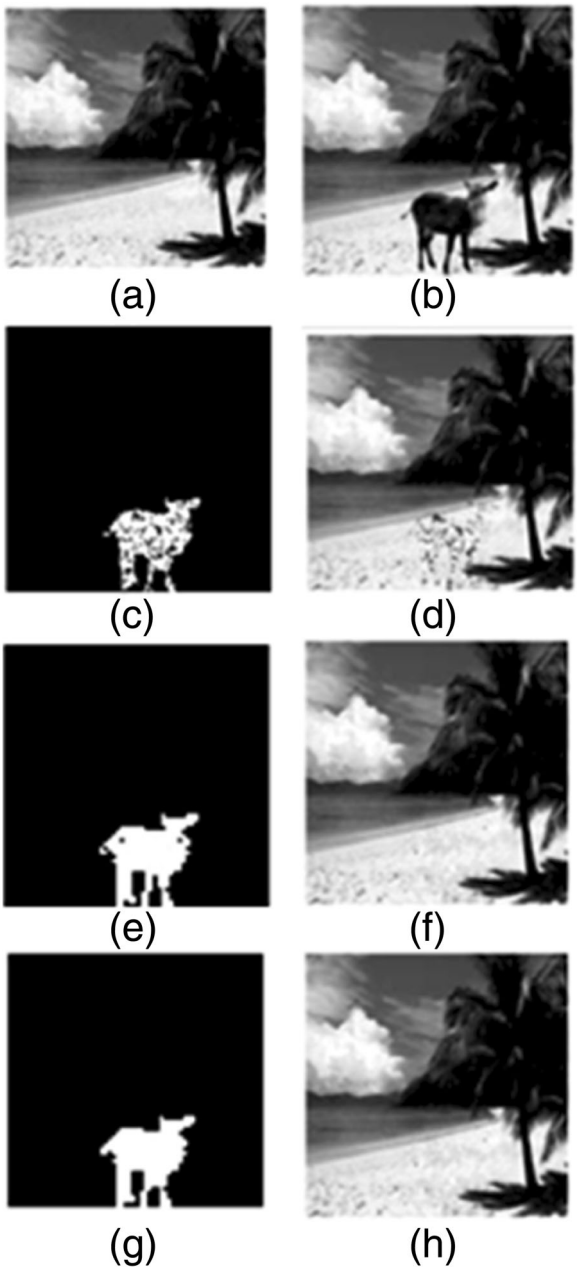

Fig. 9 Example of detecting tamper hierarchically [50]. **a** The watermarked beach. **b** Tampered beach. **c**, **e**, **g** Detected erroneous regions of level 1, level 2 and level 3 respectively. **d**, **f**, **h** Recovered beaches of level 1, level 2 and level 3 respectively

The first DCT coefficient $X(0, 0)$ is the DC coefficient. The DC coefficient has zero frequency in both the vertical and horizontal directions and it indicates the brightness of the image block since it corresponds to the average of the pixel values in the block. The remaining coefficients are called the DC coefficients. The AC coefficients closer to the DC coefficient have lower spatial frequencies and the frequencies increase as we move away from the DC coefficient in all directions. AC coefficients respond to grey level changes that are in the same direction as their spatial frequencies; for example, the coefficients in the first column respond to grey level changes in the vertical direction (horizontal edges) since their spatial frequencies are vertical.

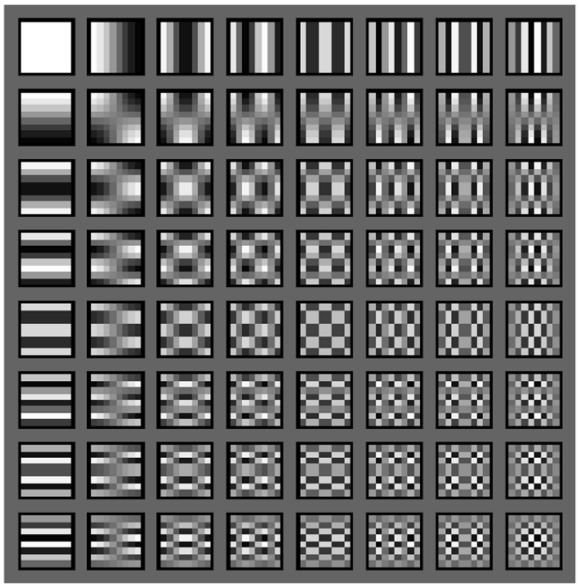

Fig. 10 Spatial frequency components of the 8 × 8 DCT [52]

The use of wavelet transform will mainly address the capacity and robustness of the Information-Hiding system features. The Haar wavelet transform is the simplest of all wavelet transforms. In this, the low-frequency wavelet coefficients are generated by averaging the two-pixel values and high-frequency coefficients that are generated by taking half of the difference of the same two pixels. The four bands obtained are LL, LH, HL and HH which are shown in Fig. 11. The wavelet transform decomposes the image into four sub-bands of different frequencies, namely approximation image (LL_k), horizontal (HL_k), vertical (LH_k) and diagonal (HH_k) details where k denotes the decomposition level.

In watermarking applications, lower decomposition levels are more vulnerable to image alteration as they have a lower proportion of energy as compared to higher decomposition levels. This energy is defined as in Eq. (13).

$$E_k = \frac{1}{N_k M_k} \sum_i \sum_j |I_k(i, j)| \tag{13}$$

Where k is the decomposition level, I_k denotes coefficients of the corresponding sub-band and N_k and M_k are sub-band dimensions. By comparing the energy of the sub-bands in the same level, i.e. as shown in Fig. 11

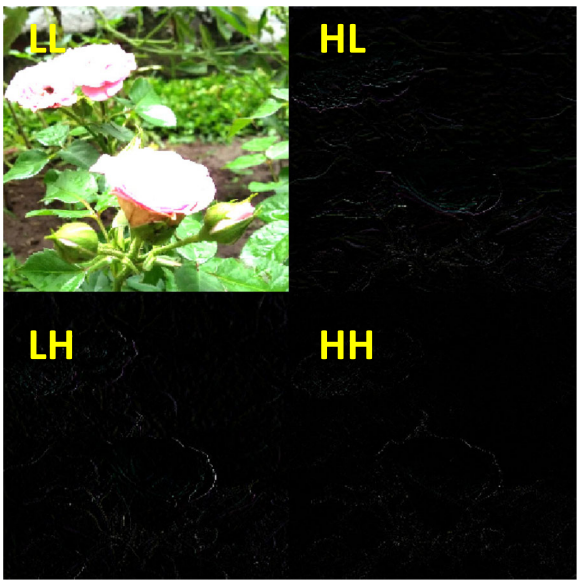

Fig. 11 Wavelet decompositions of an image

(LL_1, HL_1, LH_1 and HH_1), it can be seen that the energy accumulation in the horizontal detail (HL_k) is significantly more than those of the vertical and diagonal details, hence suggesting that this sub-band is more robust to image modification. In other words, even though the approximation image (LL_k) has the highest portion of the energy of the original image, embedding the watermark in this part will degrade the image quality. Therefore, the horizontal sub-band in each level can be chosen as the best candidate area for embedding the watermark, in order to achieve image quality preservation and offer robustness simultaneously.

In the reviews cited by [53], DWT-based methods can show more accurate visualisation results when compared to the DCT method. It is also mentioned that if we can use areas that are less sensitive to the human visual system, it will provide more opportunities to embed watermarks without degrading the quality of terracotta images and this method is the most powerful against noise. In addition, compared to DCT-based transformations, wavelet produces fewer visual artefacts because this technique does not require that images be decomposed into blocks. However, there are weaknesses in this method: (1) oscillation: one of the complexities of wavelet-based processing is the oscillation of wavelet coefficients around the singularity. This is because wavelet is a band-pass function. (2) Shift variance: another factor, which complicates the processing of wavelet, is that a small shift in the signal greatly disturbs the oscillation pattern of the wavelet coefficients around the singularity. (3) aliasing: whenever wavelet and scale coefficients change, inverted DWT cannot cancel aliasing and this causes artefacts in the reconstructed signal. (4) Lack of directionality: lack of directional selectivity greatly complicates the modelling and processing of geometric image features such as mountains and edges. This deficiency has been solved by using a dual-tree complex wavelet transform (DTCWT) [53].

2.3 Comparison and discussion

Comparison of different existing self-embedding fragile watermarking techniques in the spatial domain is presented in Table 2. Meanwhile, Table 3 shows the comparison of self-embedding fragile watermarking techniques in the frequency domain. Overall, from these tables, it can be observed that most of the fragile watermarking techniques are in a spatial domain, using LSB embedding technique.

Based on the previous sections—summarised in Tables 2 and 3—eight criteria can be categorised to compare each watermarking approach based on the embedding domain. Most of the methods differentiate in two major watermark components: authentication/check bits and recovery bits. For tamper recovery, the watermark generation which is selected using the frequency domain [2, 42, 56] has better results than the watermark generation using spatial domains [5, 14, 16], i.e. average intensity of blocks. This can be seen from the value of PSNR results after the recovery process. Estimates of fragile watermarking-based techniques are designed into the eight criteria with the following explanation: (1) watermarking approach: this term represents the technique used for image authentication

Table 3 Summarisation of some self-embedding fragile watermarking methods in the transform domain

Watermarking approach	Watermark generation	Segmented block size (pixels)	Embedding technique	Watermark payload (bpp)	PSNR watermarked image (dB)	Attacks	PSNR of restored images (dB)
[41]Han 2013	Dual watermarking scheme: Fragile: LSB, robust: DWT	2 × 2	LSB and DWT	2	N. A	General attack	N. A
[4] Chamlawi 2009	Recovery: image digest using the Huffman coding on IDCT coefficients	N. A	IDCT and IWT	N. A	38–44	Malicious and incidental attacks	NA
[54] Zhang 2018	Mean value of each overlapping	2X2	DWT	1–2	22–24	Cropping attacks	27–36
[55] Xin 2016	NA	8 × 8	DWT	N.A	44 dB	Malicious attacks	30 dB

Fig. 12 Three standard test images. **a** Lena, **b** Airplane, **c** Baboon

and recovery using fragile watermarking; (2) watermark generation: this term represents the kind of watermark generation used in each image authentication and recovery scheme; (3) segmented block size: this term indicates the block size (in $n \times n$ pixels) of each non-overlapping or overlapping block division in scheme addresses; (4) embedding technique: this term represents the kind of embedding techniques used in each scheme; (5) watermark payload: this term is related to the number of hidden information embedded; (6) PSNR watermarked image: this term represents the visual quality of the watermarked image in each scheme. PSNR larger than 40 dB is considered to be the threshold of a very good visual quality of the reconstructed image. While the PSNR lower than

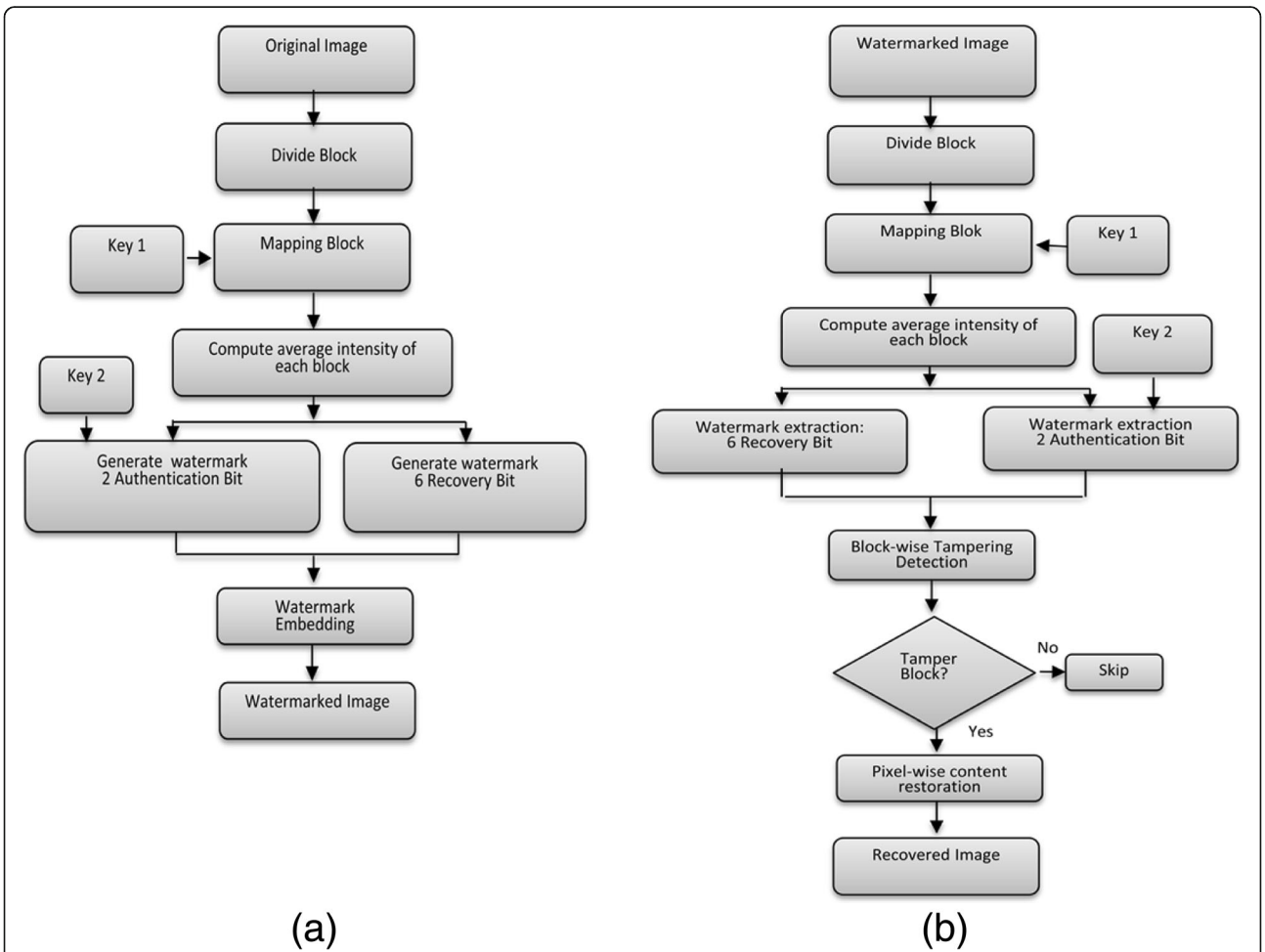

Fig. 13 Block diagram of (**a**) watermark embedding and (**b**) extraction process for Wang et al.'s method [24]

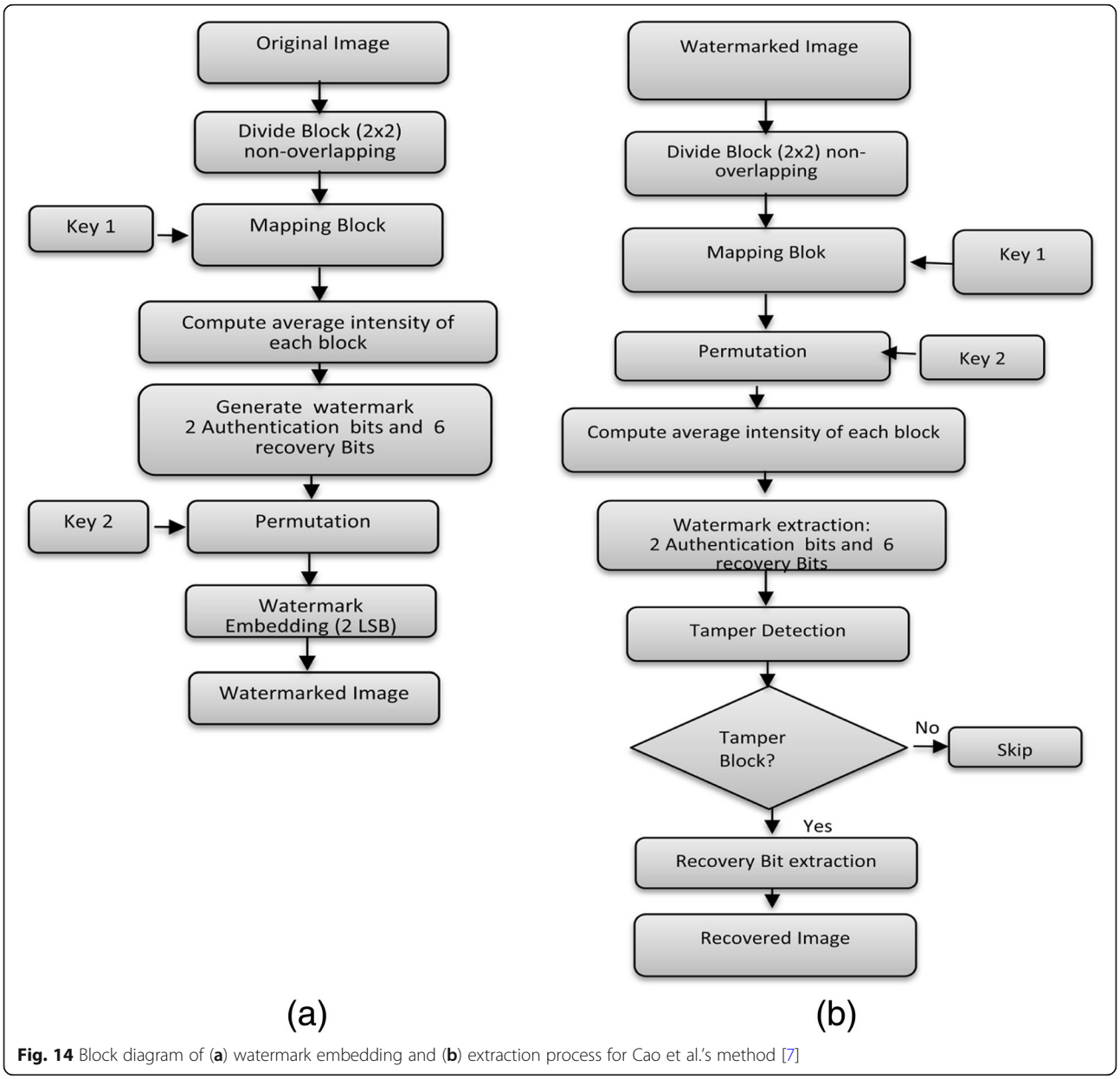

Fig. 14 Block diagram of (**a**) watermark embedding and (**b**) extraction process for Cao et al.'s method [7]

30 dB would be considered as inadequate image quality. (7) Attacks: this term represents the kind of attacks used in each image authentication and recovery scheme. There are various possible malicious or accidental attacks that may be imposed on the watermark. The availability of a wide range of image processing allows for attacks on watermarking system resilience. From Tables 2 and 3, most of the attacks include the general tampering (cropping) and the collage attack. (8) PSNR of restored image: this term represents the visual quality of the restored image in each scheme.

A comparative explanation of the advantages and disadvantages of each method is as follows. In a paper [2], Qin et al. have proposed a self-embedding watermarking scheme that is adaptive to the number of insertion bits, where for more complex blocks they are inserted more bits than the smooth blocks. Besides, this method uses one LSB similar to Zhang [6], thus the watermarked image is better. On the contrary, the ability to restore is reduced and unable to withstand content-based attacks. Similarly, the Sreenivas's method [38] proposes a payload watermark for different restorations depending on the level of similarity to the average value for blocks of 2×2 pixels. The lack of this method requires a long time because finding similarities with seven conditions and the detection process requires additional methods.

Kiatpapan [5] has proposed a simple method with dual watermarking. A limitation of the proposed method was

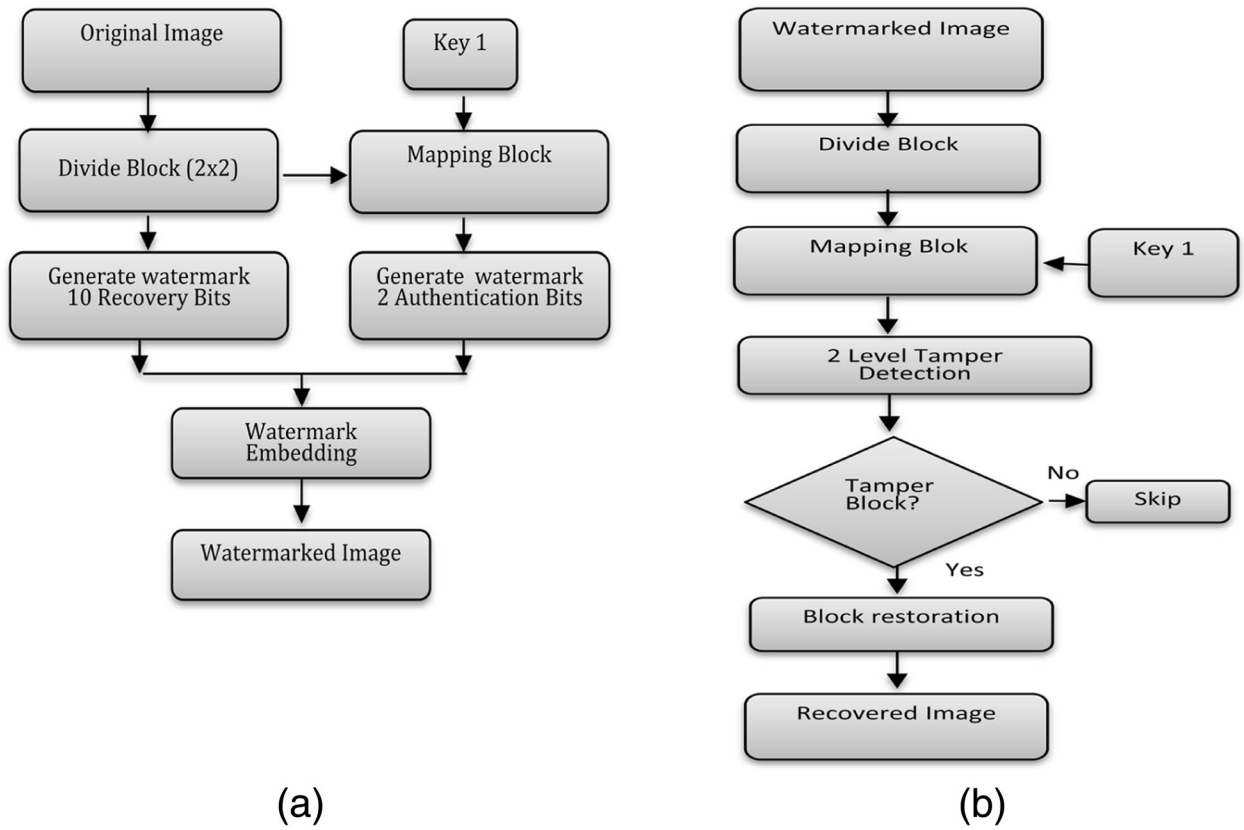

Fig. 15 Block diagram of (**a**) watermark embedding and (**b**) extraction process for Singh et al.'s method [1]

that it will not be able to recover image tamper if the tampering takes place globally across the entire original image. Meanwhile, Singh's method [1] and Dholes' method [42] were used as block-wise mechanisms with different authentication bit generation. Singh has used two level tampered detection procedures, localisation and recovery with the high probability. Dholes' was used in two phases, one with own image and another with watermarked image. The drawback of both approaches is not resistant to common image processing operations such as JPEG compression, contrast enhancement and filtering.

The proposed scheme in [7] uses a hierarchical reference-bits capacity according to the importance of image contents. They can improve the visual quality recovered results, especially for larger tampering rates, but the generation of reference bits which used reference sharing mechanism [6] depends on the proportion of different detection results. In addition, for larger tampering rates, the method proposed by Qin [24, 47, 48] produces good recovery results using the overlapping block-wise mechanism. The method proposed by [3, 13] uses the block-mapping operation and adopts parity

Fig. 17 Comparison results of Lena watermarked image of four schemes [1, 7, 24, 48]. **a** Original image. **b** PSNR = 37.15 dB [1]. **c** PSNR = 44.15 dB [7]. **d** PSNR = 44.27 dB [24]. **e** PSNR = 44.16 [48]

check and the intensity-relations check to thwart various malicious attacks. However, unlike He's method [16], it cannot be used on full frames of an image as it would destroy the intelligibility of the content.

Li [19] has proposed a dual-redundant ring structure and two tamper detection schemes, which improve the quality of the restored image. However, it cannot withstand mild distortions such as random noise and JPEG compression; another disadvantage is the possibility of detection errors because the blocks used are too large, so that if one sub-block is damaged then the entire block will be marked as an error. In method [56], all extracted watermark data contribute to the recovery of content, and the accuracy of the restoration coefficient depends on the amount of available watermark data. Furthermore, the paper [44] aims to show that having a known fault location, image damage can be modelled and handled as a deletion error. But reference data can be destroyed by some image processing operations, such as filtering, compressing and additional noise interference, so content recovery is disabled. A different method was proposed recently by Wu [49] where an image authentication scheme was proposed to exploit QR codes to protect the integrity of significant VQ authentication data. This method only focuses on the authentication process while the recovery process requires repetition until all blocks undergoing changes are reconstructed.

Grouping is different in Table 2 using the frequency domain. Method [4, 41, 49, 54] take advantage of robust watermarking for generating recovery bits inserted into the LL sub band on the DWT. The main drawback of this method is that computing is quite high and the watermark insertion capacity is low.

In most of the watermarking schemes mentioned above, data that represent the main content in an area are always poured into other areas for content recovery on the recipient's side [23]. When certain regions containing original information are damaged or lost, however, it is not possible to recover original content in the previous area. In other words, content recovery has failed. For all uses, the fragile watermarking method has a disadvantage where watermarks can be destroyed by image processing operations such as contrast enhancement, JPEG compression, filtering and so on. Therefore, in the future, semi-fragile watermarking methods can be developed that are strong against these attacks and have good storage capabilities [31] (Fig. 12).

2.4 Experimental evaluation

In this section, to clarify the process of watermark insertion, authentication and recovery, we conduct an experimental evaluation of four fragile watermarking techniques: Singh et al.'s [1], Cao et al.'s [7], Qin et al.'s method [24] and Qin et al.'s method [48], which used LSB method. A large number of test images sized 512 × 512 are used in our experiments to demonstrate the effectiveness of the schemes. We use three standard test images: Airplane, Lena and Baboon as shown in Fig. 12. Lena is considered as low-textured, Airplane as medium

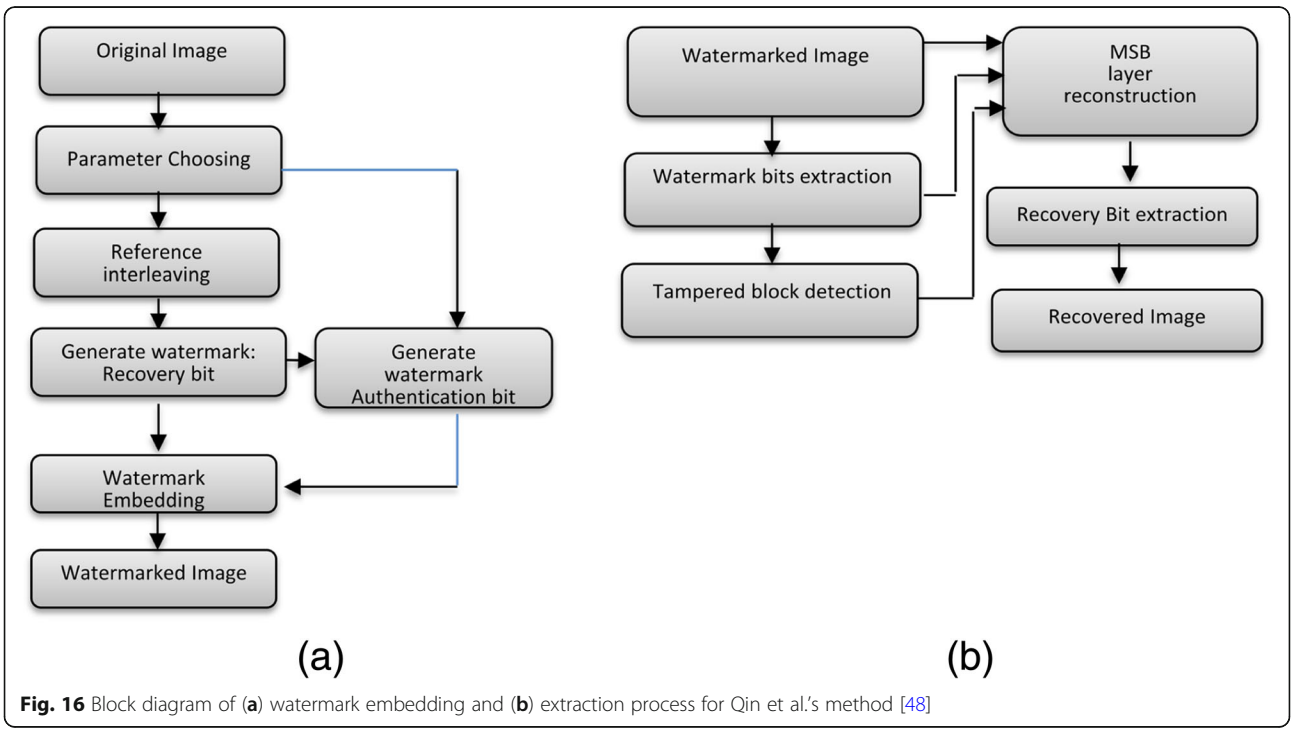

Fig. 16 Block diagram of (**a**) watermark embedding and (**b**) extraction process for Qin et al.'s method [48]

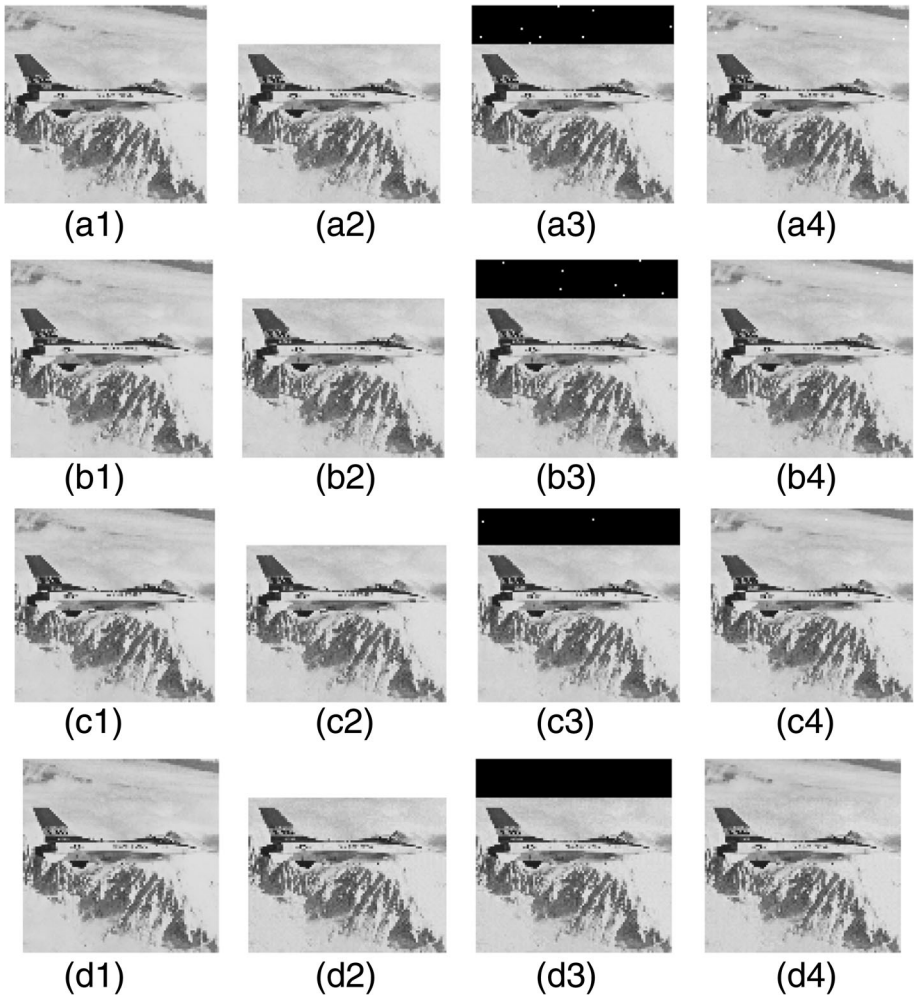

Fig. 18 Example of the 20% tampered Airplane image with cropped image of four schemes. (a1–a4) Scheme [1]; (b1–b4) scheme [7]; (c1–c4) scheme [24]; (d1–d4) scheme [48]. First column: Airplane original images, second column: tampered images, third column: tamper detection, fourth column: tamper recovery

textured and Baboon as highly textured ones. The simple block diagrams of the embedding and extraction processes for the four schemes are shown in Figs. 13, 14, 15 and 16.

From the block diagram for all of schemes, it can be underlined that there are three main phases that can be explained as follows: (1) watermark generation and insertion: in this phase, the results of imaging are generated from the host image. Furthermore, the authentication bits of each block are obtained from the slightly digested image. First, to improve safety, the permutation coefficient is used [7, 24]. Then, all the coefficient bits embedded in each block are shuffled and encrypted using a specific key [1, 7, 24, 48]. Finally, the digest image and authentication bits are merged and embedded in the host image with the LSB technique. (2) Detection and authentication of tamper: after receiving a watermarked image, for authentication, first of all, the digested image and authentication bit are extracted. Then, based on the extracted data, the image is analysed to check the authenticity of manipulation and destruction. If the image contains damaged parts, the recovery phase tries to recover the original content. (3) Recovery and reconstruction: in the final stage, based on the mapped blocks, invalid blocking (damaged areas) is restored using image imagery embedded in the host.

Furthermore, all the proposed fragile watermarking methods use non-overlapping image block sized 2×2, except the scheme proposed by Qin et al. [24], which uses overlapping image block sized 3×3. The use of smaller blocks allows better detection and recovery processes. A larger block size will result in a more sensitive damage detection. For example, even if only one pixel in a 3×3 pixel block is damaged, then the entire block will be detected as a broken block, thus the recovery process is carried out in all areas of damage, and will be replaced with other values as a whole.

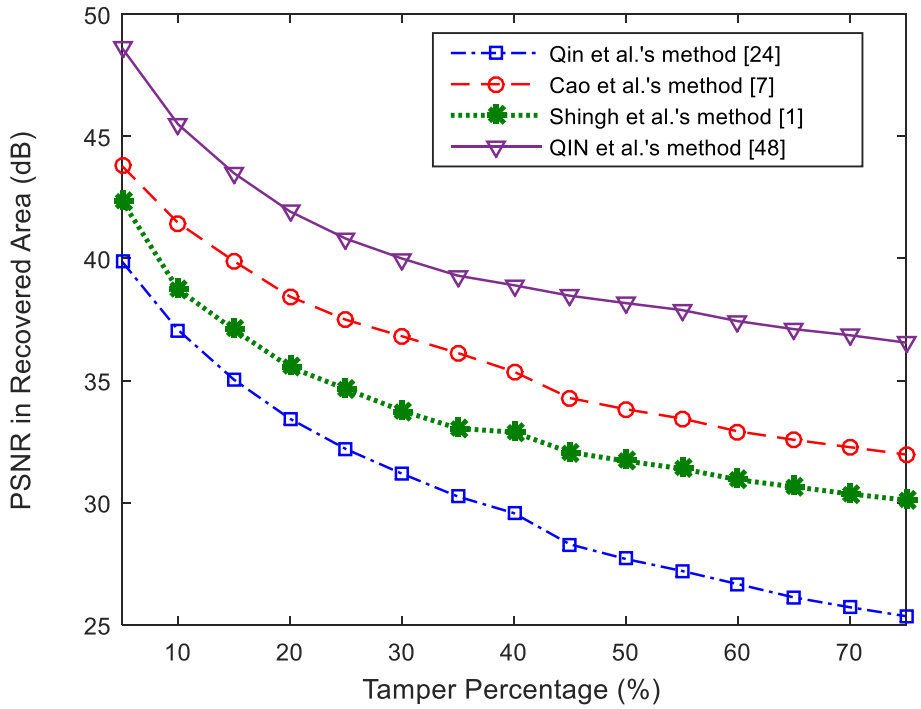

Fig. 19 Performance comparison on average PSNR of three images versus tamper cropping attacks percentage

The block diagrams show a slight difference in the watermark embedding process, wherein Singh uses three LSB as the watermark insertion point, here it can be seen with three LSB that the additional information store capacity is 12 bits, 8 bit for recovery and 4 bit for authentication, while the other methods use two LSB, so it can only hold 8 bits, 6 bit for recovery and 2 bit for authentication. The advantage of using less LSB is in the result of the watermarked image that does not show any significant change, so the value of PSNR is also high, it can be seen in Fig. 17.

The PSNR value of the watermarked image decreases with the increase of watermark payload. As shown in Fig. 17, the PSNR of the Singh et al.'s watermarked image was below the others because there were differences in the payload, where the watermark payload of Singh used 2 bpp and others used 3 bpp.

The performance of the four methods may vary by type of distortion. In this paper, we compare those using cropping attacks with various error percentages. The cropped attack is done by removing some portions of the image into another size that the attacker wants. To describe this, Fig. 18 illustrates the example of localisation of destruction for 20% of tamper in the Airplane image and their recovered image. The rows in Fig. 18 show an example of the tamper detection and recovery process of four methods, which is the first to fourth column as original images, tampered images, tamper detection and tamper recovery, respectively.

To compare the success of both methods in terms of image recovery, PSNR measurements were performed for three images in average using Eq. (1). PSNR is measured by comparison of recovered image to watermarked image. The obtained results for the four schemes can be seen in Fig. 19. The PSNR result of Qin et al.'s method [48] reaches the highest value, followed by method [1, 7, 24], because the insertion scheme proposed by Qin et al. [48] can be categorised into overlapping and non-overlapping, which are related to variable numbers from the MSB layer and the LSB layer used during watermark insertion. Based on the interleaving reference mechanism, MSB bits representing the contents of the block principle are inserted to produce reference bits, and then, are entered into LSBs. Because both the number of MSB layers and the LSB layer used affect the quality of the watermark image, affect the likelihood of complete recovery and the restored image quality, detailed

Table 4 Comparisons of tamper detection performance under cropped attack

Schemes	Lena	Airplane	Baboon
Singh et al. [1]	0.00406	0.00436	0.00374
Cao et al. [7]	0.00314	0.00326	0.00339
Qin et al. [24]	0.04450	0.04425	0.04319
Qin et al. [48]	0	0	0

analysis is given to provide the theoretical value and present optimal choice of embedding mode.

To further test the precision of tamper detection, PFD written in Eq. (10) are employed in this paper [36]. Table 4 shows the results under cropped attack, which is similar to example illustration as shown in Fig. 17. PFD is used to evaluate the tamper detection performance. The lower the PFD, the better the tamper detection performance is. Ideally, PFD of the watermarking scheme for authentication would be close to zero. It is seen that PFD of Qin et al.'s method is 0, in which all tampered blocks have been localised. Qin et al.'s results show the superiority compared with some state-of-the-art schemes, because the scheme utilises flexible numbers of the MSB layers to generate the interleaved reference bits for content recovery. They compare the recalculated authentication bits from the extracted authentication bits—if different, the block is judged as having been tampered. Otherwise, the block is marked as intact. Because the watermark concealment process for the detection process is inserted in the original block, it is possible that the comparison of block per block in the affected image with a watermarked image will be easily detected. So that for cropping attacks can be detected properly according to the location of the damage that has been raised.

3 Conclusion

Comparative characteristics of the self-embedding fragile watermarking scheme have been described. There are two watermarks which are used for authentication with recovery capability: authentication bit and restoration bit. Authentication bit is for tamper detection and localisation, while restoration bit is for tamper recovery in decoder side. In the spatial domain, the watermark is selected from the image feature itself which is then inserted in LSB bit-plane by first emptying the bits at position one LSB, two LSB and three LSB, where the bits will be used for detection in case of damage and it can be extracted to replace the tampered bit. Therefore, the watermarked image quality depends entirely on the amount of LSB replaced by the watermark in pixels, as shown in the table by increasing the bpp value, thus decreasing the PSNR value. The PSNR value of the watermarked image in the spatial domain is higher than the frequency domain because the direct spatial domain process in the pixel region of the image is transformed. In addition, from the experiment result of four methods which uses spatial domain and is generated in three different standard test images, it can be shown that the value of PFD of general tamper attack is better using average intensity of the block.

Abbreviations

bpp: Bit-per-pixel; DCT: Discrete cosine transform; DTCWT: Dual tree complex wavelets transform; DWT: Discrete wavelet transform; LBP: Local binary pattern; LRC: Longitudinal redundancy check; LSB: Least significant bit; MSB: Most significant bit; PFD: Probability of false detection; PFR: Probability of false rejection; PSNR: Peak signal-to-noise ratio

Acknowledgements

Not applicable.

Funding

This research was supported by Institut Teknologi Sepuluh Nopember (ITS) through International Publication Acceleration Program (P3I) Batch II of 2018 Grant.

Availability of data and materials

Please contact author for data requests.

Authors' contributions

The first author mainly developed and evaluated the theory analysis, experimental evaluation and wrote this manuscript. The second author managed this research project by continually advising the first author and reviewed this manuscript. The third author supported by creating a framework for thinking and editing this manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Received: 21 May 2018 Accepted: 22 April 2019
Published online: 08 May 2019

References

1. D. Singh, S.K. Singh, Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability. J. Vis. Commun. Image Represent. **38**, 775–789 (2016). https://doi.org/10.1016/j.jvcir.2016.04.023
2. C. Qin, C. Chang, P. Chen, Self-embedding fragile watermarking with restoration capability based on adaptive bit allocation mechanism. Signal Process. **92**, 1137–1150 (2012). https://doi.org/10.1016/j.sigpro.2011.11.013
3. J. Chang, B. Chen, C. Tsai, *"LBP-based fragile watermarking scheme for image tamper detection and recovery," 2013 International Symposium on Next-Generation Electronics* (Kaohsiung, 2013), pp. 173–176. https://doi.org/10.1109/ISNE.2013.6512330
4. R. Chamlawi, I. Usman, A. Khan, *"Dual watermarking method for secure image authentication and recovery," 2009 IEEE 13th International Multitopic Conference* (Islamabad, 2009), pp. 1–4. https://doi.org/10.1109/INMIC.2009.5383118
5. S. Kiatpapan, T. Kondo, in *2015 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. An image tamper detection and recovery method based on self-embedding dual watermarking (Hua Hin, 2015), pp. 1–6. https://doi.org/10.1109/ECTICon.2015.7206973
6. X. Zhang, S. Wang, Z. Qian, G. Feng, Reference sharing mechanism for watermark self-embedding. IEEE Trans. Image Process. **20**(2), 485–495 (2011). https://doi.org/10.1109/TIP.2010.2066981
7. F. Cao, B. An, J. Wang, D. Ye, H. Wang, Hierarchical recovery for tampered images based on watermark self-embedding correspondence. Displays. **46**, 52–60 (2017). https://doi.org/10.1016/j.displa.2017.01.001
8. V.C. Shruthy, S. Varghese, An efficient self-embedding watermarking scheme for colour image tamper detection and recovery. Int. J. Comput. Sci. Mob. Comput. **4**, 383–390 (2015)
9. T. Chen, H. Lu, in *2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)*. Robust spatial LSB watermarking of color images against JPEG compression (Nanjing, 2012), pp. 872–875. https://doi.org/10.1109/ICACI.2012.6463294
10. N. Wang, C. Kim, in *2009 9th International Symposium on Communications and Information Technology*. Tamper detection and self-recovery algorithm of color image based on robust embedding of dual visual watermarks using DWT-SVD (Icheon, 2009), pp. 157–162. https://doi.org/10.1109/ISCIT.2009.5341268

11. X. Tong, Y. Liu, M. Zhang, Y. Chen, A novel chaos-based fragile watermarking for image tampering detection and self-recovery. Signal Process. Image Commun. **28**(2), 301–308 (2013)
12. M. Yu, J. Wang, G. Jiang, Z. Peng, F. Shao, T. Luo, New fragile watermarking method for stereo image authentication with localization and recovery. AEU Int. J. Electron. Commun. **69**(1), 361–370 (2015). https://doi.org/10.1016/j.aeue.2014.10.006
13. M. Doyoddorj, K.H. Rhee, in *Multidisciplinary research and practice for information systems*. CD-ARES 2012. Lecture notes in computer science, ed. by G. Quirchmayr, J. Basl, I. You, L. Xu, E. Weippl. Design and analysis of a fragile watermarking scheme based on block-mapping, vol 7465 (Springer, Berlin, 2012). https://doi.org/10.1007/978-3-642-32498-7_49
14. S.D. Lin, Y.-C. Kuo, Y.-H. Huang, in *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06)*. An image watermarking scheme with tamper detection and recovery (Beijing, 2006), pp. 74–77. https://doi.org/10.1109/ICICIC.2006.414
15. Y. Huo, H. He, F. Chen, Alterable-capacity fragile watermarking scheme with restoration capability. Opt. Commun. **285**(7), 1759–1766 (2012). https://doi.org/10.1016/j.optcom.2011.12.044
16. H. He, F. Chen, H. Tai, S. Member, T. Kalker, J. Zhang, Performance analysis of a block-neighborhood-based self-recovery fragile watermarking scheme. Ieee Trans. Inf. Forensics Secur. **7**(1), 185–196 (2012)
17. X. Zhang, S. Wang, Fragile watermarking with error free restoration capability. IEEE Trans. Multimed. **10**(8), 1490–1499 (2008)
18. H. Zhang, C. Wang, X. Zhou, Fragile watermarking for image authentication using the characteristic of SVD. Algorithms. **10**(1), 1–12 (2017)
19. C. Li, Y. Wang, B. Ma, Z. Zhang, A novel self-recovery fragile watermarking scheme based on dual-redundant-ring structure q. Comput. Electr. Eng. **37**(6), 927–940 (2011)
20. S. Bravo-Solorio, A.K. Nandi, Secure fragile watermarking method for image authentication with improved tampering localisation and self-recovery capabilities. Sign. Proces **91**(4), 728–739 (2011)
21. W. Wang, A. Men, B. Yang, in *2010 2nd IEEE InternationalConference on Network Infrastructure and Digital Content*. A feature-based semi-fragile watermarking scheme in DWT domain (Beijing, 2010), pp. 768–772. https://doi.org/10.1109/ICNIDC.2010.5657886
22. R. Eswaraiah, E.S. Reddy, in *2014 Seventh International Conference on Contemporary Computing (IC3)*. ROI-based fragile medical image watermarking technique for tamper detection and recovery using variance (Noida, 2014), pp. 553–558. https://doi.org/10.1109/IC3.2014.6897233
23. P.-L. Lin, P.-W. Huang, A.-W. Peng, in *IEEE Sixth International Symposium on Multimedia Software Engineering*. A fragile watermarking scheme for image authentication with localization and recovery (Miami, 2004), pp. 146–153. https://doi.org/10.1109/MMSE.2004.9
24. C. Qin, P. Ji, X. Zhang, J. Dong, J. Wang, Fragile image watermarking with pixel-wise recovery based on overlapping embedding strategy. Signal Process. **138**, 280–293 (2017)
25. Y. Huo, H. He, F. Chen, in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. Semi-fragile watermarking scheme with discriminating general tampering from collage attack (Kaohsiung, 2013), pp. 1–6. https://doi.org/10.1109/APSIPA.2013.6694260
26. X. Qi, X. Xin, A singular-value-based semi-fragile watermarking scheme for image content authentication with tamper localization. J. Vis. Commun. Image Represent. **30**, 312–327 (2015)
27. Y. Li, L. Du, in *Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. Semi-fragile watermarking for image tamper localization and self-recovery (Wuhan, 2014), pp. 328–333. https://doi.org/10.1109/SPAC.2014.6982711
28. O. Ekici, B. Sankur, B. Coşkun, U. Naci, M. Akcay, Comparative evaluation of semifragile watermarking algorithms. J. Electron. Imaging. **13**(1), 209 (2004)
29. X. Yu, C. Wang, X. Zhou, Review on semi-fragile watermarking algorithms for content authentication of digital images. Future Internet. **9**(4), 56 (2017)
30. B. Cheng, R. Ni, Y. Zhao, in *2012 IEEE 11th International Conference on Signal Processing*. A refining localization watermarking for image tamper detection and recovery (Beijing, 2012), pp. 984–988. https://doi.org/10.1109/ICoSP.2012.6491744
31. A. Khan, A. Siddiqa, S. Munib, S.A. Malik, A recent survey of reversible watermarking techniques. Inf. Sci. **279**, 251–271 (2014). https://doi.org/10.1016/j.ins.2014.03.118
32. Y. Shi, X. Li, X. Zhang, H. Wu, B. Ma, Reversible data hiding: advances in the past two decades. IEEE Access. **4**, 3210–3237 (2016). https://doi.org/10.1109/ACCESS.2016.2573308
33. X.-L. Liu, C.-C. Lin, C.-C. Chang, S.-M. Yuan, A survey of fragile watermarking-based image authentication techniques. J. Inform. Hiding Multimedia Signal Process. **7**(6), 1282–1292 (2016)
34. C. K. R, N. Shivananda, in *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. A new fragile watermarking approach for tamper detection and recovery of document images (New Delhi, 2014), pp. 1494–1498. https://doi.org/10.1109/ICACCI.2014.6968624
35. C.-M. Wu, Y.-S. Shih, A simple image tamper detection and recovery based on fragile watermark with one parity section and two restoration sections. Opt. Photonics J. **3**(2), 103–107 (2013)
36. D. Xiao, F.Y. Shih, An improved hierarchical fragile watermarking scheme using chaotic sequence sorting and subblock post-processing. Opt. Commun. **285**(10–11), 2596–2606 (2012). https://doi.org/10.1016/j.optcom.2012.02.002
37. O. Hemida, Y. Huo, F. Chen, H. He, in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. Block-DCT based alterable-coding restorable fragile watermarking scheme with superior localization (Kuala Lumpur, 2017), pp. 846–851. https://doi.org/10.1109/APSIPA.2017.8282151
38. K. Sreenivas, V.K. Prasad, in *2016 International Conference on Information Communication and Embedded Systems (ICICES)*. Improved block encoding method for an image self-recovery approach (Chennai, 2016), pp. 1–5. https://doi.org/10.1109/ICICES.2016.7518879
39. W. Susilo, J. Tonien, A QR code watermarking approach based on the DWT-DCT technique. Lect. Notes Comput. Sci (2017). https://doi.org/10.1007/978-3-319-59870-3_18
40. A. Kunhu, H. Al-Ahmad, S.A. Mansoori, in *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. A reversible watermarking scheme for ownership protection and authentication of medical Images (Ras Al Khaimah, 2017), pp. 1–4. https://doi.org/10.1109/ICECTA.2017.8251971
41. Q. Han, L. Han, E. Wang, J. Yang, in *2013 Ninth international conference on intelligent information hiding and multimedia signal processing*. Dual watermarking for image tamper detection and self-recovery (Beijing, 2013), pp. 33–36. https://doi.org/10.1109/IIH-MSP.2013.17
42. V.S. Dhole, N.N. Patil, in *2015 International conference on computing communication control and automation*. Self embedding fragile watermarking for image tampering detection and image recovery using self recovery blocks (Pune, 2015), pp. 752–757. https://doi.org/10.1109/ICCUBEA.2015.150
43. T.-Y. Chen, M.-S. Hwang, J.-K. Jan, A secure image authentication scheme for tamper detection and recovery. Imaging Sci. J. **60**(4), 219–233 (2012). https://doi.org/10.1179/1743131X11Y.0000000018
44. X. Zhang, Z. Qian, Y. Ren, G. Feng, Watermarking with flexible self-recovery quality based on compressive sensing and compositive reconstruction. IEEE Trans. Inf. Forensics Secur. **6**, 1223–1232 (2011)
45. S. Sarreshtedari, M.A. Akhaee, A. Abbasfar, in *2015 23rd European Signal Processing Conference (EUSIPCO)*. Digital image self-recovery using unequal error protection (Nice, 2015), pp. 71–75. https://doi.org/10.1109/EUSIPCO.2015.7362347
46. B.B. Haghighi, A.H. Taherinia, A. Harati, TRLH: Fragile and blind dual watermarking for image tamper detection and self-recovery based on lifting wavelet transform and halftoning technique. J. Vis. Commun. Image Represent. **50**, 49–64 (2018)
47. C. Qin, P. Ji, C. Chang, J. Dong, X. Sun, Non-uniform watermark sharing based on optimal iterative BTC for image tampering recovery. IEEE MultiMedia. **25**(3), 36-48 (2018).
48. C. Qin, H. Wang, X. Zhang, X. Sun, Self-embedding fragile watermarking based on reference-data interleaving and adaptive selection of emb e dding mode. J.Inf. Sci. **373**, 233–250 (2016)
49. W.C. Wu, Quantization-based image authentication scheme using QR error correction. J Image Video Proc. **2017**(13) (2017). https://doi.org/10.1186/s13640-017-0163-8
50. P.L. Lin, C.K. Hsieh, P.W. Huang, A hierarchical digital watermarking method for image tamper detection and recovery. Pattern Recogn. **38**(12), 2519–2529 (2005)
51. J.C. Patra, J.E. Phua, C. Bornand, A novel DCT domain CRT-based watermarking scheme for image authentication surviving JPEG compression. Digit. Signal Process. **20**(6), 1597–1611 (2010)

52. M. Hamid, *DCT-based image feature extraction and its application in image self-recovery and image watermarking* (Thesis Concordia Univesity, Canada, 2016)
53. S.S.M. Mousavi, A. Naghsh, Watermarking techniques used in medical images: a survey. J. of Dig. Imag. **27**(6), 714–729 (2014)
54. R. Zhang, D. Xiao, Y. Chang, in *Security and Communication Networks*. A novel image authentication with tamper localization and self-recovery in encrypted domain based on compressive sensing, vol 2018, Article ID 1591206 (2018), p. 15. https://doi.org/10.1155/2018/1591206
55. J. Molina-García, R. Reyes-Reyes, V. Ponomaryov, C. Cruz-Ramos, in *2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW)*. Watermarking algorithm for authentication and self-recovery of tampered images using DWT (Kharkiv, 2016), pp. 1–4. https://doi.org/10.1109/MSMW.2016.7538148
56. A. Tiwari, M. Sharma, R.K. Tamrakar, Watermarking based image authentication and tamper detection algorithm using vector quantization approach. AEU-Int. J. Electron. C. **78**, 114–123 (2017). https://doi.org/10.1016/j.aeue.2017.05.027