# Proposed Method

We have proposed a Sudoku and MD5 hash algorithm based self-embedding fragile watermarking method to recover tampered images. Sudoku is a logic puzzle, usually played on a 9× 9 grid, based on a game called Latin Square [4]. The aim of the puzzle is to fill the 9 × 9 grid logically with 3 × 3 sub-blocks, columns and rows using the numbers 1 to 9 and the previously given numbers. In this proposed method, block-based image recovery data is embedded in four different regions of the image using the Sudoku algorithm. Then, the authentication data generated using the MD5 hash algorithm is embedded in the LSBs of the image blocks. The proposed algorithm comprises four main stages; sequentially, recovery bits embedding stage which is based on Sudoku; authentication bits embedding stage; tamper detection stage and recovery stage.

## Recovery bits embedding stage

This section comprises the generation of recover bits and embedding them into an image. Figure 1shows the block diagram of how the recovery bits are generated and how they are embedded into the image. This stage starts with setting the first and second LSBs of pixel values of the original gray-level image, I o, to zero. First LSBs are used for embedding authentication bits and second LSBs are used for storing recovery bits. The , I z,image is divided into twenty-five nonoverlapping blocks according to 5×5 Sudoku shown in the Fig. 1. Due to the random and single solution structure of Sudoku, resistance against tamper attacks is provided. Then, the image which is divided by Sudoku is grouped according to the Sudoku numbering, as formed 1, 2, 3, 4, 5. Thus, different regions of the image are used for embedding the recovery bits. Then, the sub-group which is grouped according to the Sudoku map is selected and the blocks in this sub-group are divided into non-overlapping 5 × 5 sub-blocks. For each sub-block, the average value of the sub-block is calculated to generate the recovery data, and this value is converted to a binary value to generate the recovery bits. A sample of generating the recovery bits for a sub-block is shown in Fig. 2. 24-bit quad recovery information for each sub-block is constructed by concatenating the first 6-bits of recovery bits of the other four sub-blocks as shown in Fig. 3, The generated 24-bit recovery bits are embedded into the second LSBs of the other sub-blocks in the group, excluding itself. The procedure of recovery bits embedding stage is represented as follows:

Set the first and second LSBs of original image, , pixel values to zero.

## Experimental Results and Analysis

Experiments were conducted on a group of gray-level images to verify the effectiveness of the proposed scheme. In the experiment, the sizes of the divided non-overlapping image blocks were 8 × 8, i.e., s = 8. The parameter k for block mapping was set to 2, which means two blocks map into one block and each block can also map into the other two different blocks. Obviously, the larger the parameter value of λ was set, the better performances of tampering detection and content recovery become, because larger embedding capacity can provide more valid reference bits and corresponding detailed information to recover the tampered blocks and more authentication bits can reduce the probabilities of false judgment for the tampered block and the reference bits. But, on the other hand, the larger the value of λ was set, more hiding payload were caused, which led to the degradation of visual quality for the watermarked image. During the watermark embedding procedure of our scheme, we set the parameter λ to 3 for sufficient embedding capacity, which means that three LSBs of each cover block were replaced with watermark bits and the five MSBs were preserved. Assume that the distribution of the three LSBs of cover image is uniform. Thus, the average energy of distortion caused by the watermark embedding for each pixel is:

|  |  |
| --- | --- |
| ,  where a - number of filled blocks;  b - number of empty blocks. | (2) |

The theoretical peak signal-to-noise ratio (PSNR) of the watermarked image can be calculated approximately using Eq. (9), which demonstrates the visual quality of watermarked image after embedding with λ = 3 is also satisfactory: 37.92 dB.

|  |  |
| --- | --- |
|  | (3) |

In order to meet the relationship in Eq. (2), Eq.(3), DCT coefficients of direct-current and lower frequencies representing the principle content in each block were used to produce reference bits, i.e., Q = 15. The assigned bits, i.e., r1, r2, …, r15, for the representation of the 15 DCT coefficients, i.e., c1, c2, …, c15, in each block were: 8, 6, 6, 4, 4, 4, 3, 3, 3, 3, 2, 2, 2, 2, 2, individually. Thus, the length R of the reference bits for each block was 54. The lengths L1 and L2 of the authentication bits for five MSBs and 54-bits reference bits were 40 and 22, respectively. Fig. 3 shows the standard cover image Lena sized 512 × 512 and its watermarked version. The PSNR value of the watermarked image is 37.98 dB. It can be observed from Fig. 3 that the visual distortion due to watermark embedding is imperceptible.



**Figure 3:**Lena.

For convenience of description, we divide the 512 bits of each 8 × 8 received watermarked block into three sets: 1) Φ1: 320 bits of the five MSBs, 2) Φ2: 40 bits of the authentication bits for the five MSBs, 3) Φ3: 152 bits of the reference bits and corresponding authentication bits for the two mapping blocks. The 192 bits in Φ2 and Φ3 come from the three LSBs of each block. We first conducted the testing of meaningful content tampering for the proposed scheme, in which the bits of Φ1, Φ2, and Φ3 for each modified block may all be changed randomly. Figs. 4-6 show the results of tampering detection and content recovery for our scheme. In Figs. 4-6, subfigures (a) show the original images Lena, Elaine, and Lake, and all of them are sized 512 × 512; subfigures (b) show the tampered, watermarked versions of (a) correspondingly; subfigures (c) are the tampered detection results, in which the black blocks indicate the tampered regions; subfigures (d) show the content recovery results for (b) correspondingly. The tampering percentages of subfigures (b) in Figs. 4-6 are 2.44%, 7.40%, and 9.67%, respectively, and the PSNR values of the recovered images in subfigures (d) of Figs. 4-6 are 37.39 dB, 36.11 dB, and 32.66 dB, respectively. Because the watermark

***Table 1:****Results of tampering recovery for Scenario 1.*

|  |  |  |  |
| --- | --- | --- | --- |
| Number of tampered block pairs | Recovered percentage  Of [23] | Recovered percentage of proposed scheme | PSNR of recovered results by proposed scheme(dB) |
| 50 | 48.00% | 100% | 37.14 |
| 100 | 48.50% | 100% | 36.53 |
| 250 | 46.69% | 100% | 35.08 |
| 500 | 40.90% | 100% | 33.92 |
| 750 | 36.44% | 100% | 33.92 |
| 1000 | 31.80% | 100% | 32.22 |

***Table 2:*** *Results of tampering recovery for Scenario 2.*

|  |  |  |  |
| --- | --- | --- | --- |
| Number of tampered block pairs | Recovered percentage | PSNR of the scheme(dB) | PSNR of the proposed scheme (dB) |
| 50 | 48.00% | 37.66 | 37.74 |
| 100 | 48.50% | 37.48 | 37.70 |
| 250 | 46.69% | 36.70 | 37.61 |
| 500 | 40.90% | 36.15 | 37.45 |
| 750 | 36.44% | 35.59 | 37.42 |
| 1000 | 31.80% | 35.49 | 37.20 |

## Basic Principles and Characteristics of Semi-Fragile Watermark

Semi-fragile watermarking algorithms for image authentication focus on the ability of detection, location and recovery from tamper attacks carried out on images. Because the semi-fragile watermark has a certain degree of fragility, the algorithm can realize the image authentication according to whether the watermarking data is tampered with. When the host image suffers attacks, the watermarking information will make a corresponding change as well. However, for normal image-processing operations, the semi-fragile watermark also has a certain degree of robustness, so that the normal image processing operations and malicious tampering operations can be distinguished. The framework of the semi-fragile watermarking authentication system contains the watermark embedding process and the watermark detection process. The watermark embedding process is shown in Figure 1, and the watermark detection process is shown in (the dashed part may not be needed in some cases).

At present, a lot of semi-fragile watermarking algorithms for content authentication have appeared. In general, to achieve good performance, a semi-fragile watermarking algorithm must meet the following characteristics:

* Balance between robustness and fragility

On the one hand, after experiencing malicious tampering, such as collage attack, constant average attack, and statistical analysis attack, a watermark should be able to make an alarm response to show its fragility; on the other hand, after experiencing reasonable image-processing operations, such as JPEG compression, brightness adjustment, affine transform, filtering, and noise addition, a watermark should be able to accept them to show its robustness. Therefore, the watermark embedded into the semi-fragile watermarking algorithms must balance the robustness and the fragility well;

* Invisibility

After embedding watermark into the image, the quality of the image cannot have obvious degeneration, and the embedded watermark needs to be invisible to ensure that the original image is not damaged;

* Blind detection

Sometimes people cannot find the original image in many applications, and even if sometimes the original image is provided by a third party, people still cannot believe it. So this requires that content authentication algorithms can achieve the detection process without original images;

* Localization and recoverability

If the image has been maliciously tampered with, the algorithm should be able to accurately determine and locate the tampered regions, and then recover it using the recovery data;

* Security

Attacks on the security of watermarks are intentional, and the next step is to threaten the robustness of the watermark. Semi-fragile watermarking algorithms should be able to resist malicious tampering, and cannot be easily detected, imitated, tampered with or fabricated by unauthorized parties;

* Collision resistance

Means that for two different messages, it should be difficult to find the same digest;

* Preimage resistance

Given only a hash value (message digest), it should be difficult to find any message hashing (or preimage) that generates that digest;

* Second preimage resistance

Given some message, it should be difficult to find a different message having the same hash.

## Characteristics of fragile watermarking schemes

The characteristics of fragile watermarking schemes are imperceptibility, data payload, computational cost, blind detection, tamper detection, and tampered image recovery. These properties are as follows:

1. Imperceptibility: Imperceptibility implies that embedded watermarks should be invisible. This measures the similarity between the original image and the watermarked image. 19308 Multimedia Tools and Applications (2021) 80:19307–19333;
2. Data payload: Data payload refers to the number of bits that can be embedded in the cover image;
3. Computational cost: It refers to the amount of time taken for watermark generation, embedding, extraction, tamper detection, and recovery;
4. Blind detection: This implies that watermark extraction and tamper detection can be done in the absence of the original cover image;
5. Tamper detection: Fragile watermarking scheme should detect any change in the image content and able to locate the tampered regions precisely;
6. Tampered image recovery: Once the tampered regions are located, a fragile watermarking scheme should be able to recover the tampered regions to the original contents of the image.