

# A Secure Image Authentication Scheme with Tamper Localization and Recovery

Neena Raj N. R.<sup>a</sup> and Shreelekshmi R.<sup>b,1</sup>

<sup>a</sup> *Department of Computer Science and Engineering, College of Engineering Trivandrum (Affiliated to APJ Abdul Kalam Technological University), Thiruvananthapuram - 695016, Kerala, India*

<sup>b</sup> *Department of Computer Science and Engineering, Government Engineering College (Affiliated to APJ Abdul Kalam Technological University), Thrissur - 680009, Kerala, India*

**Abstract.** This paper proposes a secure image authentication scheme that can locate the tampered regions, recover the lost contents and hide application-specific sensitive data. In this scheme, an encrypted watermark that comprises tamper localization, recovery and application-specific information is placed in the selected pixels, which is extracted and decrypted to identify the tampered regions, recover the tampered regions approximate to original image contents and extract the hidden data. The watermark is highly secure and sensitive to any modification in the image. The proposed scheme ensures lossless recovery of the original image and data from an untampered image. The experimental results show that this scheme generates watermarked image of high quality and has high resistance to copy-move, image splicing, vector quantization and collage attacks. As compared with state-of-the-art schemes, the proposed scheme provides better recovered image quality under extensive tampering.

**Keywords.** Image authentication, Fragile watermarking, Tamper localization, Image recovery, Data hiding

## 1. Introduction

Rapid growth in communication and digital imaging technology enables anyone to share images on various platforms like e-healthcare, e-commerce, e-governance, social media, etc. Security of these images is a serious concern since the availability of powerful image manipulation software enables anyone to tamper images without leaving visual traces. Tampering the image contents misleads its purpose in medical diagnosis, forensic investigation, military applications, etc. These situations demand content authentication and verification of image integrity.

Fragile watermarking is an effective solution for strict content authentication since any modifications in the protected image lead to changes in the embedded watermark [1,2]. D. Singh and S. K. Singh proposed a fragile watermarking scheme [3] in which the recovery information is generated using Discrete Cosine Transform (DCT). They proposed another scheme [4] using Block Truncation Coding (BTC). In these schemes,

---

<sup>1</sup> Corresponding Author: Shreelekshmi R., Department of Computer Science and Engineering, Government Engineering College, Thrissur - 680009; E-mail: shreelekshmir@gmail.com

the watermark generated for a block is placed in another block known as a mapping block to resist collage attack [5]. This causes misclassification of untampered blocks if their mapping block is tampered. Low tamper localization accuracy causes degradation in the recovered image quality.

To increase recovery rate under extensive tampering, Haghighi et al. [6] generated recovery information using Lifting Wavelet Transform (LWT) and halftoning. This scheme cannot resist collage attack. Qin et al. [7] proposed a scheme in which the recovery information is generated using an Optimal Iterative Block Truncation Coding (OIBTC) algorithm and distributed over the image. This scheme provides high quality recovered image under extensive tampering. But this scheme cannot resist collage attack. Rajput and Ansari [8] proposed a recovery scheme in which four copies of the recovery information are placed in the different parts of the image. The watermarked image quality is low since the watermark is placed in the 4 Least Significant Bits (LSB) of image pixels. This scheme is vulnerable to four scanning attack since the watermark is less secure [5]. Shen et al. [9] generated authentication information using Singular Value Decomposition (SVD) and recovery information using block average.

Gul and Ozturk [10] proposed a pixel-wise authentication scheme in which recovery information is generated using the average value of image blocks. Kim and Yang [11] proposed a scheme based on Absolute Moment Block Truncation Code (AMBTC). The recovery information is computed using AMBTC and placed in the LSBs using Optimal Pixel Adjustment Process (OPAP). This scheme provides low recovered image quality under extensive tampering. The schemes [3,4,6,7,8,9,10,11] cannot recover original image from an untampered received image. The schemes [12,13] can hide sensitive data like Electronic Patient Record (EPR) along with tamper localization. But these schemes are vulnerable to copy-move, Vector Quantization (VQ), collage and chosen cover-image attacks since the watermark is not protected [5].

Most of the existing schemes have limitations such as low watermarked image quality, low tamper detection accuracy, less security and poor recovery ability under extensive tampering. The majority of schemes focus on grayscale images only. In this scenario, we propose a fragile watermarking scheme that can locate the tampered regions, recover the lost contents and also hide data in the image. The scheme can support both grayscale and color and can be effectively used in images carrying sensitive information like medical, biometric. The major highlights of the proposed scheme include:

1. High watermarked image quality.
2. High security.
3. High tamper detection accuracy.
4. Lossless recovery of the original image contents and sensitive information from the untampered watermarked image.
5. High recovered image quality under extensive tampering.
6. High resistance to copy-move, image splicing, VQ and collage attacks.

The remaining part of this paper is organized as follows. Section 2 presents the proposed scheme. Section 3 discusses experimental results and performance comparison with state-of-the-art schemes before concluding in section 4.

## 2. Proposed Scheme

The proposed scheme creates a cover image  $C$  from the original image  $I$  of size  $h \times w \times c$  with  $b$  bits per channel where  $h$ ,  $w$  and  $c$  represent height, width and number

of channels in  $I$  respectively, using Pixel to Block conversion (PTB) technique [12]. The cover image is divided into  $N$  blocks of size  $4 \times 4$  and a watermark of length 24 bits is generated for each channel in the image block. The watermark contains an 8-bit Tamper Localization Code (TLC), 14-bit Self Recovery Code (SRC) and two bits from the application-specific sensitive data. The TLC is computed from the image pixels using SVD [14] and logistic map to make it sensitive to the changes in the image contents. The SRC contains the recovery information of two different blocks which are randomly selected using a block mapping sequence generation procedure. The watermark is encrypted using the logistic map and is embedded into the 2 LSBs of 12 pixels in the image block. The remaining 4 pixels are kept unchanged to ensure the recovery of image pixels in the original image from untampered regions of the watermarked image. A smoothing function is used to enhance the watermarked image quality. The tamper localization is performed in multilevel to increase the localization accuracy and resist various tampering attacks. The recovery is carried out in two levels for improving the recovery rate under extensive tampering. The major phases in the proposed scheme are watermarked image generation, tamper localization and image recovery.

### 2.1. Watermarked image generation

The watermarked image generation phase includes chaotic sequence generation, block mapping sequence generation, cover image generation, watermark generation, encryption and embedding.

1. Chaotic sequence generation: In the proposed scheme, a chaotic sequence  $S$  of length  $31N$  is generated using a logistic map  $x_{n+1} = \mu x_n(1 - x_n)$ , where  $n \geq 0$ ,  $0 \leq x_0 \leq 1$  and  $0 \leq \mu \leq 4$  [15]. The value of  $\mu$  is selected from the range  $[3.57, 4]$  to ensure the chaotic nature of  $S$ . Initial condition  $x_0$  and the control parameter  $\mu$  together forms  $k_1$ , which is used as a key in the chaotic sequence generation. Further  $S$  is partitioned into  $SM, ST, SE$  of length  $N, 4N, 26N$  respectively from front to rear.
2. Block mapping sequence generation: A look-up table  $L$  of size  $\frac{h}{2} \times \frac{w}{2}$  is constructed to store terms in the block mapping sequence. Initially,  $L$  is assigned with  $1, 2, \dots, N$  in row-major order. Odd columns in the left half are swapped with corresponding columns of the right half to maximize the recovery rate. Then  $L$  is partitioned into sub-blocks  $TL, TR, BL, BR$  each of size  $\frac{h}{4} \times \frac{w}{4}$ . The cells in each sub-block are scrambled using  $SM$  to provide security and randomness in block mapping. To eliminate mapping to the same block, a cell in the left half of  $L$  which is mapped to itself is swapped with the corresponding cell in the right half. Similarly, cells in the right half which are mapped to itself are swapped with cells in the left half.
3. Cover image generation: PTB conversion technique [12] is used to generate the cover image  $C$  of size  $2h \times 2w \times c$  from the original image  $I$ . Further  $C$  is divided into  $4 \times 4$  blocks  $C_1, C_2, \dots, C_N$  in row-major order.
4. Watermark generation and embedding: Watermark of length 24 bits is generated for each channel  $d$  in  $C_i$ 's mapping block  $C_j$  and placed in  $C_j$  after encrypting the watermark.
  - (a) TLC computation: The TLC is computed by using pixels in the image block and a randomly selected chaotic terms in  $ST$ . It ensures sensitivity of TLC towards changes in image contents and key which in turn helps to resist copy-move,

image splicing, VQ and collage attacks. To compute TLC, the sequence  $ST$  is partitioned into  $N$  sets and a set is chosen at random using random numbers in  $[1, N]$  generated based on key  $k_2$ , image identifier and  $d$ . The block  $C_j$  is divided into  $2 \times 2$  sub-blocks and matrix Aof size  $2 \times 2$  is formed using the first pixel from each sub-block of  $C_j$  in row-major order.  $A$  is changed to  $A'$  by applying XOR operation to each element with  $[(ST_l \times 10^{k_3}) \bmod 2^b]$  where  $ST_l, l = 1, 2, 3, 4$  are the terms in selected subset of  $ST$ . Then  $msv' = [msv \times (2^{2b} - 1) / 2(2^b - 1)]$  is computed, where  $msv$  is the maximum singular value of  $A'$  obtained by applying SVD.  $msv'$  is converted into binary format of length  $2b$  and partitioned into 8 groups. The members of each group are XORED together to form 8 bit TLC.

- (b) SRC generation: Applying the same procedure in TLC computation, matrices  $A$  and  $B$  are created from the block  $C_i$  and its partner block  $C_k$  respectively. Here  $k = i + \frac{N}{2}$  if  $i \leq \frac{N}{2}$  and  $i - \frac{N}{2}$  otherwise. The 7 Most Significant Bits (MSBs) of  $A$ 's average and  $B$ 's average together form 14 bit SRC.
- (c) Watermark formation: A 24 bit watermark  $W = TLC \parallel SRC \parallel SD$  is computed where  $SD$  represents two bits from sensitive data.
- (d) Watermark encryption: The generated watermark  $W$  for each channel  $d$  is encrypted using the sequence  $SE$ . The sequence is partitioned into  $N$  sets and a set  $R$  is selected using random numbers in  $[1, N]$  generated based on key  $k_4$ , image identifier and  $d$ . Compute  $t_l = [(R_l \times 10^{k_5}) \bmod 2^{12}]$  for  $l = 1, 2$  in which key  $k_5 > 3$ . The binary values of  $t_1$  and  $t_2$  are concatenated to form  $t$ . Then  $t$  and  $W$  are XORED to obtain  $T$ . A permutation  $p$  is computed based on the new indices obtained after sorting the last 24 terms in  $R$ . The encrypted watermark  $W'$  is obtained by applying  $p$  to the bits in  $T$ . Since the chaotic sequence  $SE$  is sensitive to the control parameter and initial condition the encrypted watermark is highly key sensitive and secure.
- (e) Watermark embedding: The mapped block  $C_j$  is divided into  $2 \times 2$  sub-blocks and the first pixel of each sub-block is classified as unchangeable pixels and remaining as changeable pixels. The encrypted watermark  $W'$  is placed in 2 LSBs of changeable pixels. A smoothing function is applied to improve the watermarked image quality. The difference between the changeable pixel and the unchangeable pixel is computed. The value of changeable pixel is decreased by 4 if the result is 3 and increased by 4 if the difference is  $-3$ .

## 2.2. Tamper detection and localization

The tamper detection and localization procedure includes watermark extraction, decryption and a multilevel tamper localization process. The suspected image  $G$  is divided into non overlapping blocks  $G_i, i = 1, 2, \dots, N$  of size  $4 \times 4$ .

1. Watermark extraction: The watermark  $E$  for each channel is extracted from the 2 LSBs of 12 changeable pixels in the block  $G_i$  in row-major order.
2. Watermark decryption: To decrypt the extracted watermark  $E$  corresponding to the mapped block of  $G_i, t$  and  $p$  are computed as per the watermark encryption procedure in section 2.1. The inverse permutation of  $p$  is applied to  $E$  and the result is XORED with  $t$  to obtain the decrypted watermark. The first 8 bits of the decrypted

watermark from the TLC, the next 14 bits indicate SRC and the remaining 2 bits represent bits from the sensitive data.

3. Tamper localization: Initially each block  $G_i$  is considered as valid.

*Level 1:* The block  $G_i$  is divided into  $2 \times 2$  sub-blocks and the first sub-block is chosen. The 2 LSBs from a changeable pixel are extracted and inserted into the 2 LSBs of the unchangeable pixel. Then the smoothing function is applied. If the obtained value is different from the value of changeable pixel, the block is considered invalid. Otherwise, repeat the process with the remaining changeable pixels. If no mismatch is found, repeat the process with the next sub-block. This level helps to resist image splicing attack.

*Level 2:* For each valid block after the first level, the extracted TLC is compared with the computed TLC. If a mismatch is found, the block is considered invalid. Since TLC depends on block contents and position this level helps to resist copy-move attack.

*Level 3:* For each block that remained as valid after the second level, if at most one among the neighboring blocks is valid, the block is changed as invalid.

*Level 4:* For each valid block  $G_i$ , the average value of unchangeable pixels is computed and the 7 MSBs are stored in  $a$ . Then  $a$  is compared with the first 7 bits in the SRC of the mapped block if  $G_i$  is located in the upper half, otherwise with the last 7 bits in the SRC. If a mismatch is found  $G_i$  is changed as invalid. For a valid block, if the number of invalid neighboring blocks is greater than or equal to the number of invalid neighboring blocks of its mapped block, the block is considered invalid. If only less than 3 neighboring blocks of an invalid block are invalid, the block is turned as valid and if more than 5 neighboring blocks of a valid block are invalid, it is considered invalid. Since embedded recovery information enables block-wise dependency, this level helps to resist VQ and collage attacks.

A block is considered as tampered if it becomes invalid either in level 3 or 4.

### 2.3. Image recovery

The recovered image  $R$  of size  $h \times w \times c$  is initialized with zeros and divided into  $2 \times 2$  blocks  $R_i, i = 1, 2, \dots, N$ .

*Level 1:* For each  $i$ , if  $G_i$  is not tampered, then the unchangeable pixels in the sub-blocks constitute pixels in  $R_i$ . For a tampered block  $G_i$  with untampered mapping block  $G_j$ ,  $R_i$  is obtained from the retrieved SRC of  $G_j$ . If  $G_i$  is located in the upper half,  $b - 7$  zeros are appended to the first 7 bits in SRC to form  $b$  bits intensity. Otherwise, the intensity is obtained from the last 7 bits in SRC. The new intensity value is assigned to  $R_i$ . If  $G_i$  and  $G_j$  are tampered and the mapping block  $G_k$  of  $G_i$ 's partner block is not tampered, then SRC is retrieved from  $G_k$  and  $R_i$  is formed as per the previous case.

*Level 2:* The blocks which are not recovered in the first stage are recovered using an inpainting algorithm [16].

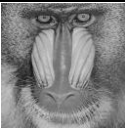
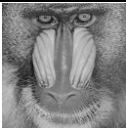
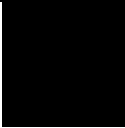
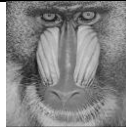




















## 3. Experimental results

To evaluate the performance of the proposed scheme for tamper detection and recovery, the experiments are conducted on 96 grayscale and 61 color images selected from CVG-UGR database [17]. The proposed scheme is implemented in MATLAB R2017b

environment. In the experiments, the parameters are randomly initialized as  $k_1 = (x_0, \mu) = (0.8566, 3.998)$ ,  $k_2 = 100$ ,  $k_3 = 6$ ,  $k_4 = 101$ ,  $k_5 = 6$  and image identifier is the file name in the database.

*Watermarked image quality analysis:* Peak Signal to Noise Ratio (PSNR) is used to measure the imperceptibility of the watermark in the cover image. The average PSNR of watermarked image for grayscale images is 45.37 dB and is enhanced to 47.46 dB after applying smoothing function. In the case of color images, it is enhanced to 47.32 dB from 44.98 dB.

*Resistance against tampering attacks:* To measure the resistance of the proposed scheme against tampering attacks, we used three metrics viz. True Positive Rate (TPR) or Tamper Detection Rate (TDR), True Negative Rate (TNR) and precision (p) [5]. PSNR is used to measure the similarity between recovered and original images. Copy-move, image splicing, VQ and collage attacks [5] are simulated in each watermarked image at various tamper ratios. Figure 1 shows the visual performance of proposed scheme on tamper localization and recovery on sample images from CVG-UGR database [17].

Watermarked image	PSNR (dB)	Tampered image	Tamper ratio	Tamper localization	TPR TNR p	Recovered image	PSNR (dB)
	47.62		0%		100 100 100		$\infty$
	47.61		10%		100 100 100		40.16
	47.61		30%		100 100 100		36.86
	47.62		50%		99.99 100 100		31.04
	47.62		70%		99.99 100 100		26.1
	47.61		90%		100 100 100		23.3

**Figure 1.** Visual performance of the proposed scheme on tamper localization and recovery on sample images from CVG-UGR database [17].

Table 1 shows the average TPR, TNR and p under various tamper ratios on grayscale images and color images separately. The obtained results are above 99.9% and 100% in

almost all cases. This shows that the proposed scheme has high resistance against copy-move, image splicing, VQ and collage attacks.

**Table 1.** Tamper detection performance against various attacks

Attacks	Metric	Grayscale images					Color images				
		Tamper ratio (%)									
		10	30	50	70	90	10	30	50	70	90
Copy-move	TPR	99.9933	99.9926	99.9965	99.9917	99.995	100	100	100	99.9997	100
	TNR	100	100	100	100	100	100	100	100	100	100
	p	100	100	100	100	100	100	100	100	100	100
Image Splicing	TPR	99.9959	99.9976	99.9968	99.9991	99.9983	100	100	100	100	100
	TNR	100	100	100	100	100	100	100	100	100	100
	p	100	100	100	100	100	100	100	100	100	100
VQ	TPR	99.9964	99.994	99.9972	99.9979	99.9968	100	100	100	100	100
	TNR	100	100	100	100	100	100	100	100	100	100
	p	100	100	100	100	100	100	100	100	100	100
Collage	TPR	100	99.9969	99.9969	99.9997	99.9978	100	100	100	100	100
	TNR	100	100	100	100	100	100	100	100	100	100
	p	100	100	100	100	100	100	100	100	100	100

*Comparison with state-of-the-art schemes:* The performance of the proposed scheme is compared with recent schemes [12,7,6,8,4] in terms of watermarked and recovered image quality. To compare the recovery performance, each watermarked image is replaced with white pixels at various tamper ratios from left to right. Table 2 shows the average PSNR of the watermarked image and recovered image on grayscale images. From the table, it is evident that the watermarked image quality of the proposed scheme is higher than other schemes. The proposed scheme and scheme proposed by Parah et al. [12] provide lossless recovery of original image from an untampered watermarked image. It is also observed that the recovered image quality of the proposed scheme is better than the other schemes under extensive tampering.

**Table 2.** Performance comparison of proposed scheme with state-of-the-art schemes

Schemes	Watermarked image quality	Recovered image quality with various tamper ratio									
		0	10	20	30	40	50	60	70	80	90
Parah et al. [12]	46.39	∞	16	12.76	10.75	9.32	8.2	7.3	6.47	5.76	5.14
Qin et al. [7]	44.12	44.12	38.18	35.82	33.8	32.41	31.26	19.7	18.81	13.32	12.77
Haghighi et al. [6]	46.41	42.81	35.25	32.74	30.8	29.22	28.32	22.62	18.49	15.56	13.05
Rajput and Ansari [8]	32.52	28.22	24.93	22.59	21.69	20.49	21.22	10.77	7.1	5.39	4.8
D. Singh and S. K. Singh [4]	37.24	37.24	16.37	12.99	10.97	9.56	8.43	7.42	6.6	5.84	5.2
Proposed scheme	47.46	∞	40.2	36.32	34.01	32.16	30.79	28.74	26.91	25.07	22.87

4. Conclusion

A highly secure fragile watermarking based image authentication scheme for tamper localization, recovery of tampered regions and hiding sensitive information is proposed in this paper. The generated fragile watermark is highly sensitive to any modification in the image. This scheme ensures lossless recovery of the original image and hidden data from an untampered received image. Experimental results show that the proposed scheme generates watermarked image with high perceptual quality and can locate the tampered regions precisely. This scheme has high resistance against copy-move, image splicing, VQ and collage attacks. It has better recovered image quality under extensive

tampering as compared with state-of-the-art schemes and can be applied to color, grayscale and DICOM images. The proposed scheme can be used in sensitive applications which need image authentication scheme that ensures 100% recovery of image and hidden data, high security, precise tamper localization and high recovery rate.

## Acknowledgement

Authors extend gratitude to the Department of Higher Education, Government of Kerala, for granting the research fellowship.

## References

- [1] Raj NN, Shreelekshmi R. Blockwise fragile watermarking schemes for tamper localization in digital images. In 2018 International CET Conference on Control, Communication, and Computing (IC4); 2018. p. 441-446.
- [2] Raj NN, Shreelekshmi R. Security Analysis of Hash Based Fragile Watermarking Scheme for Image Integrity. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT); 2019. p. 651-654.
- [3] Singh D, Singh SK. DCT based efficient fragile watermarking scheme for image authentication and restoration. *Multimedia Tools and Applications*. 2017; 76(1): p. 953-977.
- [4] Singh D, Singh SK. Block Truncation Coding based effective watermarking scheme for image authentication with recovery capability. *Multimedia Tools and Applications*. 2019; 78(4): p. 4197-4215.
- [5] Raj NN, Shreelekshmi R. A survey on fragile watermarking based image authentication schemes. *Multimedia Tools and Applications*. 2021; 80: p. 19307-19333.
- [6] Haghighi BB, Taherinia AH, Harati A. TRLH: Fragile and blind dual watermarking for image tamper detection and self-recovery based on lifting wavelet transform and half-toning technique. *Journal of Visual Communication and Image Representation*. 2018; 50: p. 49-64.
- [7] Qin C, Ji P, Chang CC, Dong J, Sun X. Non-uniform watermark sharing based on optimal iterative BTC for image tampering recovery. *IEEE MultiMedia*. 2018; 25(3): p. 36-48.
- [8] Rajput V, Ansari IA. Image tamper detection and self-recovery using multiple median watermarking. *Multimedia Tools and Applications*. 2020; 79(47): p. 35519-35535.
- [9] Shen JJ, Lee CF, Hsu FW, Agrawal S. A self-embedding fragile image authentication based on singular value decomposition. *Multimedia Tools and Applications*. 2020; 79(35): p. 25969-25988.
- [10] Gul E, Ozturk S. A novel pixel-wise authentication-based self-embedding fragile watermarking method. *Multimedia Systems*. 2021;: p. 1-15.
- [11] Kim C, Yang CN. Self-embedding fragile watermarking scheme to detect image tampering using AMBTC and OPAP approaches. *Applied Sciences*. 2021; 11(3): p. 1-21.
- [12] Parah SA, Ahad F, Sheikh JA, Bhat GM. Hiding clinical information in medical images: a new high capacity and reversible data hiding technique. *Journal of biomedical informatics*. 2017; 66: p. 214-230.
- [13] Geetha R, Geetha S. Embedding electronic patient information in clinical images: an improved and efficient reversible data hiding technique. *Multimedia Tools and Applications*. 2020; 79(19): p. 12869-12890.
- [14] Klema V, Laub A. The singular value decomposition: Its computation and some applications. *IEEE Transactions on automatic control*. 1980; 25(2): p. 164-176.
- [15] Mandal MK, Banik GD, Chattopadhyay D, Nandi D. An image encryption process based on chaotic logistic map. *IETE Technical Review*. 2012; 29(5): p. 395-404.
- [16] Shen J, Chan TF. Mathematical models for local nontexture inpaintings. *SIAM Journal on Applied Mathematics*. 2002; 62(3): p. 1019-1043.
- [17] The CVG-UGR Image Database. [Online]. [Accessed on January 2021]. Available from: <https://ccia.ugr.es/cvg/dbimagenes/index.php>.