

Semi Fragile Watermarking Technique using IWT and a Two Level Tamper Detection Scheme

Nandhini Sivasubramanian and Gunaseelan Konganathan

Department of Electronics and Communication Engineering, College of Engineering, Guindy, Anna University, Chennai, Tamil Nadu, India

Keywords: Semi Fragile Watermarking, Integer Wavelet Transform, Image Processing, Tamper Detection, Hamming Distance.

Abstract: A semi fragile watermarking technique using a two level thresholding scheme for tamper detection is proposed. The proposed embedding technique uses two level IWT (integer wavelet transform) to embed the authentication watermark. The authentication watermark generated from the approximate coefficients is stored in the detail coefficients using least significant substitution to form the watermarked image. The proposed tamper detection technique for identifying attacks in the watermarked image is a two level thresholding scheme using normalized hamming similarity (NHS) and a tamper detection map. The performance of the proposed technique was evaluated for a variety of content preserving manipulations and malicious attacks. The proposed technique produces a better performance in terms of an increased PSNR (Peak Signal to Noise Ratio) of the watermarked image and by localizing the malicious attacks when compared to the existing techniques. The significant performance of the proposed semi fragile watermarking technique is due to the combined results from both the NHS and the tamper detection map which helps in localizing the malicious attacks and identifying the incidental manipulations. Also, the authentication watermark which is a copy of the original image helps in identifying the tampered regions in the attacked watermarked image.

1 INTRODUCTION

The present digital age of communication calls for a secured way for communicating the confidential information from one remote terminal to another. Watermarking is one of the important techniques for communication as it authenticates the received data and also helps in identifying the attacks to the data. Watermarking can be classified into fragile and semi fragile. Fragile watermarking is sensitive even to a single pixel change in the watermarked image and hence making it unsuitable for watermarking images in a noisy environment. On the other hand, semi fragile watermarking is tolerant to incidental manipulations to the watermarked image which are called content preserving Manipulations. The incidental manipulations include addition of noise to the watermarked image, image compression, Blurring etc. Semi fragile watermarking techniques are also sensitive to deliberate malicious attacks to the watermarked image making it suitable for using it in noisy environment.

Most of the existing semi fragile watermarking

techniques rely on Discrete Cosine transform (DCT) or Discrete wavelet transform (DWT) to hide the watermark. The strategy which is used in semi fragile watermarking techniques is to embed the features of an image as a watermark. Some of the existing DWT based semi fragile watermarking techniques are discussed in this section. DWT based watermarking technique (Hang and Park,2003) embeds the just noticeable feature as a watermark. Hu and Han(2005) embed the features generated from the low frequency wavelet coefficients. A DWT based Zernike moments is used as a feature in (Liu et al., 2005). Hang and Sun (2003) embed the semi fragile watermark by combining it with the human visual model. Some techniques quantize the wavelet coefficients to embed the watermark. Preda (2013) embeds the watermark by quantizing the second level DWT coefficients. Tsai and Chien (2008) embeds the watermark into the second level DWT coefficient using two different quantization parameters. Preda et al., (2015) embeds the watermark by quantizing the mean of a group of second level coefficients. The drawback of all these

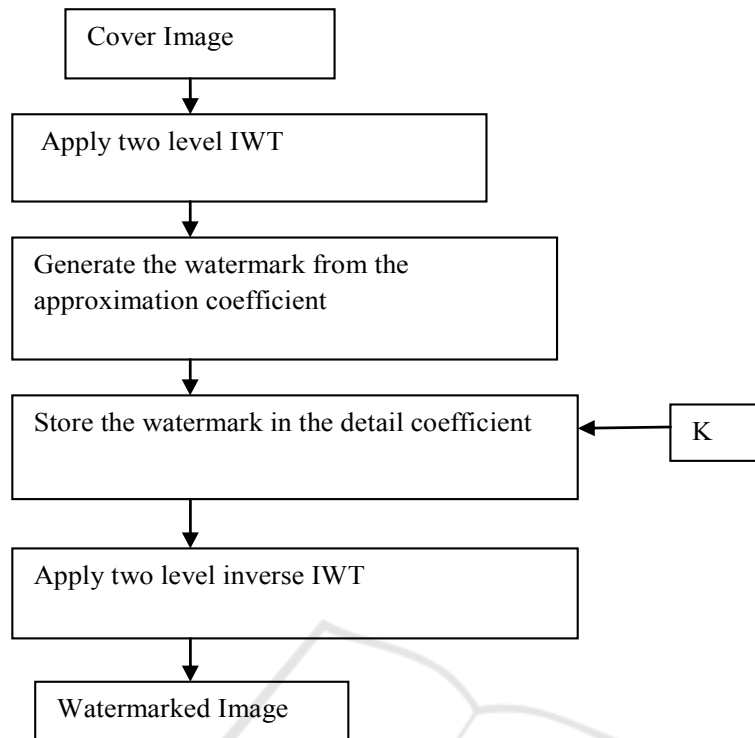


Figure 1: Flow chart of the proposed embedding technique.

schemes is that they are tolerant only to JPEG compression and the effect of other content preserving manipulations is not discussed thoroughly. There are only a few semi fragile watermarking techniques that are tolerant to a variety of incidental manipulations. Tiwari et al., (2017) proposed a novel watermarking technique based on vector quantization and modified index key modulation. Benrhouma et al., (2015) proposed a technique based on cat map and DWT. Qi and Xin (2011) used a non traditional quantization method to modify one chosen approximation coefficient. Lai(2011) used singular value decomposition and Tiny GA for semi fragile watermarking purpose. In some of these approaches the PSNR value of the watermarked image is very less and some approaches do not discuss the effect of geometric attacks on the watermarked images.

The proposed technique tries to address the above drawbacks by proposing an embedding technique that preserves the visual quality of the watermarked image and by proposing a tamper detection technique for testing the watermarked image to different content preserving manipulations including geometric attacks.

2 FRAMEWORK OF THE PROPOSED TECHNIQUE

2.1 Proposed Embedding Technique

A flow chart of the proposed embedding Technique is shown in Figure 1. Let the size of the cover image, I used in the proposed technique be $M \times M$. In order to obtain the watermark and embed it, integer wavelet transform (IWT) is used to decompose the cover image. Equation (1) represents the first level decomposition of the cover image using IWT results in one approximation coefficient CA and three details coefficient CH , CV , CD which are of size $(M/2) \times (M/2)$. The detail coefficient CH is again decomposed according to equation (2) to obtain four sub bands AA , AH , AV and AD which are of size $(M/4) \times (M/4)$.

$$[CA, CH, CV, CD] = iwt2(I) \quad (1)$$

$$[AA, AH, AV, AD] = iwt2(CH) \quad (2)$$

In the proposed technique the approximation coefficient sub band AA is used to generate the watermark which will be used for authentication at

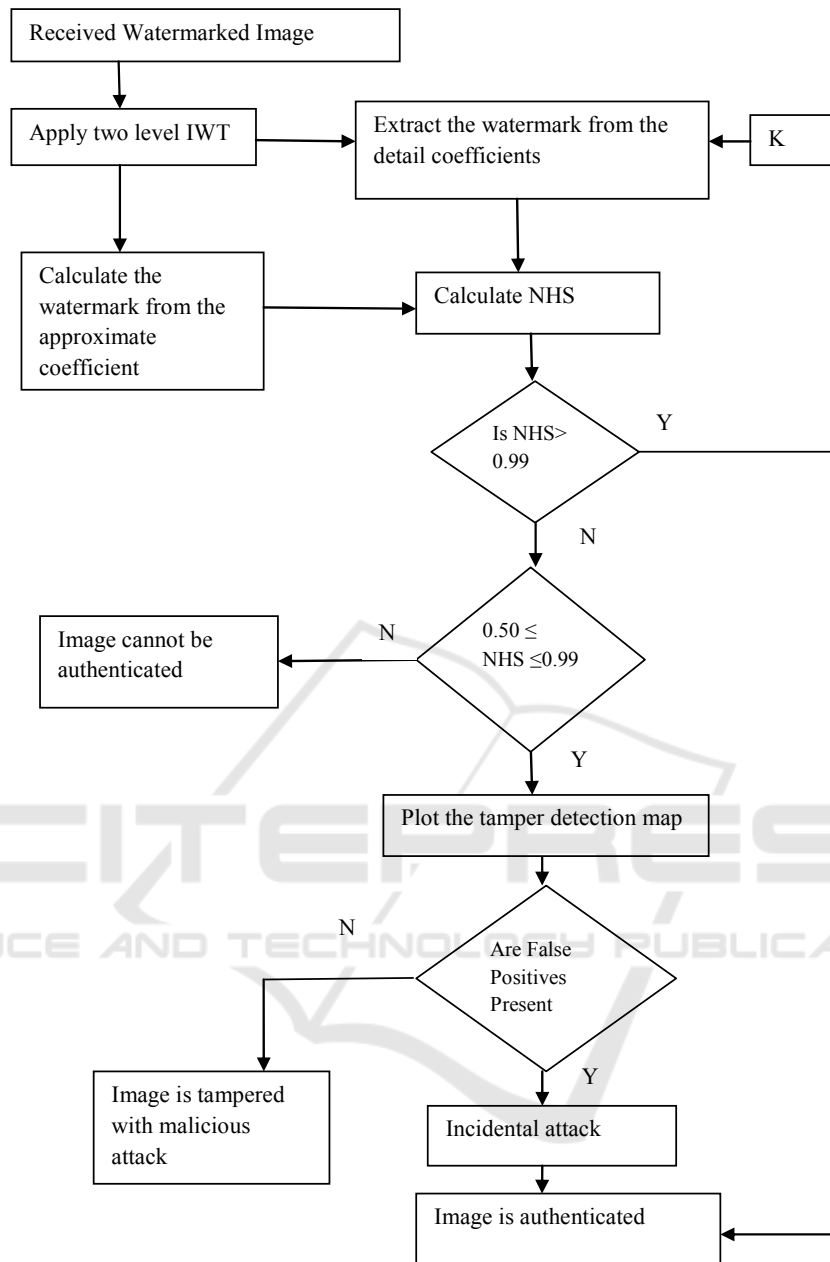


Figure 2: Flow chart of the proposed tamper detection technique.

the receiver end. The watermark, W which is of size $(M/4) \times (M/4)$ is obtained using equation (3).

$$W(i, j) = \text{xor}(\text{dec2bin}(AA(i, j))) \quad (3)$$

$$1 \leq i, j \leq (M / 4)$$

In equation (3) dec2bin represents the decimal to 8-bit binary conversion of a pixel at the position (i, j) and xor represents the logical exoring of the resultant bits to obtain the watermark at the position (i, j) .

$$RW = W \oplus K \quad (4)$$

In order to improve the security of the generated watermark, W is exored with shared secret key matrix K to form RW which is shown in equation (4). The shared secret key is a randomly generated matrix of ones and zeros which is of size $(M/4) \times (M/4)$. The resultant watermark RW is embedded into the detail coefficient sub band AV using least significant bit substitution which are shown by equations (5),(6) and (7).

$$B(1:8) = dec2bin(AV(i, j)) \quad (5)$$

$$B(8) = RW(i, j) \quad (6)$$

$$AV'(i, j) = bin2dec(B) \quad (7)$$

In equation (5), every element of AV is converted into binary bits and the watermark at its corresponding position is embedded into the least significant bit of AV which is B(8) to obtain AV'. The final step shown in equations (8) and (9) is the image reconstruction through inverse IWT to obtain the watermarked image WI.

$$CH' = iwt2(AA, AH, AV', AD) \quad (8)$$

$$WI = iwt2(CA, CH', CV, CD) \quad (9)$$

2.2 Proposed Tamper Detecting Technique

A flow chart of the proposed Tamper detection Technique is shown in Figure 2. Suppose the received image WI is tampered via incidental manipulations or malicious attacks. The proposed tamper detection technique to differentiate an incidental/content preserving manipulation from a malicious attack is explained below:

The first step is the two level decomposition of the received image WI using IWT which are shown in equations (10) and (11).

$$[CA1, CH1, CV1, CD1] = iwt2(WI) \quad (10)$$

$$[AA1, AH1, AV1, AD1] = iwt2(CH1) \quad (11)$$

In order to identify the tampered portions of the received watermarked image, watermarks CW and EW are to be obtained. EW which is of size (M/4) x (M/4) is the extracted watermark from the least significant bits of AV1 as shown in equations (12) and (13). CW which is of size (M/4) x (M/4) is the calculated watermark from AA1 using equation (14).

$$B(1:8) = dec2bin(AV1(i, j)) \quad (12)$$

$$EW(i, j) = B(8) \oplus K(i, j) \quad (13)$$

$$CW(i, j) = xor(dec2bin(AA1(i, j))) \quad (14)$$

$$1 \leq i, j \leq (M/4)$$

Normalized hamming similarity (NHS) (Lu et al,2005) is calculated between CW and EW using equation (15) in order to know the effectiveness of the attack on the watermarked image.

$$NHS = 1 - \frac{HD(CW, EW)}{N \times N} \quad (15)$$

In equation (15), HD is the hamming distance between CW & EW and N x N is their corresponding size. Hamming distance represents the number of positions at which CW and EW differs and this variation is shown using the tamper detection map. Using HD, NHS is calculated whose value ranges from 0 to 1. NHS value of 1 indicates that both CW and EW are identical and there is no attack on the watermarked image. Therefore, higher values of NHS signify that the calculated watermark is more similar to that of the embedded watermark. In order to distinguish the incidental manipulations from that of the malicious attacks a threshold of 0.99 is fixed on the NHS value (Tiwari et al., 2017). The significance of this threshold is that a value of NHS higher than 0.99 implies that the watermarked image is free from malicious attacks and it is automatically authenticated. If the value of NHS is less than 0.99 and greater than 0.50, then a tamper detection map is plotted to ascertain the nature of attacks.

In order to plot the tamper detection map, at first the tampered regions have to be identified. The tampered regions are obtained from the hamming distance calculated between CW and EW. Hamming distance represents the corresponding positions where the calculated and embedded watermarks mismatch. In other words as CW is obtained from AA1, the corresponding positions from HD can be directly mapped onto AA1. At this stage, the elements of AA1 will be labeled either as authenticated or tampered. In order to refine the tamper detection process neighbourhood approximation is used.

Example 1: Illustration when a tampered pixel is identified as authenticated.

Tampered	Authenticated	Authenticated
Authenticated	Tampered	Tampered
Tampered	Authenticated	Authenticated

Example 2: Illustration when a tampered pixel is identified as tampered.

Tampered	Tampered	Authenticated
Authenticated	Tampered	Tampered
Tampered	Authenticated	Authenticated



Figure 3: Some of the cover images used for testing (a) Baboon (b) Peppers (c) Lena (d) Goldhill (e) Fishing Boat (f) Barbara.

The labeling of the eight neighbours of an element in AA1 is taken into account to finalize whether an element is tampered or not. As shown in example 1, if the number of tampered neighbours surrounding a tampered element is less than three then the corresponding element is identified as authenticated. As shown in example 2, if the number of tampered neighbours surrounding a tampered element is more than three then the corresponding element is identified as tampered. By this way the labeling of the elements in AA1 is fine tuned to plot the tamper detection map. The tamper detection map shows the spread of tampered and authenticated elements in AA1. In order to detect malicious attacks from the incidental manipulations, it is important to identify any pattern in the tamper detection map. An identification of a well defined pattern outlining an area in the tamper detection map clearly indicates that the attack is malicious (Benrhouma et al., 2015). If the potentially tampered elements are scattered all over the detection map like a random noise and if it does not contain any isolated tampered coefficients then the elements are false positives and should be considered as authentic (Preda et al., 2015). The final step is the reconstruction of the received image using equations (16) and (17).

$$CH1' = iwt2 (AA1', AH1, AV1, AD1) \quad (16)$$

$$RI = iwt2 (CA, CH1', CV, CD) \quad (17)$$

3 RESULTS AND DISCUSSION

The cover images used for testing the proposed tamper detection technique are of size 512x512. Some of the cover images used are shown in Figure 3. PSNR (peak signal to noise ratio) is calculated using equation (18) between the cover and the watermarked images to access the visual

quality of the visual quality of the watermarked images.

$$PSNR = 10 \times \log_{10} \left(\frac{255 \times 255}{MSE} \right) \quad (18)$$

Where $MSE = \frac{\sum_{i=1}^M \sum_{j=1}^M (X_{i,j} - Y_{i,j})^2}{M \times M}$

It can be shown from table 1 that the average PSNR value using the proposed embedding technique exceeds the acceptable value of 38 dB (Voloshynovskiy et al., 2001). The efficiency of the proposed technique was tested for a variety of content preserving manipulations and malicious attacks.

Table 1: PSNR of the watermarked images.

Cover Image	PSNR of the watermarked image
Lena	41.80 dB
Baboon	31.96 dB
Barbara	39.63 dB
Peppers	42.01 dB
Gold Hill	41.21 dB
Airplane(F-16)	41.92 dB
Sailboat on Lake	38.75 dB
Fishing boat	41.39 dB
Elaine	40.05 dB

Table 2: NHS values for various watermarked image with salt and pepper noise (sigma: 0.01).

Cover Image	NHS value
Lena	0.9261
Barbara	0.9249
Elaine	0.9247
Airplane(F-16)	0.9244
Fishing Boat	0.9261
Peppers	0.9230
Sailboat on Lake	0.9268

Table 3: NHS values for various watermarked image with rotation (degree: 45).

Cover Image	NHS value
Lena	0.7414
Barbara	0.7433
Elaine	0.7347
Airplane(F-16)	0.7381
Fishing Boat	0.7374
Peppers	0.7476
Sailboat on Lake	0.7375

Table 4: NHS values for various watermarked image with image brightening (Contrast Limits: 0.1 & 0.6).

Cover Image	NHS value
Lena	0.6135
Barbara	0.5554
Elaine	0.6342
Airplane(F-16)	0.8339
Fishing Boat	0.5806
Peppers	0.6274
Sailboat on Lake	0.6818

3.1 Evaluation of the Proposed Technique in Terms of Incidental Manipulations

In order to prove the efficiency of the proposed tamper detection technique in terms of incidental manipulations, a variety of content preserving attacks were considered. An attack is classified as incidental if the NHS value is greater than 0.99. If the NHS value is between 0.50 and 0.99 then the tamper detection map is to be considered for identifying it. Initially, the salt and pepper noise was added to the

watermarked image and the corresponding NHS value was calculated. It can be inferred from table 2 that for various images the average NHS value after adding salt and pepper noise comes to 0.90. It can also be inferred from table 3 and 4 that rotating an watermarked image by 45 degrees and adjusting the contrast parameters produces an average NHS value between 0.5 and 0.9. So, in order to correctly identify it as an incidental manipulation tamper detection map was plotted as can be shown in table 7. The first row of table 7 shows the tamper detection map when salt and pepper noise is added to the watermarked image of 'peppers'. The second row of table 7 shows the tamper detection map when the contrast of the watermarked airplane image was adjusted to 0.1 and 0.6. It can be inferred from the tamper detection map that the tampered pixels are scattered all over the image and it does not produce a defined pattern. Due to the above reasons, the contrast adjustment manipulation and the addition of salt and pepper noise is identified as incidental. In the same way when the watermarked image was attacked by various incidental manipulations like speckled noise, gamma correction, wiener filtering and motion blurring the proposed tamper detection technique produced NHS value between 0.5 and 0.9 as shown in table 5. The fifth row of table 7 shows the tamper detection map when the watermarked 'cameraman' image was manipulated by using wiener filtering (with sigma :0.01). As the tampered pixels are not isolated and are scattered all over the image, the wiener filtering attack can be conclusively identified as incidental. The Possible parameter values for the content preserving manipulations for which the

Table 5: NHS values for various values of content preserving manipulations for 'lena'.

Incidental Manipulation	Parameter Value	NHS value
Salt and Pepper Noise	Sigma: 0	1
	Sigma:0.1	0.5914
Gamma Correction	Gamma: 0	1
	Gamma: 2	0.5367
Wiener Filtering	Filter size:3x3	0.5140
Speckle Noise	Sigma: 0	1
	Sigma: 0.1	0.5068
Gaussian Blur	Sigma:4	0.6178
Image Brightening	Contrast Limits:0.3 & 0.7	0.5923
Motion Blur	Len: 5, theta: 45	0.5145
	Len :20,theta:45	0.5240
Rotation	Degree:6	0.5855
	Degree:45	0.7433
	Degree:80	0.62

proposed tamper detection technique will identify as incidental and not malicious is given in table 6 .The better performance of the proposed technique is because a copy of the image in the form of watermark is used for tamper detection.

3.2 Evaluation of the Proposed Technique in Terms of Malicious Attacks

The efficiency of the proposed tamper detection technique was also tested for malicious attacks like object addition and deletion. The main objective of the proposed two level thresholding is to properly identify malicious attacks from incidental manipulations. The malicious attacks was found to produce a NHS value that was greater than 0.9. Since the proposed technique is a two level thresholding process, an object addition or deletion is clearly outlined in the tamper detection map. This results in identifying it as a malicious attack. As shown in the third and the fourth column of the table 7, an object addition or deletion to the original image clearly

outlines the tampered part which shows where the malicious attack had taken place. The better performance of the proposed tamper detection technique is due to the two level thresholding of NHS and tamper detection map to identify incidental manipulations from malicious attacks. Further the proposed embedding technique almost embeds a copy of the original image by using a watermark of size 128x128 which helps in identifying the tampered elements at the receiver end.

Table 6: List of Incidental manipulations and its parameters.

Manipulations	Parameters
Salt and pepper Noise	Sigma:0-0.1
Speckle Noise	Sigma:0-0.1
Gaussian Blur	Sigma:2-5
Motion Blur	Len:5-20,theta:45
Gamma Correction	Gamma:0.5-1.5
Rotation	Degree:5-80
Wiener Filtering	Size:3x3
Image Brightening	Contrast Limits:0.3 & 0.7

Table 7: Tamper Detection map for various types of attacks.



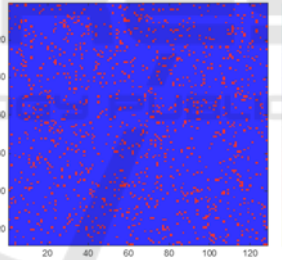


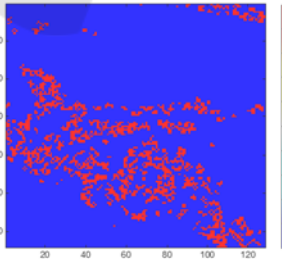
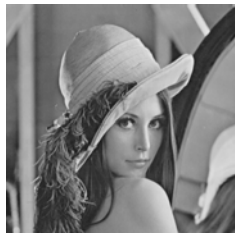
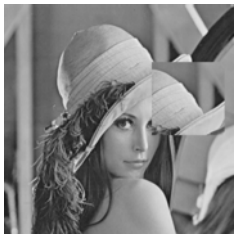
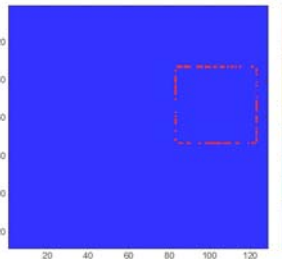
	Cover Image	Attacked Image	Tamper Detection Map	Classification
1.				Incidental
2.				Incidental
3.				Malicious

Table 7: Tamper Detection map for various types of attacks (cont.).

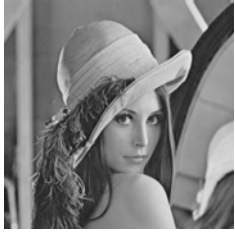
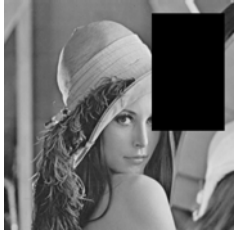
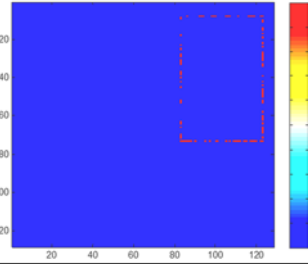


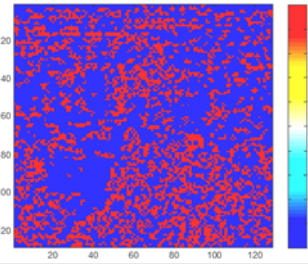
	Cover Image	Attacked Image	Tamper Detection Map	Classification
4.				Malicious
5.				Incidental

Table 8: Comparison of the characteristics of the proposed technique with various methods.

Paper	Technique	Maximum PSNR(dB)	Tamper Localization	Attacks Classification
Shen and Chen, 2012	DWT technique	30	---	JPEG compression, Mean and Median Filtering, Noise
Preda, 2013	DWT based approach	40	Yes	JPEG compression, Filtering
Li et al., 2015	Two level DWT	36	Yes	JPEG compression, Gaussian Noise.
Zhang et al., 2016	DWT based approach	40	---	JPEG compression, Salt & Pepper and Gaussian Noise, Speckle Noise, Image Rescaling
Shojanazeri et al., 2017	DWT and Zernike Moments	40.9	Yes	JPEG compression, Rotation, Scaling, Translation, Additive Noise.
Proposed	IWT based technique	42	Yes	Salt and pepper Noise, Speckle Noise, Gaussian Blur, Motion Blur, Gamma Correction, Rotation, Wiener Filtering, Image Brightening

Finally, table 8 compares the characteristics of the proposed technique with the existing methods.

4 CONCLUSION

A semi fragile watermarking technique using integer wavelet transform and a two level thresholding scheme to identify attacks in the watermarked image is proposed. Due to the usage of LSB substitution to embed the authentication watermark, the degradation in the visual quality of the watermarked

image is reduced. As a result, the PSNR of the watermarked images using the proposed embedding technique is greater when compared to the existing techniques. On analyzing the proposed tamper detection technique to a variety of content preserving manipulations like addition of noises, blurring, filtering, geometric attacks, image brightening it is found that the image authenticity is correctly verified. When malicious attacks like object addition and object deletion was tested on the watermarked image, the tampered pixels was clearly outlined in the tamper detection map. The better

performance of the proposed technique was due to the two level thresholding scheme of NHS and tamper detection map to identify the tampered portions in the watermark image.

REFERENCES

- H. Kang, and J. H. Park, "A semi-fragile watermarking using JND," in *Proceedings of the Pacific Rim Workshop on Digital Steganography (STEG 2003)*, 2003, pp. 127–131.
- Y. P. Hu, and D. Z. Han, "Using two semi-fragile watermarks for image authentication," in *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, China, 2005*, pp. 5484–5489.
- H. Liu, J. Lin, and J. Huang, "Image authentication using content based watermark," in *Proceedings of IEEE International Symposium on Circuits and Systems, Kobe, Japan, 2005*, pp. 4014–4017.
- H. Yang, and X. Sun, "Semi-fragile watermarking for image authentication and tamper detection using HVS model," in *Proceedings of International Conference on Multimedia and Ubiquitous Engineering, Seoul, South Korea, 2007*, pp. 1112–1117.
- R. O. Preda, "Semi-fragile watermarking for image authentication with sensitive tamper localization in the wavelet domain," *Measurement*, vol.46, no.1, pp.367–373, January 2013.
- M. J. Tsai, and C. C. Chien, "A wavelet-based semi-fragile watermarking with recovery mechanism," in *Proceedings of the IEEE International Symposium on Circuits and Systems, Seattle, WA, USA, 2008*, pp. 3033–3036.
- R. O. Preda, I. Marcu, and A. Ciobanu, "Image authentication and recovery using wavelet-based dual watermarking," *UPB Science Bulletin, Series C*, vol.77, no.4, pp.199–212, 2015.
- A. Tiwari, M. Sharma, and R.K. Tamrakar, "Watermarking based image authentication and tamper detection algorithm using vector quantization approach," *AEU -International Journal of Electronics and Communications*, vol.78, pp.114–123, August 2017.
- O. Benrhouma, H. Hermassi, and S. Belghith, "Tamper detection and self-recovery scheme by DWT watermarking," *Nonlinear Dynamics*, vol.79, no.3, pp.1817–1833, February 2015.
- X. Qi, and X. Xin, "A quantization-based semi-fragile watermarking scheme for image content authentication," *Journal of Visual Communication and Image Representation*, vol.22, no.2, pp.187–200, February 2011.
- C. C. Lai, "A digital watermarking scheme based on singular value decomposition and tiny genetic algorithm," *Digital Signal Processing*, vol.21, no. 4, pp.522–527, July 2011.
- Z. M. Lu, D. G. Xu, and S. H. Sun, "Multipurpose image watermarking algorithm based on multistage vector quantization," *IEEE Transactions on Image Processing*, vol.14, no.6, pp. 822–831, June 2005.
- S. Voloshynovskiy, S. Pereira, T. Pun, J.J. Eggers, and J.K.Su, "Attacks on digital watermarks: classification, estimation based attacks and benchmarks," *IEEE Communications Magazine*, 2001, 39 (8), 118–126, August 2001.
- H. Shen, and B. Chen, "From single watermark to dual watermark: a new approach for image watermarking," *Computers & Electrical Engineering*, vol.38, no.5, pp.1310–1324, September 2012.
- C. Li, A. Zhang, Z. Liu, L. Liao, and D. Huang, "Semi-fragile self-recoverable watermarking algorithm based on wavelet group quantization and double authentication," *Multimedia Tools and Applications*, vol.74, no.23, pp.10581–10604, December 2015.
- Z. Zhang, C. Wang, and X. Zhou, "Image watermarking scheme based on DWT-DCT and SSVD," *International Journal of security and its applications*, vol.10, no.10, pp. 191–206, October 2016.
- H. Shojanazeri, W. A. W. Adnan, S. M. S. Ahmad, and S. Rahimipour, "Authentication of images using Zernike moment watermarking," *Multimedia Tools and Applications*, vol.76, no.1, pp.577–606, January 2017.