ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2021.0334

SEMI-FRAGILE BLIND WATERMARKING MECHANISM FOR COLOR IMAGE AUTHENTICATION AND TAMPERING

Hiral A. Patel[1] and Dipti B. Shah[2]

[1]Department of Computer Science, Sutex Bank College of Computer Applications and Science, India
[2]Department of Computer Science, G.H. Patel P.G. Department of Computer Science and Technology, India

Abstract

Digital Image watermarking data hiding technique is proposed in paper to provide authentication to color image. It has capability of tamper detection and tamper localization. The suggested system is blind watermarking system which doesn't require original information at destination. The system is robust against non-malicious attacks where as fragile with malicious attacks. The watermark is generated based on content of original image and it is embedded using IWT and QIM techniques. To the destination place for authentication purpose, embedded watermark is extracted. For classification of image, extracted and generated watermark from watermarked image are compared and based on this differential information, tampering is analyzed. The image is classified as authentic or tampered as per clustering of pixels within differential information. If image is declared as tampered then tampered region is identified using clustering detail. The PSNR is achieved up to 43.57dB and NCC is near to 1. As per numbers of parameters concern and as per experiments, it is found that proposed system is the improved version of existing systems.

Keywords:

Image Watermarking, Image Authentication, Tamper Detection, Tamper Localization, Color Image

1. INTRODUCTION

Digital images are essential part of human life. At each stage, people need these files to transfer from one location to another. With fast development of IT sector, these images can easily modified using powerful image processing tools like Photoshop, Corel-draw. During image transmission, many operations are applied on image which helps in transferring image. This type of image manipulation is known as Non-malicious modification which doesn't harm anyone. Some users intentionally modify image which can harm someone is known as Malicious modification. This malicious modification is one of the electronic security issues which are called tampering. Using watermarking concept, tampering issue can be solved.

To address this issue, many digital watermarking based tamper detection, localization and recovery techniques have been suggested by researchers. Numbers of systems were suggested for gray-scale images [1]-[12] and for color images [13]-[20]. Some watermarking systems were developed for tamper detection and localization [1] [2] [5] [7] [8] [19] where as some systems also tried to recover original image from tampered image [3] [4] [6] [9]-[17] [20].

The main factor of watermarking system is watermark. This watermark can be individual file or it can be generated based on content of original image. With authentication system, if watermark is generated based on content of original image is more suitable as compare to using individual file. To embed watermark within original image, frequency domain is more suitable than spatial domain because of its robustness nature [21]. DFT, DCT and Wavelet transforms are different methods of frequency transform. Wavelet transforms are more suitable because of its multi resolution description feature [22]. QIM technique is used for implementing concept of blind watermarking system [14] [23] and it also gives protection against JPEG Compression [23].

Tampering can be detected by comparing original and extracted watermark where original information is required at destination place [2] [5]-[8]. On the other hand, for detection of tampering, extracted watermark and generated watermark from watermarked image are compared so original image is not required at destination [3] [4] [10] [12] [13] [20]. To find difference between these two binary images: bit to bit checking, XOR or absolute difference method is used. Image is tampered or not that can easily find using this difference directly [1] [2] [6]-[8] [10] [11] [14] or some extra post processing can apply to binary image for getting more accurate outcomes.

For post processing, morphological operations can be applied to image which can remove noise from binary image [13] [16] [17]. Median filter can be applied to the image for noise removal [17]. Image filling can be applied to binary image to fill holes [13]. Eight neighbor pixel or block testing is also used to identify whether the pixel is tampered or not [3]-[5] [9] [12] [20].

For temper detection, process has been divided into multiple levels to improve the accuracy level [9] [12] [15]. In these papers, at first level, watermark difference was found based on 2×2 sub-blocks based on bit to bit matching. In paper [15], at second level, all sub-blocks' code are combined and compared if these are not exactly matched then the 4×4 block was treated as tampered. At third level, 20 bits recovery code was compared for providing more security to collage attack. In papers [9] [12], at second level, for 4×4 block, if any of its sub-block or minimum 2 sub-blocks are tampered then this 4×4 block was also treated as tampered. At third level, they applied neighbor pixel testing method. If more than 5 neighbors were tampered then this 4×4 block is also considered as tampered.

If any attack is not applied on image then obviously watermark differential image becomes blank. In watermark difference image, if pixels are scattered all over image then it is found that non-malicious attack is applied on image and if pixels are grouped at specific location then it is found that malicious attack is applied on image [1] [3]. Classification of tamper detection can be possible on these bases. If clusters are found within the image then there is a need to identify the location of those clusters to find tampered region. Authors are inspired with this concept and are tried to solve tamper detection and localization. Post processing is also applied on differential image for getting more accurate outcomes.

The rest of paper is organized as: section 2 discusses the proposed algorithm. Section 3 shows the experimental results. Section 4 expresses the comparison of proposed system with existing systems and finally conclusion is discussed.

2. PROPOSED ALGORITHM

The proposed watermarking system is divided into numbers of processes which are discussed in detail. The detailed framework is discussed in [24].

2.1 WATERMARK GENERATION PROCESS

The watermark is generated from features of original image. For extracting features of original image, first the color image is converted to gray-scale image and IWT is applied. LL sub-band of image is used to generate watermark. LL sub-band is divided into non-overlapping 4x4 blocks which also internally divided into 2×2 blocks. The mean of 4x4 block and all 2×2 sub-blocks are calculated individually. If mean of 4×4 block is greater than the means of 2×2 block then it is considered as 0 else 1. Like this for each 4×4 block 4 watermark bits are generated. Now all watermark bits of 4×4 blocks are combined and is used as original watermark (WM) of size 64×64×4.

2.2 WATERMARK EMBEDMENT PROCESS

The generated watermark is embedded within blue channel of color image. DWT is applied on blue channel. LL sub-band is used for embedding watermark. It is divided into 4×4 non-overlapping blocks. Watermark is embedded at specific position see Fig.1. Also the QIM method is applied here.

The QIM rules is given below:

$$\tilde{y} = Rounddown\left(\frac{y}{\Delta}\right) \times \Delta \qquad (1)$$

If *WM*=0

$$y' = C_0 = \begin{cases} y & \text{if } y \in \left[\tilde{y}+\alpha, \tilde{y}+\frac{\Delta}{2}-\alpha\right] \\ \tilde{y}+\alpha & \text{if } y \in \left[\tilde{y}, \tilde{y}+\alpha\right] \\ \tilde{y}+\frac{\Delta}{2}-\alpha & \text{if } y > \tilde{y}+\frac{\Delta}{2}-\alpha \end{cases} \qquad (2)$$

If *WM*=1

$$y' = C_1 = \begin{cases} \tilde{y}+\frac{\Delta}{2}+\alpha & \text{if } y < \tilde{y}+\Delta-\alpha \\ \tilde{y}+3\times\frac{\Delta}{4} & \text{if } y \in \left[\tilde{y}+\frac{\Delta}{2}-\alpha, \tilde{y}+\frac{\Delta}{2}\right] \\ \tilde{y}+\frac{\Delta}{2}+\alpha & \text{if } y \in \left[\tilde{y}+\frac{\Delta}{2}, \tilde{y}+\frac{\Delta}{2}+\alpha\right] \\ y & \text{if } y \in \left[\tilde{y}+\frac{\Delta}{2}+\alpha, \tilde{y}+\Delta-\alpha\right] \\ \tilde{y}+\Delta-\alpha & \text{if } y > \tilde{y}+\Delta-\alpha \end{cases} \qquad (3)$$

where *y* is pixel value, Δ is quantization interval and *α* is scaling factor. All 4×4 blocks are merged and inverse DWT is applied. The blue color channel is replaced within original image's blue color channel and this image is used as watermarked image (WMD).

	1		2
	3		4

Fig.1. Pixel Positions within 4×4 block

2.3 WATERMARK EXTRACTION AND GENERATION PROCESS

At destination place, WMD is received. The embedded watermark is extracted from WMD i.e. EWM and again watermark is generated from WMD i.e. GWM.

Select WMD image and apply DWT to blue plane of WMD. LL sub-band is divided into 4×4 blocks. Apply QIM de-quantization process and extract the watermark.

De-quantization rules is given below:

$$R_1 = Rounddown\left(\frac{y'+\frac{\Delta}{2}}{\Delta}\right) \times \Delta \qquad (3)$$

$$R_2 = Rounddown\left(\frac{y'}{\Delta}\right) \times \Delta \qquad (4)$$

$$VM = \begin{cases} 0 & if\ R_1 = R_2 \\ 1 & Otherwise \end{cases} \qquad (5)$$

All watermark bits of 4×4 blocks are combined and this data is used as EWM of size 64×64×4.

To generate watermark from the WMD image, follow the steps of Watermark Generation Process. Consider watermark as GWM of size 64×64×4.

2.4 TAMPER DETECTION PROCESS

This process is used to detect tampering. This process returns the classification of WMD as "Image is Authentic", "Image is Authentic with Non-malicious attacks" or "Image is tampered". The steps are as follow:

Step 1: First of all, compare EWM and GWM using XOR operation and resulting data is binary image named DIFF_I.

Step 2: The above DIFF_I image has EWM and GWM image differential information about each 4×4 blocks. So for each 4×4 block, two types of information are tested: one for 4×4 block tamper testing which helps for tamper localization and another for 2× sub-block tamper testing which helps for tamper detection.

a. ***2×2 Sub-Blocks Tamper Testing***: DIFF_I image has differential information for each 4×4 blocks. Now for each block, 4 bits are stored which individually used to check the tampering within 2× sub-blocks of same 4×4 block. The first sub-block's bit is stored at first position and so on. Here each bit of the 4×4 block is compared individually. If bit is 0 then the respective sub-block is considered as valid so 0 bit otherwise sub-block is considered as tampered so 1 bit is set for 2×2 sub-block. Let this 2×2 block tamper tested result image is named as TAM_2. It is 128×128 sized binary image. It is used for classification of WMD.

b. ***4×4 Block Tamper Testing***: Compare 4 bits pattern of each 4×4 block from DIFF_I image. If all bits are 0 then block is considered as valid block so 0 bit otherwise it is considered as tampered block so 1 bit is set. Let consider this 4×4 block tamper tested result image as TAM_4. It is of 64×64 sized binary image and it is used for tamper localization purpose if image is classified as Tampered.

Step 3: For the classification of watermarked image WMD, select TAM_2 image.

Step 4: Find out clusters of TAM_2 image. Let total numbers of clusters are N.

Step 5: If attack is not applied to image then numbers of clusters are 0. If non-malicious attack is applied to image then pixels are spread all over image so numbers of clusters are more. If malicious attack is applied to image then pixels are not scattered all over the image but it is found within specific area or in a group so numbers of clusters are less. So the total numbers of clusters (N) are compared with the threshold value (T).

Step 6: The watermarked image is classified as per the numbers of clusters (N). If N is zero then the image is considered as “Authentic” but if the N is greater than threshold value (T) then the image is considered as “Authentic with non-malicious attack” and if N is less than threshold value (T) then the image is considered as “Tampered.”

2.5 TAMPER LOCALIZATION PROCESS

If the image is classified as “Authentic or Authentic with non-malicious attacks” then this tamper localization process is not executed. If image is classified as “Tampered” then only Tamper Localization process will be executed. This process identifies tampered region. The steps are as follow:

Step 1: Select TAM_4 binary image.

Step 2: The post processing operation is applied on TAM_4 binary image using 8 neighboring blocks and if all 8 neighboring blocks are tampered then the current block is also considered as tampered. Like this all block related post processing is applied and holes of the image are filled.

Step 3: Finds numbers of clusters of TAM_4 image.

Step 4: Calculate total numbers of pixels in each cluster. Now find clusters having only 1 pixel and set this pixel value as 0 in TAM_4 image. It helps in removing extra noise.

Step 5: Store positions of each pixels of cluster. Now set boundary outside all these clusters which shows tampered region where tampering was applied within WMD. This region will help if the recovery process will be proceeded.

3. EXPERIMENT RESULTS

An algorithm is implemented using Matlab R2017a. Color images of 512×512 size are selected for testing algorithm which are shown in Fig.2. The respective watermarked images are shown in Fig.3. The algorithm is implemented with Lossless format of image i.e. with PNG and TIFF. The size of the generated watermark and extracted watermark is 64×64×4 = 16384 bits.

Fig.2. Original Images

Fig.3. Watermarked Images

Table.1. PSNR and NCC Results

Images	PSNR	NCC
Lena.png	42.9730	0.9998
Pepper.png	44.1457	1
Flower.png	45.9153	1
Lena.tif	42.1917	1
Baboon.tif	42.6095	1
Average	**43.5670**	**0.99996**

Table.2. Tamper Detection Performance with Non-malicious attacks

Attacks	Classification
No attack	Authentic
Gaussian Noise	Not Tampered
Salt and Pepper	Not Tampered
JPEG compression	Not Tampered
Low Pass Filter	Not Tampered
Median Filter	Not Tampered
Sharpen Image	Not Tampered

Table.3. Tamper Detection and Localization Performance with Malicious attacks

Attacks	Classification
Text addition with background	Tampered
Text addition without background	Tampered
Object removal	Tampered
Copy move	Tampered
Splicing	Tampered

From the Fig.2 and Fig.3, it is clear that after embedding the watermark within the image, the generated watermarked image is imperceptible with human eye. No visible difference found between original and watermarked image.

Table.4. Comparison with Existing system

Parameters	[5]	[19]	[20]	Proposed Method
Image	Grey	Color	Color	Color
PSNR	41.79 dB	48 dB	44.63 dB	43.57 dB
System	Semi-fragile	Fragile	Fragile	Semi-fragile
Extraction Process	Blind	Non-Blind	Blind	Blind
Malicious Attack	Text Addition Cropping	Object removal Collage Cropping	Object Removal	Copy move Splicing Object removal Text addition
Non-malicious attack	Gaussian Noise Salt and pepper JPEG Compression Filtering Rotation Blur	-	-	Gaussian Noise Salt and pepper JPEG Compression Low Pass filter Median Filter Sharpen image
Issue	Tamper detection Tamper localization	Tamper detection Tamper localization	Tamper detection Tamper localization recovery	Tamper detection Tamper localization

Imperceptibility between Original and Watermarked image is measured using PSNR and Robustness between original and extracted watermark is measured using NCC based on my embedment and extraction algorithms are demonstrated Table.1.

For testing the Authenticity of the image, different attacks were applied on watermarked images. Non malicious attacks like Gaussian Noise, Salt and Pepper, JPEG Compression, Low Pass Filter, Median Filter, and Sharpen Image are selected for experiment.

Malicious attacks like text addition with background and without background, object removal, copy move and splicing attacks are applied using Adobe Photoshop to the watermarked image by applying above attacks at different locations and with different size.

In this paper, only the outputs related with Lena.png image is shown. The Table.2 demonstrated the results with Non malicious attacks and Table.3 shows the results with malicious attacks.

In Table.2, different non malicious attacks' results are demonstrated. As per TAM_2 image, watermarked image is properly classified as "Authentic" when no attack is applied and "Not Tampered" when non-malicious attacks are applied. And if the malicious attacks are applied to the image, then the image is classified as "Tampered". The related information is demonstrated in Table.3.

The Table.3 demonstrated different malicious attacks' results. As per TAM_2 image, watermarked image is properly classified as "Tampered" with malicious attack. When the image is classified as tampered then Tamper Localization Process is proceed further and using TAM_4 binary 4×4 blocks' image, tampered region is found. As malicious attacks, text addition with background and without background, copy move, splicing and object removal (10% of image) are demonstrated.

4. COMPARISON WITH EXISTING SYSTEM

The proposed system is compared with existing systems as per some of parameters and results are shown in Table.4. The comparison of proposed system is done with three systems named Archana's system [5], Nirali's system [19] and Molina's system [20]. The Archana's system was developed for grayscale image who achieved PSNR up to 41.79dB, Nirali's system was developed for color images and achieved PSNR up to 48dB and Molina's system was developed for color image who achieved PSNR up to 44.63dB.

The proposed system achieved PSNR up to 48dB. Also proposed system is blind watermarking system which doesn't require original image content at destination place. The system has capability of providing protection against malicious attacks and it allows non-malicious attacks which are frequently required during transmission. Only Archana's system allows both of these type of protection against attacks but this system was worked only with greyscale image only. With color image, two systems were there but these are not providing security against Non-malicious attacks.

As per the overall comparison, it is clear than the suggested proposed system is better than existing systems.

5. CONCLUSION

The proposed watermarking system is semi fragile based watermarking system which is robust against the non-malicious attacks where as fragile with malicious attacks. The proposed system is also blind watermarking system so it doesn't require the original content at the time of extraction at destination place. The PSNR is achieved up to 43.57 dB which shows that the watermarked image is imperceptible from the human eye where as NCC is near to 1 which shows that it provides more protection against attacks. The tamper detection process can easily classify the watermarked image without support of the original image. If not, a single attack is applied on the image then the image is classified as "Authentic Image". If any attack is applied on the image then it is classified as the type of manipulation. It is classified as "Image is authentic with non-malicious attack" if some unintentional manipulation like Gaussian noise, salt and pepper, JPEG compression, low pass filter, median filter or sharpen image is applied on image. It is classified as "Image is tampered." if image is intentionally manipulated like splicing, copy move, object removal, text addition with or without background. Even the proposed algorithm has capability to identify the tampered region if the image is tampered. In future, authors will try to recover the original image from the tampered one.

REFERENCES

[1] Kommini Chaitanya, Kamalesh Ellanti and E. Harshavardhan Chowdary, "Semi-Fragile Watermarking Scheme based on Feature in DWT Domain", *International Journal of Computer Applications*, Vol. 28, No. 3, pp. 42-46, 2011.

[2] D. Vaishnavi and T.S. Subashini, "Image Tamper Detection based on Edge Image and Chaotic Arnold Map", *Indian Journal of Science and Technology*, Vol. 8, No. 6, pp. 548-555, 2015.

[3] L.V. Lintap, "A Semi-Fragile Watermarking Scheme for Image Tamper Localization and Recovery", *Journal of Theoretical and Applied Information Technology*, Vol. 42, No. 2, pp. 287-293, 2012.

[4] Li Chunlei, "Semi-Fragile Self-Recoverable Watermarking Scheme for Face Image Protection", *Computers and Electrical Engineering*, Vol. 54, pp. 484-493, 2016.

[5] Archana Tiwari and Manisha Sharma, "An Efficient Vector Quantization Based Watermarking Method for Image Integrity Authentication", *Proceedings of International Conference on Progress in Intelligent Computing Techniques*, pp. 215-225, 2017.

[6] Cheonshik Kim, Dongkyoo Shin and Ching-Nung Yang, "Self-Embedding Fragile Watermarking Scheme to Restoration of a Tampered Image using AMBTC", *Personal and Ubiquitous Computing*, Vol. 22, No. 1, pp. 11-22, 2018.

[7] D. Vaishnavi and T.S. Subashini, "Fragile Watermarking Scheme based on Wavelet Edge Features", *Journal of Electrical Engineering and Technology*, Vol. 10, No. 5, pp. 2149-2154, 2015.

[8] Sanjay Rawat and Balasubramanian Raman, "A Chaotic System based Fragile Watermarking Scheme for Image Tamper Detection", *AEU-International Journal of Electronics and Communications*, Vol. 65, No. 10, pp. 840-847, 2011.

[9] Phen Lan, Chung Kai Hsieh, and Po Whei Huang, "A Hierarchical Digital Watermarking Method for Image Tamper Detection and Recovery", *Pattern Recognition*, Vol. 38, No. 12, pp. 2519-2529, 2005.

[10] Rosales-Roldan, "Watermarking-Based Image Authentication with Recovery Capability using Halftoning Technique", *Signal Processing: Image Communication*, Vol. 28, No. 1, pp. 69-83, 2013.

[11] Hanen Rhayma, "Semi Fragile Watermarking Scheme for Image Recovery in Wavelet Domain", *Proceedings of International Conference on Advanced Technologies for Signal and Image Processing*, pp. 1-12, 2018.

[12] B. Feng, Bin, "A Novel Semi-Fragile Digital Watermarking Scheme for Scrambled Image Authentication and Restoration", *Mobile Networks and Applications*, Vol. 25, No. 1, pp. 82-94, 2020.

[13] Paween Pongsomboon, Toshiaki Kondo and Yoshiyuki Kamakura, "An Image Tamper Detection and Recovery Method using Multiple Watermarks", *Proceedings of International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1-13, 2016.

[14] K.R. Chetan and S. Nirmala, "Intelligent Multiple Watermarking Schemes for the Authentication and Tamper Recovery of Information in Document Image", *Proceedings of International Conference on Advanced Computing and Communication Technologies*, pp. 183-193, 2018.

[15] Sajjad Dadkhah, "An Effective SVD-Based Image Tampering Detection and Self-Recovery using Active Watermarking", *Signal Processing: Image Communication*, Vol. 29, No. 10, pp. 1197-1210, 2014.

[16] Behrouz Bolourian Haghighi, Amir Hossein Taherinia and Ahad Harati, "TRLH: Fragile and Blind Dual Watermarking for Image Tamper Detection and Self-Recovery based on Lifting Wavelet Transform and Halftoning Technique", *Journal of Visual Communication and Image Representation*, Vol. 12, No. 2, pp. 1-18, 2017.

[17] Sawiya Kiatpapan and Toshiaki Kondo, "An Image Tamper Detection and Recovery Method based on Self-Embedding Dual Watermarkin", *Proceedings of International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1-8, 2015.

[18] Na Wang and Chung-Hwa Kim, "Tamper Detection and Self-Recovery Algorithm of Color Image based on Robust Embedding of Dual Visual Watermarks using DWT-SVD", *Proceedings of International Conference on Communications and Information Technology,* pp. 1-6, 2009.

[19] Nirali N. Jani and Ashish N. Jani, "Image Segmentation Level Key Driven Image Tampering Detection and Localization Enhancement", *Elixir Digital Processing*, Vol. 115, pp. 49954-49957, 2018.

[20] Javier Molina-Garcia, "An Effective Fragile Watermarking Scheme for Color Image Tampering Detection and Self-Recovery", *Signal Processing: Image Communication*, Vol. 81, pp. 715-725, 2020.

[21] Prabhishek Singh and R.S. Chadha, "A Survey of Digital Watermarking Techniques, Applications and Attacks", *International Journal of Engineering and Innovative Technology*, Vol. 2, No. 9, pp.165-175, 2013.

[22] Vidyasagar M. Potdar, Song Han and Elizabeth Chang, "A Survey of Digital Image Watermarking Techniques", *Proceedings of International Conference on Industrial Informatics*, pp. 175-183, 2005.

[23] A. Zaid and A. Ouled, "Improved QIM-Based Watermarking Integrated to JPEG2000 Coding Scheme", *Signal, Image and Video Processing,* Vol. 3, No. 3, pp. 197-207, 2009.

[24] Hiral A. Patel and Dipti B. Shah, "Digital Image Watermarking Mechanism for Image Authentication, Image Forgery and Self Recovery", *International Journal of Electronics Engineering*, Vol. 11, No. 1, pp. 140-143, 2019.