

A FRAGILE WATERMARKING ALGORITHM FOR IMAGE TAMPERING DETECTION AND CONTENT RECOVERY BASED ON CHAOS

D. Shehwar and M. Iram

Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan
Corresponding author's E-mail: shehwardurre@gmail.com

ABSTRACT: With the exponential advancement in networking technologies and multimedia tools, image authentication has become a major challenge. This paper presents a new fragile watermarking method for tampering detection and content recovery. Using 2 x 2 sized image blocks, eight bit watermark is derived by encoding the low frequency component of the discrete cosine transform (DCT) coefficients of the block. The watermark is embedded in the two least significant bit planes of the cover image. A non-linear chaotic sequence based on Arnold cat map is used for block mapping. Experimental results indicate that the proposed algorithm performs better as compared to the existing state of the art methods.

Keywords: Image authentication, Fragile watermarking, Tamper detection, Chaos.

(Received 17-06-19

Accepted 04-12-19)

INTRODUCTION

Due to rapid advancement in multimedia tools and Internet technology, exchanging, downloading and storing digital information and multimedia data has become easier and frequent (*et al.*, 2016). Even though it is appropriate to send data on the internet in digital form, many security issues have arisen. The easy access of sophisticated and powerful signal and image processing tools has made it easier to spread, duplicate and alter the contents of digital data in illegal means. Therefore, image authentication has become a significant aspect of image processing (*Xia et al.*, 2016, *Qin et al.*, 2016) to ensure the security and integrity of multimedia information. Digital signature based schemes cannot locate and recover the tampered regions (*Qin et al.*, 2016). To address these issues, fragile watermarking is a promising solution for content authentication (*Li et al.*, 2007; *Bravo-Solario et al.*, 2011).

Digital watermarking inserts some information called watermark in the cover image. Applications of watermarking include copyrights protection, broadcast monitoring, content authentication and ownership verification (*Singh and Singh*, 2017).

Digital watermarking is broadly categorized into three classes on the basis of embedding purposes namely robust, semi-fragile and fragile watermarking (*Xiao and Shih*, 2012). A fragile watermark is sensitive and is damaged even if a slight intentional or unintentional modification occurs (*Bravo-Solario et al.*, 2011). A robust watermark can resist manipulations (*Song et al.*, 2012). The third category, semi-fragile watermarking possesses the features of both fragile and robust watermarks (*Preda*, 2013) i.e. it can resist

legitimate distortions while it can detect unauthorized distortions. The prime applications of fragile and semi-fragile watermarking include content authentication, tamper localization and integrity verification. Some fragile watermarking methods can recover the damaged areas to a certain degree. In the scheme put forward by (*Chang et al.*, 2006), the watermark is placed in the least significant bits of the principal pixel of each block after dividing the cover image into blocks of size 3 x 3 pixels. The number of least significant bits is selected on the basis of the type of the correlative block. Although the watermarked images have high quality and any modification in images can be detected and localized using this scheme, there is no recovery option. (*Zhang and Wang*, 2008) presented a novel scheme which can superbly recover the tampered image. A watermark comprising of check-bits and recovery bits is inserted in the cover image in this scheme. However, the condition for error free restoration of original image is that the tampering percentage should be less than 3.2% of the original cover image. *Lee et al.* (2008) suggested a dual watermarking method which not only provided tamper detection but also recovery option. By inserting dual watermarks, the second copy of watermark gives another chance to recover a block in case first copy gets corrupted. However, there are some draw-backs in this method. Firstly, a linear transform is used for block mapping which is not reliable enough. Secondly, the host image is not protected from content-only attack and collage attack. Lastly, the embedding payload is increased.

Hsu et al. (2010) used the concept of probability in watermarking to raise detection precision. This scheme performs better in terms of detection precision. However, it does not provide any recovery

option. Zhang *et al.* (2010) put forward two watermarking methods based on a reference sharing mechanism. In first method, the principal data of the cover image is used to acquire the watermark. If tampering rate is below 24%, the original image can be restored. In other method, the cover image is disintegrated into three levels using a hierarchical mechanism. Since dissimilar restoration capabilities are employed in this technique, therefore, the second technique performs well in terms of recovery capability. (Chang *et al.* 2011) proposed a method to protect images using a novel hybrid scheme based on logistic map and Hamming code. The scheme presented by (Tong *et al.* 2013) is based on spatial domain. The watermark for each 2×2 block is generated by using block characteristics and embedded in the mapping block. Although the scheme works well for tamper detection but the usage of three least significant bit planes for embedding reduces the quality of watermarked image. Zhang *et al.* (2015) presented a watermarking method with improved restoration capability is presented. The original cover image is compressed by using a procedure for categorizing the blocks based on the DCT coefficients. The quality of the recovered images is good as demonstrated by the experimental results.

The watermarking technique put forward by (Ansari *et al.*, 2016) is based on singular values of an image for effective tamper detection. The singular values are computed for each image block which are later used for tamper localization. The average recovery of this scheme is 28 dB for a tampering rate of 50 %. Ali *et al.* (2014) proposed a watermarking approach based on block truncation coding. The tampering detection capability of this scheme is poor plus the restoration capability was not satisfactory. Kiatpapan *et al.* (2015) put forward a dual watermarking scheme in which two similar watermarks are placed in the cover image. The quality of watermarked and recovered images is good in this scheme but it increases blocky artifacts. Hsu *et al.* (2016) exploited the concept of smoothness to discern various types of blocks. Hence, the efficiency of information hiding is increased by effectively placing the watermark in the cover image. A DCT based image authentication scheme is proposed by Singh and Singh (2016). After dividing the cover image into fixed size blocks, a watermark composed of authentication and recovery bits is generated for all blocks. Although the scheme has high quality of recovered image along with exceptional tamper detection accuracy, but watermarked image quality is decreased due to usage of three least significant bit planes. Fang *et al.* (2017) presented a watermarking technique with hierarchical recovery. The main idea behind this scheme is to assign a higher recovery priority to higher MSB layers. The quality of recovered

image is enhanced considerably in this scheme. Wang *et al.* (2018) presented an alterable capacity watermarking method based on SVD transform. After splitting the host image into 2×2 blocks and applying SVD on all blocks, the image blocks are categorized as smooth and textured blocks. The length of recovery watermark depends on the type of block. A watermarking scheme based on hash function for image authentication is suggested by (Gul and Ozturk, 2019). Although this scheme detects all tampered regions, it has no recovery capability, because hash function is irreversible.

As mentioned above, previous fragile methods watermarking approaches have low qualities of watermarked and recovered images. Although these schemes can detect the tampered areas, but the localization is weak and reconstruction quality of the main content is just acceptable. Hence, it is desired to develop a watermarking method that can maintain high quality of watermarked images and can recover the tampered image.

To increase security and image quality, a watermarking algorithm based on chaos is presented. To recover the tampered area, the watermark needs to contain the information of original cover image. DCT is used to generate the watermark. DCT is an image compression technique having low memory requirement and less computational complexity (Britanak *et al.*, 2007). A non-linear block mapping based on Arnold cat map is proposed to increase the percentage of data recovery. A chaotic sequence is used to produce the block mapping (Ali *et al.*, 2014). Chaotic systems are governed by nonlinear dynamics and show deterministic behavior that is highly sensitive to change in initial conditions. The outcomes of such systems seem random and are uncorrelated.

In general, the novelties of the proposed algorithm include using DCT for watermark derivation and Arnold map for block mapping sequence generation. The quality of watermark generated using DCT is high. Moreover, usage of small 2×2 sized blocks increases the localization accuracy and removes blocky artifacts.

MATERIALS AND METHODS

The proposed algorithm comprises of three main phases i.e. watermark derivation and embedding, tamper detection and content recovery procedures. A description of the main phases of the proposed method is given below.

Watermark derivation and embedding: Assume that the cover image has m rows and n columns. Let N denote the image size ($N = m \times n$). After dividing the image into 2×2 size blocks, a vector of 8 bits is

generated for each block. The watermark is derived using the six most significant bit planes of the original cover image while the two least significant bit planes are used to accommodate the watermark. The watermark consists of two types of bits i.e. authentication bits and recovery bits. DCT is used to generate watermark recovery bits. Then, the authentication bits are derived by computing the parity check bits of the recovery bits.

The watermark of each 2 x 2 block is then embedded in the two least significant bit planes of cover image. The details of this phase are described below.

3.1.1 Derivation of recovery bits: The purpose of recovery bits is to restore the content of a tampered region. The recovery bits are derived from the content of the image itself. To generate the recovery mark, one level DCT is applied on each block of host image. The LL band coefficients are modified using Algorithm 1. The advantage of using this proposed algorithm is to decrease the difference between two values, thereby enhancing the quality of watermarked image. Suppose that the value of LL band coefficient is $331_{10}(101001011_2)$. Using the proposed method, the 2 LSBs of coefficient are ignored and replaced with zero in recovery stage. Without using this technique, the value of coefficient will be $328_{10}(101001000_2)$. Therefore, the difference between coefficients is equal to three. The LSBs are adjusted in this algorithm in such a way that the value of coefficient becomes $332_{10}(101001100_2)$, reducing the difference between the original coefficient and the adjusted watermark to one. Ultimately, 6 MSBs of the adjusted coefficient as kept are recovery bits while the three LSBs are ignored.

For each 2 x 2 block, six recovery bits are considered. The two LSBs of the host image I are replaced with zero, that is

$$I' = I \text{ and } (11111100)_2 \quad (1)$$

- i. Block Division: Divide the image I' into N non-overlapping 2 x 2 blocks. X_i is the i-th block which is denoted as:

$$X_i = \begin{bmatrix} x_{i1} & x_{i3} \\ x_{i2} & x_{i4} \end{bmatrix}, i=1, 2, \dots, N \quad (2)$$

- ii. Discrete Cosine Transform: L_i is the DCT coefficient of each block X_i .

$$L_i = \text{DCT2}(X)_i = \begin{bmatrix} L_{i1} & L_{i3} \\ L_{i2} & L_{i4} \end{bmatrix}, \quad (3)$$

- iii. The 6 recovery bits are obtained by

$$r_j = \sqrt{L_{i1}} / 2^{6-j} \text{ mod } 2, j=1, 2, \dots, 6 \quad (4)$$

Algorithm 1. Generation of Recovery bits

Input: LL band coefficients of nth block pixels of cover image

Output: Recovery mark of 6 bits

```

1:  procedure Recovery mark generation, where
    coeff illustrates the LL band DCT coefficients
2:  var=0
3:  coeffout = 0
4:  if LLcoef < 0 then
5:  coeff = abs(LLcoef)
6:  var = 1
7:  coeff = coeff + 2
8:  coeffout = bitand(coeff, 252)
9:  end if
10: if var = 1 then.
    
```

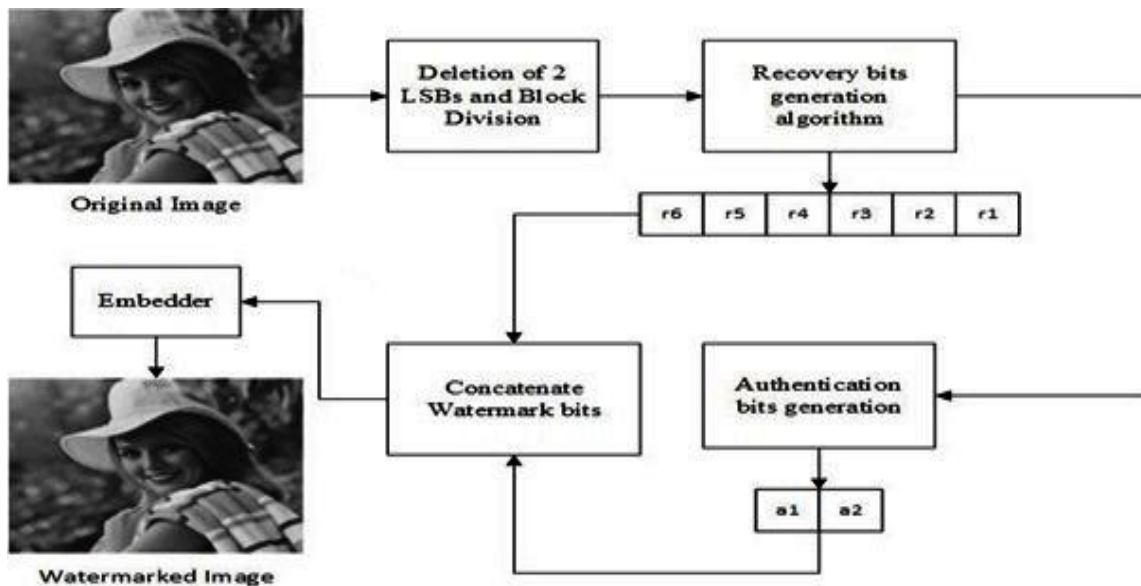


Figure-1. Watermark embedding procedure.

```

11:   coeffout = -coeffout
12:   end if
13:   Recoverymark = coeffout
14:   return Recoverymark
15:   end procedure
    
```

Derivation of authentication bits: The authentication bits are derived from the recovery bits and are used for tampering detection. Two authentication bits are considered for authenticating each 2 x 2 block. The authentication bits a1 and a2 are derived as below:

$$a1 = r_1 \oplus r_1 \oplus r_2 \oplus r_3 \oplus r_4 \oplus r_5 \oplus r_6 \quad (5)$$

where, \oplus represents the exclusive OR operation.

The value of a1 can either be 1 or 0. Another flag bit a2 is generated to pad the 8-bits section. a1 and a2 both are used for tampering detection.

$$a2 = \begin{cases} 1 & \text{if } a1 = 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

The eight bits generated for each block consists of 6 bit recovery information and 2 authentication bits (total bits are 6+2=8).

Watermark Embedding: In this step, the 8-bit watermark W_i is placed in the two least significant bit planes of each block. Fig. 1 shows the procedure of watermark embedding.

Firstly, the two authentication bits are placed in each block X_i to obtain the watermarked block $X_{i(w)}$ on the basis of following equation.

$$x_{ij} = x_{ij} + 2W_{i(j+3)} + W_{i(j+4)}, j=4 \quad (7)$$

Then, the recovery bits are embedded in the two least significant bit planes of the corresponding mapping block X_i , as explained by Eq. 8:

$$x_{ij} = x_{ij} + 2W_{ij} + W_{i(j+3)}, j=1,2,3 \quad (8)$$

Fig. 2 depicts the visual representation of a watermarked image block where p0, p1, p2 and p3 represent four pixels of a 2 x 2 image block and each row denotes eight bits b7, b6, ..., b0 of a pixel.

	b7	b6	b5	b4	b3	b2	b1	b0
p0							r6	r3
p1							r5	r2
p2							r4	r1
p3							a1	a2

Fig-2. Visual representation of a watermarked image block.

Tamper detection: The integrity of the watermarked image will be tested whether it has been manipulated intentionally or by incidental modifications. The

extracted and calculated authentication bits are compared to determine the authenticity of a block.

Assume J is the image received by the receiver which may be tampered or untampered. At the receiver side, following steps are performed for tamper detection.

Step1 Divide the received image into 2 x 2 sized non-overlapping blocks. Label all blocks as valid blocks.

Step2 Extract the 8 bit watermark from all image blocks.

Step3 Recalculate the authentication bits using the six most significant bit planes of each 2 x 2 block.

Step4 If generated authentication bits of a block are exactly equal to the authentication bits extracted from that block, then label the block as a valid block.

Step5 If there is a discrepancy between the retrieved authentication bits and recalculated authentication bits, then label the block as an invalid block.

Step6 Set the intensity value of the invalid blocks to maximum. The locations of tampered blocks are represented by a binary matrix $loc = \{loc_i | i = 1, 2 \dots N\}$ as shown in the following formula. G_{a1} and G_{a2} denote the recalculated authentication bits and E_{a1} and E_{a2} represent the extracted bits.

$$loc_i = \begin{cases} 0 & G_{a1} = E_{a1}, G_{a2} = E_{a2} \\ 1 & \text{otherwise} \end{cases} \quad (9)$$

Self-Recovery Scheme: After tampering detection process, all the blocks are judged as tampered or untampered. If a block is judged as untampered, then there is no need of further processing. And a block declared tampered need to pass through the recovery module. The recovery process is explained below.

Step1 Traverse all the blocks of the tampered image. If a block is labelled as a valid block, then move to the next block else go to Step (2).

Step2 For an invalid block B, find its mapping block.
Step3 If the mapping block C is a valid block, then retrieve the 6 recovery bits from block C using equation 10 and obtain the discrete cosine transform coefficients.

$$l_c^* = [\sum_{j=1}^6 r_j^* \times 2^{6-j}]^2 \quad (10)$$

Step4 Pad three zeros at the end of the six bits of LL band coefficient and apply one-level inverse DCT to the resulting bands. The matrix acquired by IDCT is shown below.

$$L_c^* = \text{IDCT2} \begin{bmatrix} l_c^* & 0 \\ 0 & 0 \end{bmatrix} \quad (11)$$

Let L_c^* substitute the block B. Label block B as a valid block.

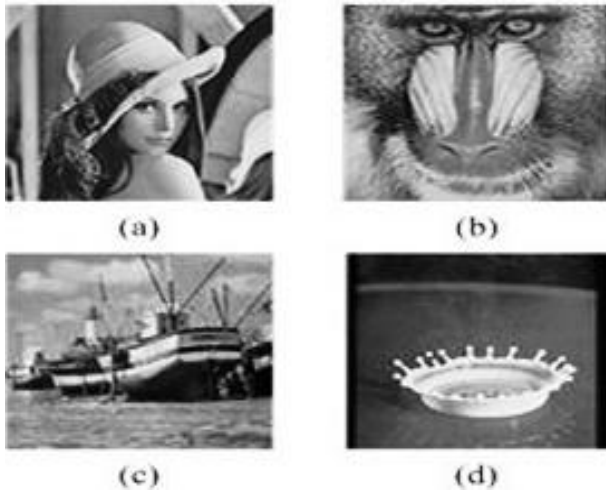
Step5 If the mapping block C is tampered, then recover the remaining erroneous blocks by substituting the mean value of eight valid neighbor blocks surrounding block B.

Step6 Label the block B as a valid block after its complete restoration.

Step7 Combine all the blocks to form the recovered image.

RESULTS AND DISCUSSIONS

The proposed scheme used grayscale images as original cover image as illustrated in Fig. 3. The test images are chosen from the MISC database. The programming environment for testing experimental results is MATLAB R2018 a. Experimental results of the proposed algorithm are discussed and the evaluation is performed in terms of essential properties of imperceptibility security, fragility and recovered image quality.



Figur-3. Four test images (a) Lena, (b) Baboon, (c) Boat, (d) Splash.

Perceptual Quality: Perceptual quality refers to the quality of the image after inserting a watermark in it. Good perceptual quality indicates that the watermarked image is almost similar to the original image. The perceptual quality is gauged using the objective metric peak signal to noise ratio (PSNR). The mathematical formula of PSNR is given below

$$PSNR = 10 \times \log_{10} (255^2 / MSE) \quad (12)$$

The mean square error (MSE) is defined as

$$MSE = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n (e(i, j))^2 \quad (13)$$

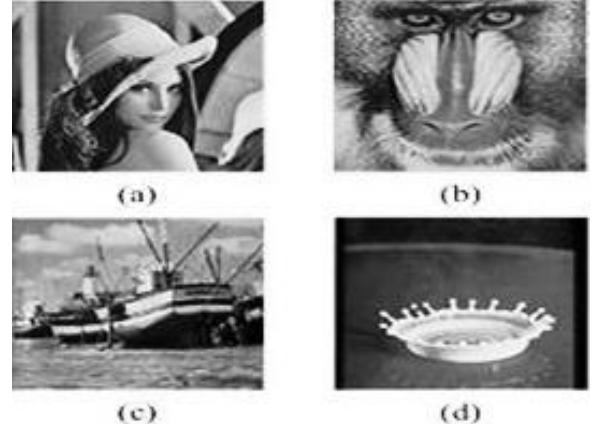


Figure-4. Images with watermark (a) Lena, $PSNR_w=44.14$, (b) Baboon, $PSNR_w=44.16$ (c) Boat, $PSNR_w= 44.17$ (d) Splash, $PSNR_w=44.25$

Table-1: PSNR (dB) of watermarked images of proposed method and related works.

	Proposed	Tong <i>et al.</i> , 2013	Hsu <i>et al.</i> , 2016	Chang <i>et al.</i> , 2011
Cover Image	$PSNR_{(w)}$	$PSNR_{(w)}$	$PSNR_{(w)}$	$PSNR_{(w)}$
Lena	46.26	40.73	40.95	37.762
Baboon	46.28	40.71	40.92	37.874
Sailboat	46.26	40.58	40.99	37.332
Peppers	46.33	40.67	40.92	37.858

Fig. 4 shows the watermarked versions of the original test images. The PSNR value after watermark embedding is greater than 46 dB indicating that visual distortions are imperceptible. The PSNR values of images using this scheme and some related works are listed in Table 1. The embedding PSNR values using the proposed approach are higher, with an average value of 46.28 dB, which is greater than that of (Chang *et al.*, 2011, Tong *et al.*, 2013, Hsu *et al.*, 2016; Singh and Singh, 2016).

Performance of tamper localization: To evaluate the performance of tamper localization, Photoshop was used to modify watermarked images to make forgery images. General tampering attacks such as image cropping, object addition, content removal, copy-move and text addition as applied on watermarked images. The images after various attacks are illustrated in Fig.5. The results of tamper detection are illustrated in Fig.6 which depicts that the proposed algorithm can accurately localize the tampered regions.

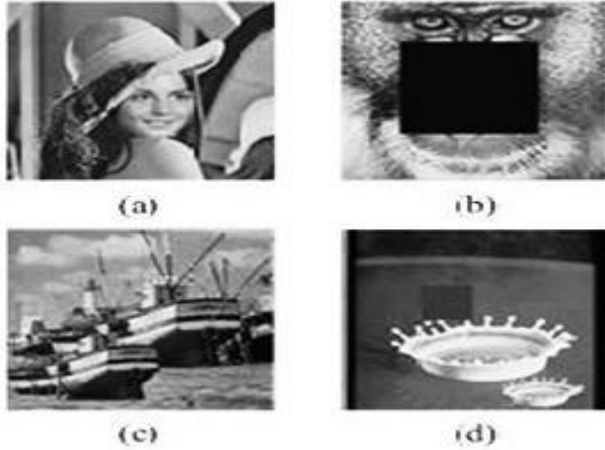


Figure- 5. Tampered Images (a) Lena, Content modification attack (b) Baboon, Content removal i]attack(c) Boat, Object addition attack (d) Splash, Copy-move attack.

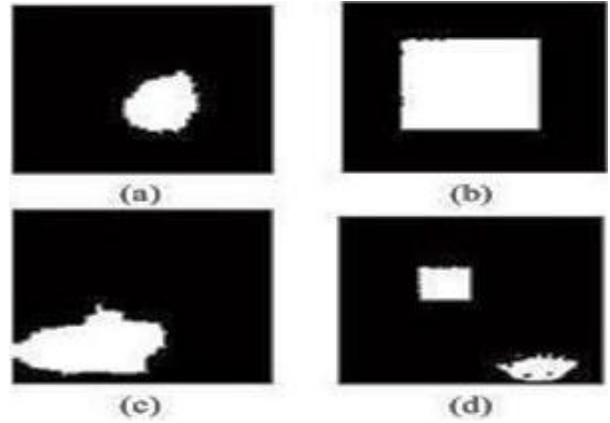


Figure -6. Tamper detection results for Fig. 11(a-d).

Performance of tamper recovery: Fig. 7 shows tampered “Sailboat” images after applying different intensities of content removal attacks. The recovered images are shown in Fig. 8.



Figure-7. Tampered Sailboat at various tampering percentages



Figure-8. Recovered Images of Fig. 7

Also, cropping attacks are also applied on other images. The PSNR values of recovered images

are listed in Table 2.



Figure-9. Tampered Elaine at various tampering percentages.



Figure-10. Recovered Images of Fig. 9.

It is observed that the PSNR value of recovered images is satisfactory. Even if the tampering percentage is 50%, the PSNR values of recovered images are higher than 32.16 dB. So, the quality of recovered images is rather satisfactory using the proposed method.

Table-2. PSNR (dB) values of recovered images for different tampering rates.

Cover Image	Tampering Rates				
	10	20	30	40	50
Lena	40.39	36.49	34.70	33.32	32.16
Splash	38.25	35.36	33.61	32.57	31.24
Elaine	39.59	36.77	34.79	33.53	32.79
Peppers	40.75	36.36	34.67	33.37	32.53
Baboon	40.04	36.51	34.45	32.76	31.86

Conclusion: In this article, a watermarking algorithm is proposed to ensure the security of digital images. Results of experiment reveal that the performance of suggested method is improved than several other similar methods in terms of perceptual invisibility and is efficient in execution time because of using simple parity check operations and comparisons.

REFERENCES

- Ansari, I.A., M. Pant and C.W. Ahn (2016). SVD based fragile watermarking scheme for tamper localization and self-recovery. *Int. J. Mach. Learn. Cyb.* 7(6): 1225-1239.
- Ali, M., C.W. Ahn and M. Pant (2014). A robust image watermarking technique using SVD and differential evolution in DCT domain. *Optik-Int. J. Light Elect. Optics.* 125(1): 428-434.

- Britanak, V., P. Yip and K.R. Rao (2007), Discrete Cosine and Sine Transforms. New York: Academic.
- Bravo-Solorio, S. and A.K. Nandi, (2011). Secure fragile watermarking method for image authentication with improved tampering localization and self-recovery capabilities. Sig. Proc. (4): 728-739.
- Chang, C., K.N. Chen, C.F. Lee and L.J. Liu (2011). A secure fragile watermarking scheme based on chaos-and-hamming code. J. Sys. Software. 84(9): 1462-1470.
- Chang, C.C., Y.S. Hu and T.C. Lu (2006). A watermarking-based image ownership and tampering authentication scheme. Patt. Recog. Lett. 27(5): 439-446.
- Fang, C., B. An, J. Wang, D. Ye and H. Wang (2017). Hierarchical recovery for tampered images based on watermark self-embedding. Displays. 46: 52-60.
- Gul, E. and S. Ozturk (2019). A novel hash function based fragile watermarking method for image integrity. Multimedia Tools Apps: 1-18.
- Hsu, C.S. and S.F. Tu (2016). Image tamper detection and recovery using adaptive embedding rules. Meas. 88: 287-296.
- Hsu, C.S. and S.F. Tu (2010). Probability-based tampering detection scheme for digital images. Optics Comm. 283(9):1737-1743.
- Shapiro, J.M. (1993). "Embedded image coding using zero trees of wavelet coefficients" IEEE Trans. S 41(12) :3455-3462.
- Kiatpapan, S. and T. Kondo (2015). An image tamper detection and recovery method based on self-embedding dual watermarking in Electrical Engg./Electronics, Comp., Telecomm. Info.Tech. (ECTI-CON), 12th Int. Conf. IEEE.
- Lee, T.Y. and S.D. Lin (2008). Dual watermark for image tamper detection and recovery. Patt. recog. 41(11): 3497-3506.
- Li, Q. and N. Memon (2007). Security models of digital watermarking. Multimedia Content Analysis Mining: 60-64.
- Peng, F., R.S. Guo, C.T. Li and M. Long (2010). A semi-fragile watermarking algorithm for authenticating 2D CAD engineering graphics based on log-polar transformation. Computer-Aided Design. 42(12): 1207-1216.
- Preda, R.O. (2013). Semi-fragile watermarking for image authentication with sensitive tamper localization in the wavelet domain. Meas. 46(1): 367-373.
- Qin, C., X. Chen, D. Ye, J. Wang and X. Sun (2016). A novel image hashing scheme with perceptual robustness using block truncation coding. Info. Sci. 361: 84-99.
- Singh, D. and S.K. Singh (2017). DCT based efficient fragile watermarking scheme for image authentication and restoration. Multimedia Tools Apps. 76(1): 953-977.
- Singh, D. and S.K. Singh (2016), Effective self-embedding watermarking scheme for image tampered detection and localization with recovery capability. J. Visual Comm. Img. Represent. 38: 775-789.
- Song, C., S. Sudirman and M. Merabti (2012). A robust region-adaptive dual image watermarking technique. Journal of Visual Communication and Image Represent. 23(3): 549-568.
- Song, C., S. Sudirman and M. Merabti (2009). Recent advances and classification of watermarking techniques in digital images. in Proceedings of post graduate network symposium.
- Tong, X., Y. Liu, M. Zhang and Y. Chen (2013). A novel chaos-based fragile watermarking for image tampering detection and self-recovery. Sig. Proc.: Image Comm. 28(3): 301-308.
- Wang, C., H. Zhang and X. Zhou (2018). A self-recovery fragile image watermarking with variable watermark capacity. Appl. Sci. 8(4):548.
- Wu, C.M., Y.C. Hu, K.Y. Liu and J.C. Chuang (2014). A novel active image authentication scheme for block truncation coding. Int. J. Signal Process. Image Process. Patt. Recog. 7(5): 13-26.
- Xia, Z., X. Wang, X. Sun, Q. Liu and N. Xiong (2016). Steganalysis of LSB matching using differences between nonadjacent pixels. Multimedia Tools Apps. 75(4): 1947-1962.
- Xiao, D. and F.Y. Shih (2012). An improved hierarchical fragile watermarking scheme using chaotic sequence sorting and subblock post-processing. Optics Comm. 285(10): 2596-2606.
- Zhang, X. and S. Wang (2008). Fragile watermarking with error-free restoration capability. IEEE Transac. Multimedia. 10(8): 1490-1499.
- Zhang, X., S. Wang, Z. Qian and G. Feng (2010). Reference sharing mechanism for watermark self-embedding. IEEE Transac. Image Proc. 20(2): 485-495.
- Zhang, X., Y. Xiao and Z. Zhao (2015). Self-embedding fragile watermarking based on DCT and fast fractal coding. Multimedia Tools Apps. 74(15): 5767-5786.
- Usc-sipi image database-miscellaneous. URL <http://sipi.usc.edu/database/database.php?volume=misc>.