Journal of Engineering and Applied Sciences 12 (6): 1582-1589, 2017
ISSN: 1816-949X
© Medwell Journals, 2017

A Semi-Fragile Watermarking Based Image Authentication

Suhad Ahmed Ali, Majid Jabbar Jawad and Mohammed Abdullah Naser
Department of Computer Science, College of Science for Women, Babylon University, Babylon, Iraq

Abstract: This study has suggested a semi fragile watermarking technique for proving the authenticity of image. The suggested technique is also used for detecting any tampering may be done on the image. The suggested technique has extracted the feature distribution (digital signature) image from the cover image to be as a watermark. The digital signature image represents the important edges which are extracted from the coefficients of a Discrete Wavelet Transform (DWT). Later, the created signature (watermark) is embedded in the cover itself. The suggested scheme could decide whether the image is tampered or not in addition to localize the tampered region. After applying the suggested technique, the results had shown that it could detect and localize detect and localize any tampered region in image. In addition, the suggested technique could decide whether the tampered region is attacked intentionally or unintentionally.

Key words: Digital still image, discrete wavelet transform, digital watermarking, semi-fragile watermarking, image authentication

INTRODUCTION

The digital images are used in several purposes such as indicator or evidence in some cases (car accidents, politician's scandals, medicals and others) (Cruz *et al*., 2009; Manickam *et al*., 2013). Due to availability of improved digital processing techniques, anyone can make unauthorized changing on an image in order to reuse it in other applications. Accordingly, the necessity of ensuring the authenticity and integrity of digital images is required. The authentication systems can be classified into two types: the first one is based on digital signature while the second one is based on digital watermark (Tiwari and Sharma, 2012). A digital signature either can be an encrypted value or signed hash value of image contents and/or image properties. The drawback of signature based methods is that they can delimited whether a digital image has been modified or not but they cannot localize the modified region. To overcome this problem, digital watermarking based methods has been suggested.

Digital watermarking technique means embedding digital data called watermark in digital content such audio, images, video and text called cover or host. It has introduced as a powerful tool for increasing the security (Sumalatha *et al*., 2012).

Now a days, digital watermarking appears as an efficient tool to ensure integrity and authenticity verification. According to watermarking properties, watermarking techniques can be classified into several types namely: robust, fragile and semi-fragile (Shukla and Mehta, 2015). A robust watermark can be exploited for protecting copyright of digital product while a fragile and semi fragile watermark can be exploited for confirming the authenticity of digital product (Lintao *et al*., 2012). Semi fragile watermarking combines the characteristics of fragile and robust watermarking. Semi fragile watermarking is robust against unintentional modification while fragile against intentional modification.

This research suggests a semi fragile watermarking technique for detecting any modification may be done on the image (intentional or unintentional) and localize the modified regions.

Literature review: Herein, some proposed methods related to our suggested method. Clara have introduced a semi fragile watermarking scheme for doing some activities such as detecting image integrity, detecting areas suffered some modifications of content and restoring the tampered areas. In the suggested scheme, the digital image is divided into two regions namely Region of Interest (ROI) and Region of Embedding (ROE). The first one is important region which must be protcted against malicious modification. The second one is the rest of the image where in which the watermark is embedded. After applying DCT, the watermark is created from the coefficients of ROI and embedded in the DCT coefficients ROE. The suggested scheme is checked according several criteria such as imperceptibility, quality of restored regions robustness against unintentional manipulations (such as JPEG type) and accuracy of tampered area detection.

Coressponding Author: Suhad Ahmed Ali, Department of Computer Science, College of Science for Women, Babylon University, Babylon, Iraq

In order to check the authenticity of an image Qi *et al*. (2009) proposed a semi fragile watermarking technique. This technique can detect any intentional and unintentional tampering may be done on image successfully. In addition the suggested technique localize the tampered regions. In this method, two complementary watermarks are generated to enhance the authentication process. These watermarks are created based on the content of the cover in order to localize the tampered regions. The watermarks are created from DWT coefficients of the cover image and also embedded in the DWT coefficients of the cover.

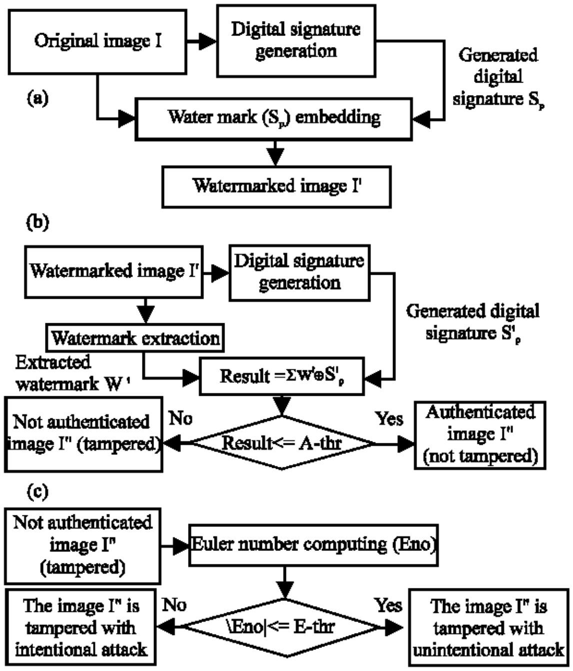
Preda *et al*. (2015) have introduced a semi fragile watermarking scheme which can detect the tampered regions in images and restore amount of the original content of the modified regions. The proposed system used two different watermarks. The first one is a semi-fragile watermark which is used for image authentication. The second one is a robust watermark which is used for image restoration. These watermarks are generated from the DWT domain of the host itself and also embedded in the DWT domain. The results illustrated that the proposed method achieves good results such as the image quality and identifies image tampering of up to 20% of the original image.

Lu *et al*. (2015) have suggested a semi fragile watermarking scheme for checking the authenticity of medical image. In this scheme, a watermark is generated from tag information and embedded into the cover(image). In this scheme, two procedures is be done. The first one is dividing the image into two regions namely, Region of Interest and (ROI) and Non-Interest (RONI). The second procedure is applying 2-level DWT is on the ROI. The low-frequency sub band (LL2) of the DWT is exploited as a host for hiding the watermark. This scheme is very important in the medical applications.

MATERIALS AND METHODS

The suggested method: The suggested method is illustrated in Fig. 1. The procedures of the suggested method is listed as follows.

The generation of digital signature: Figure 2 depicts the block diagram of digital signature generation.



(a)
Original image I
Digital signature generation
Generated digital signature S_p
Water mark (S_p) embedding
Watermarked image I'
(b)
Watermarked image I'
Digital signature generation
Generated digital signature S'_p
Watermark extraction
Extracted watermark W '
Result =Σw'⊕S'$_p$
Not authenticated image I'' (tampered)
No
Result<= A-thr
Yes
Authenticated image I'' (not tampered)
(c)
Not authenticated image I'' (tampered)
Euler number computing (Eno)
The image I'' is tampered with intentional attack
No
\Eno|<= E-thr
Yes
The image I'' is tampered with unintentional attack

Fig. 1: Steps of the suggested system; a) The watermark insertion process; b) The authenticity process and c) The verification process

LLI	HLI
LHI	HHI

Fig. 2: One level DWT decomposition

Discrete Wavelet Transform (DWT): The DWT has been presented as a highly efficient and flexible method for sub band decomposition of signals. In the frequency domain; it decomposes a signal depending on a family of basis functions obtained through translation and dilation of a mother wavelet. When one dimensional DWT is applied on an image, an image is decomposed into two parts which is coarse and detail elements by using Low-pass (L) and High-pass (H) filters. Image signal is decomposed into four sub-bands when two dimensional DWT is applied on rows at first and then on columns of an image. These subbands are approximation sub-band (LL) and details subbands (LH, HL and HH). For applying multilevel DWT; the approximation subband is used for next level of decomposition. Figure 3 shows single level decomposition of DWT.

Generation of digital signature from wavelet transform: A digital signature is generated from the original image using wavelet transform; therefore, a technique based on edge detection will be applied. A wavelet edge pixel E(i, j)

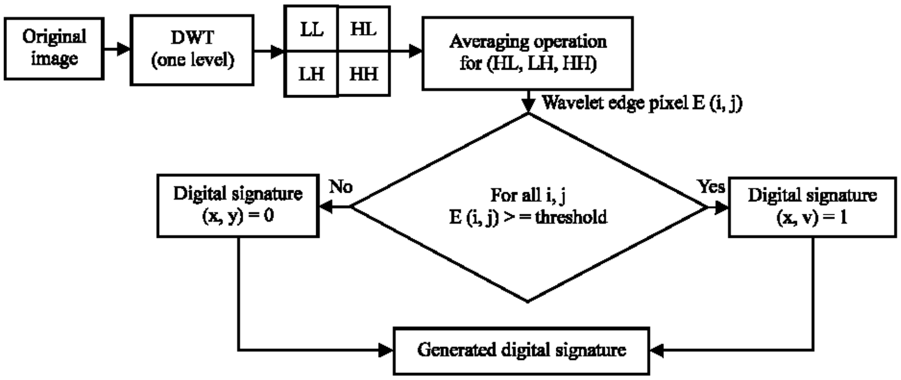

Original image
DWT (one level)
LL
HL
LH
HH
Averaging operation for (HL, LH, HH)
Wavelet edge pixel E (i, j)
For all i, j
E (i, j) > = threshold
No
Yes
Digital signature (x, y) = 0
Digital signature (x, v) = 1
Generated digital signature

Fig. 3: Digital signature generation

at pixel in location i and j can be define by taking the average of corresponding pixels in three (details) high frequency subbands (LH, HL and HH) according to Eq. 1:

$$E(i,j) = \sum_{n=1}^{3} (D_n(i,j))^2/3 \qquad (1)$$

where, E(i, j), n and D_n (i, j) are wavelet edge, number of high frequency subbands (details) high frequency subbands, respectively. The digital signatures is generated according to Eq. 2:

$$S_p(i,j) = \begin{cases} 1 & E(i,j) \geq \alpha \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

Where:

$S_p(i, j)$ = Generated digital signature of pixels
α = A threshold value

The value of α is determined depending on the statistical measurements of S_p (i, j) as follows:

$$\alpha = m + k \times \sigma \qquad (3)$$

where, m, σ and are mean, standard deviation of candidate original image pixels array and parameter, respectively. The value of k depends on the mean and standard deviation values. Figure 4a-d shows examples of the generated signatures.

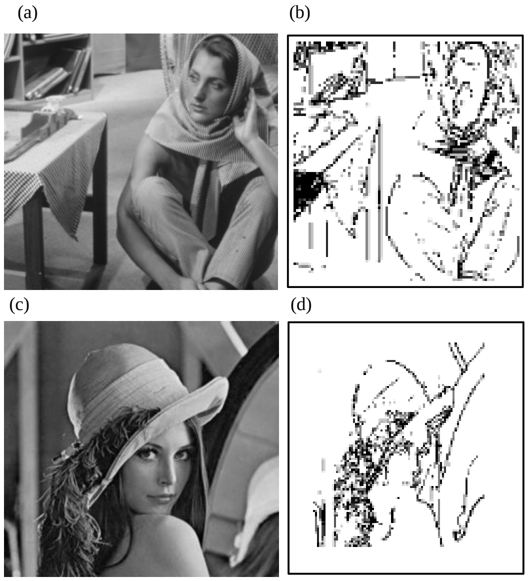


Fig. 4: a, c) Cover imagesl and b, d), generated signatures

Watermark embedding: After implementing the procedure of section (3.1.2), the embedding procedure is being done. The detail of the embedding procedure is illustrated in Algorithm 1.

Algorithm 1 (The embedding algorithm):

```
Input: I, W
(I: cover image, W: generated digital signature (binary image which is generated in section (3.1.2))
Output: I'
    (I': watermarked image)
    Dividing the cover (I) into blocks of size 2×2
    for n = 1 to no. of blocks
        Compute the one level decomposition (the result is LL, HL, LH and HH)
        Making quantization procedure according to the following equation:
            $q(n) = \left[\frac{LL(n)}{Qstep}\right]$    (4)
        if mod(q(n), 2) = W(n) then
            $LL(n) = (q(n) * qstep) + \left(\frac{qstep}{3}\right)$    (5)
        Else
            $LL(n) = ((q(n)+1)) * qstep + \left(\frac{qstep}{3}\right)$    (6)
            {adjust the quantize low-frequency wavelet coefficient}
        endif
        applying inverse wavelet transform on subbands (LL, HL, HI, HH) on block
            {to get reconstructed block)
End for
Reconstructing the watermarked image from reconstructed blocks (I')
```

Watermark extraction: The embedding and extraction processes are reverse. Algorithm 2 shows watermark extraction.

Algorithm 2 (The extracting algorithm):

```
Input: I'
    (I': watermarked image))
Output: W'
    (W': extracted watermark)
    divide the watermarked image I' into non overlapping blocks of size 2×2.
    for n = 1 to no. of blocks
      Compute the one level decomposition (the result is LL', HL', LH' and HH')
      Making quantization procedure according to the following equation:
```

$$q'(n)=\left[\frac{LL'(n)}{Qstep}\right] \qquad (7)$$

```
      extract the embedding signature according to the following equation
```

$$sim(k)=mod(q'(i,j),2) \qquad (8)$$

```
  {Where i = 1, …, m/2, j = 1, ……., n/2, k = 1, ….., q×q}
  convert sim into two dimensional extracted watermark image W'.
End for
Reconstructing the watermark image from reconstructed blocks (W')
```

Authentication process: In this process the differences between extracted watermark W' and generated digital signature S'_p is determined according to the following steps.

Algorithm 3 (The authentication algorithm):

```
Input: W', S'p
    (W': extracted watermark image, S'p: generated signature)
Output: Decision
    generate watermark S'p from watermarked image I' as described in section (3.1.2)
    extract watermark W' using algorithm 2.
    divide each of extracted watermark W' and digital signature S'p into 2×2 block.
    apply Xor logical operation between each extracted watermark W' and digital signature
    blocks S'p to determine whether the block is tampered or not.
    compute the number of non-match bits (nbits) between each extracted watermark W' and
  generated blocks S'p according to Eq. 9:
```

$$nbits=\sum W\oplus S'p \qquad (9)$$

```
    compare the value of nbits with threshold (thr).
{Note: The value of threshold is set to 2 in all experiments and is determine through trial and errors}.
```

The authentication process:

```
    If nbits≤thr then
        The block is not tampered (authenticated)
    Else
        The block is tampered (not authenticated)
```

Verification process: This process is used for checking whether the errors occurred in watermarked image are resultant of intentional attack or unintentional attacks. The difference between the extracted watermark W' and generated watermarked S'_p is computed is done to illustrate whether the attack is intentional or unintentional. The Euler number Eno is computed in order to measure the distributed of errors pixels in watermarked image. The computation Euler is deepened on pixels connectivity. The 8-connective is depended in this study. The Euler number represents the total number of objects in the image minus the total number of holes in those objects. This number is checked with threshold to detect whether the block is authenticated or tampered using Eq. 10:

$$\text{Image is tampered with}\begin{Bmatrix}\text{Unintenional attack if}\,|Eno|\le thr\\ \text{Intenional attack if}\,|Eno|\le thr\end{Bmatrix} \qquad (10)$$

Benchmarking criteria: Several benchmarking criteria have been used, for testing the efficiency of the suggested scheme. These criteria are listed as follows.

Peak Signal to Noise Ratio (PSNR): The PSNR measures the fidelity of the watermarked image by comparing it with the original image (Huynh-Thu and Ghanbari, 2008). The PSNR is defined as:

$$PSNR=10\log_{10}\left[\frac{(1_{MAX})^2}{MSE}\right] \qquad (11)$$

where, IMAX is the maximum grey level of the image. In this case, the value of IMAX is up to 255. MSE is stand for Mean Square Error and can be defined in Eq. 12:

$$MSE=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}(p_1(i,j)-(p_2(i,j))^2 \qquad (12)$$

where, p_1 (i, j) and p_2 (i, j) represent two images, m, n represent the dimensions of two images.

Structural Similarity Index Measure (SSIM): The Structural Similarity Index Measure (SSIM) is used for evaluate the difference between the cover image and the watermarked image. The evaluation of the difference consists of comparisons namely; the luminance of each image is compared. The contrast of each image is compared and the structure is to be compared. So, SSIM algorithm is dealing with three components: luminance, contrast and structure independently. The value of SSIM function is computed according to Eq. 10 which proposed by Wang *et al.* (2004):

$$SSIM=\frac{(2\mu_x u_y+c_1)(2\sigma_{xy}+c_2)}{\left(\mu_x^2\mu_y^2+c_1\right)\left(\sigma_x^2\sigma_y^2+c_2\right)} \qquad (13)$$

where, μ, σ and σ_{xy} are mean, variance and covariance of the images and c_1, c_2 are the stabilizing constants. The value of SSIM is ranged from 0-1. Then mean of MSSIM is computed according to Eq. 14 (Wang *et al*., 2004; Navas *et al*., 2007):

$$\text{MSSIM} = \frac{1}{m}\sum_{i=1}^{m}\text{SSIM}(x_i, y_i) \qquad (14)$$

where, x and y represents the cover and the watermarked image, respectively x_i, y_i represent the images contents at the ith local window.

Normalization Correlation (NC): The Normalization Correlation (NC) is another benchmark a standard method used for evaluating the similarity of correlation between the embedded and extracted watermark. The NC is defined as follows (Navas *et al*., 2007; Kutter and Petitcolas, 1999):

$$\text{NC} = \frac{\sum_{i=1}^{M} W_i W'_i}{\sqrt{\sum_{i=1}^{M} W_i^2}\sqrt{\sum_{i=1}^{M} W'^2_i}} \qquad (15)$$

where, w and w', represents the embedded and the extracted watermark, respectively.

RESULTS AND DISCUSSION

This study discusses the efficiency of the suggested watermarking method. Several standard 8-bit grayscale images are used with the suggested scheme. In addition, different kinds of attacks is applied with the scheme. The effect of scaling factors on the perceptual invisibility is tested with PSNR and SSIM. Table 1 shows the effect of scaling factors on the perceptual invisibility of the watermarked images where the chosen values are 4, 6, 8 and 12. Table 1 shows that the proposed watermarking scheme has acceptable invisibility for the different scaling factors. Figure 5 shows that the distortion produced by

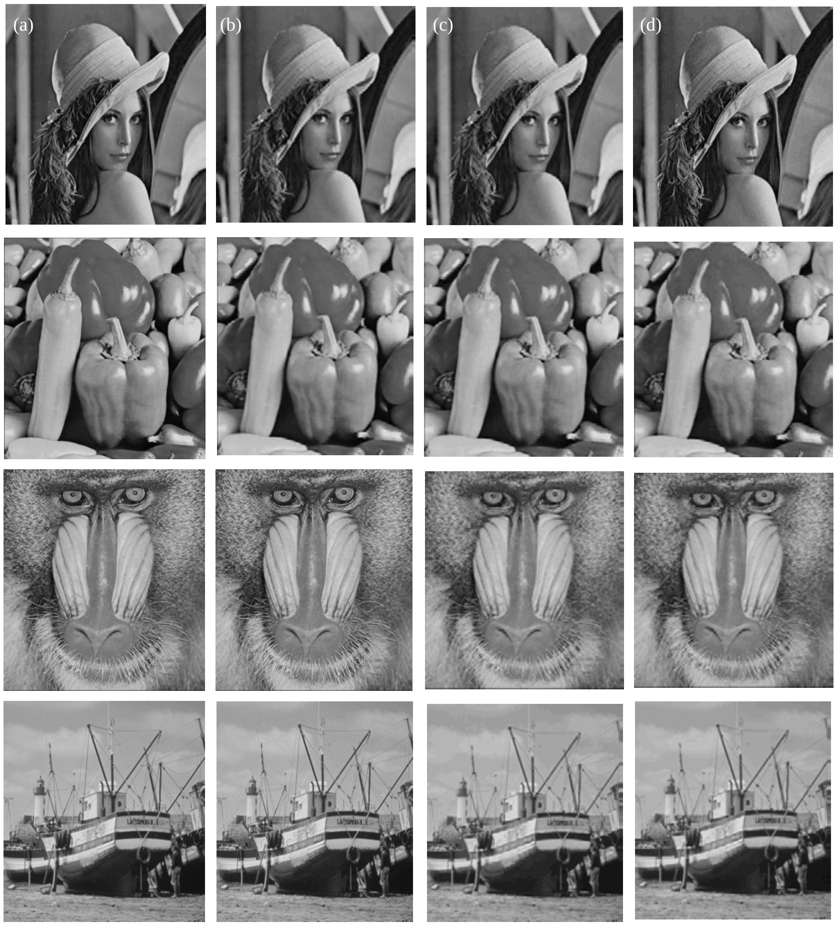


Fig. 5: Different test images and their watermarked versions: a) Original images; b) Watermarked images with scale factor of 4; c) Watermarked images with scale factor 6 and d) Watermarked images with scale factor 8

Table 1: The PSNR and SSIM with different scaling factors and different standard images

Scaling Factors (SF)	PSNR	SSIM	NC
Lenna image			
4	45.0738	0.9922	1.0000
6	41.5022	0.9841	0.9999
8	39.2825	0.9733	1.0000
12	35.8429	0.9495	0.9999
Peppers image			
4	45.0492	0.9916	1.0000
6	41.3681	0.9823	0.9999
8	39.2832	0.9717	1.0000
12	35.8015	0.9436	1.0000
Baboon image			
4	45.0707	0.9964	1.0000
6	41.4133	0.9925	1.0000
8	39.4039	0.9873	1.0000
12	35.7414	0.9719	1.0000
Boat image			
4	45.1024	0.9921	1.0000
6	41.5214	0.9859	0.9999
8	39.3935	0.9779	1.0000
12	35.7458	0.9598	1.0000

the new watermarking scheme is undetectable in all tested images with scaling factors (4, 6 and 8). Visually, the watermarked image is relatively close to the original image. In all our tests we will use (scaling factor = 6).

In order to approve the efficiency of the suggested scheme, the watermarked image is attacked with different attacks such as JPEG compression, salt and pepper noise, Gaussian noise, smoothing, median filtering and histogram equalization.

Robustness against JPEG compression: In this test, the watermark image is subject to JPEG compressor at different quality factors. Compression attack is one of the most important attacks. When the compression is applied on the watermark image, its quality will be decreased and it became difficult to recover the embedding watermark. The proposed method has good robustness against the JPEG compression attack. Table 2 shows the efficiency of the suggested scheme where the value of PSNR, SSIM and NC is of acceptable value.

Robustness against Salt and Pepper and Gaussian noise attacks: In this test, a watermarked is attacked with salt and pepper and Gaussian noises. The values of PSNR, SSIM and NC show that the suggested scheme is robust against Salt& Pepper and Gaussian noise.

Tamper detection: In this test, we show the ability of the proposed system to detect the location of tampered regions (malicious attack) into the image where small part of watermarked image is replaced with another part. Figure 6 shows this tampering.

Table 2 shows that the proposed system has resistance against JPEG lossy compression with different quality factor (Q). Table 3 shows that the proposed system has resistance against Salt and Pepper and Gaussian noises.

Table 2: The NC values under JPEG lossy compression with different quality factor (Q) (scale factor = 6, test image is Lena)

Quality factor (Q)	NC	PSNR	SSIM
20	0.6560	30.2793	0.8764
30	0.6807	31.8968	0.9089
40	0.6933	32.7448	0.9247
50	0.7122	33.4971	0.9428
60	0.7549	36.9392	0.9704
70	0.8270	42.1038	0.9859
80	0.8961	43.0970	0.9885
90	0.9746	45.9206	0.9932

Table 3: The NC values under density levels for salt and pepper and Gaussian noises (scale factor = 6, test image is Lena)

Density	PSNR	SSIM	NC
Salt and pepper			
0.001	35.2686	0.9803	0.9951
0.02	22.1442	0.6637	0.9343
0.1	15.1766	0.2544	0.8387
Gaussian noise			
0.001	20.2059	0.3820	0.6434
0.02	19.9964	0.3806	0.6390
0.1	20.2490	0.3830	0.6437

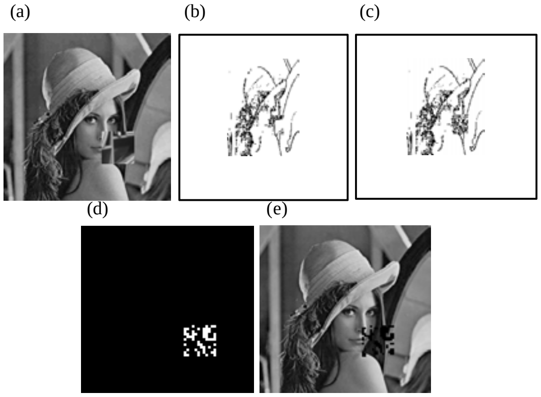


Fig. 6: Tampered detection process 1: a) tampered watermarked image; b) generated signature; c) extracted watermark; d) tampered region and e) detected tampered region

Figure 7 shows another example of tampering attack. In this attack, the pixels of watermarked image is tampered by replacing it with other intensities. The scheme can detect and localize the tampered region.

Another test is done for tampering on watermarked image. In this test, text is adding to the watermarked image as shown in Fig. 8. The text "LENA" is adding at the top of the image. Figure 9 shows the detected tampered region and tamper image is shown in Fig. 10.

In addition, the proposed watermark algorithm could detect the small content modification on image as shown from Fig. 11 where the clasp region is removed from Lena image.

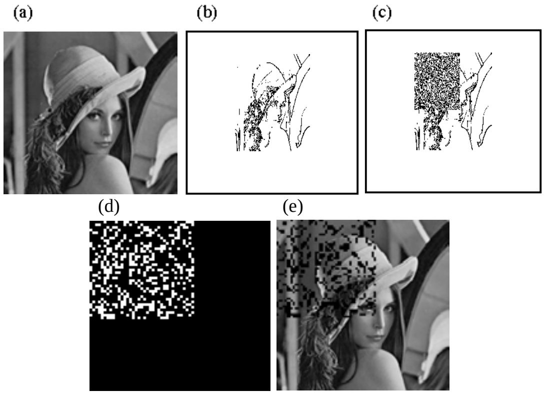


Fig. 7: Tampered detection process 2: a) tampered watermarked image; b) generated signature; c) extracted watermark; d) tampered region and e) detected tampered region

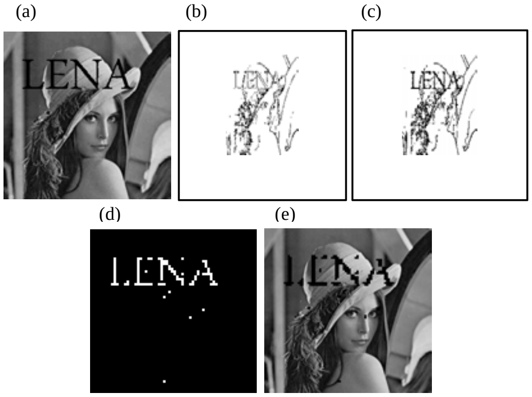


Fig. 8: Tampered detection process 3: a) tampered watermarked image; b) generated signature; c) extracted watermark; d) tampered region and e) detected tampered region

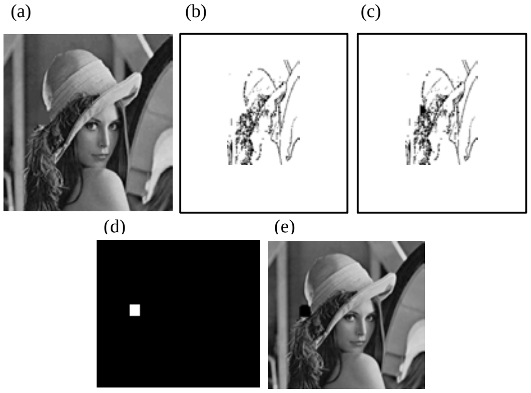


Fig. 9: Cropping attack: a) intentional attack b) generated signature c) extracted watermark d) difference between two watermarks (c, b)

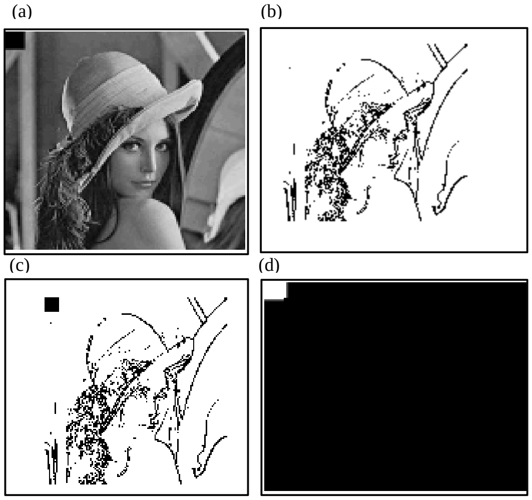


Fig. 10: Tampered detection process 4: a) tampered watermarked image; b) generated signature; c) extracted watermark d) tampered region and e) detected tampered region

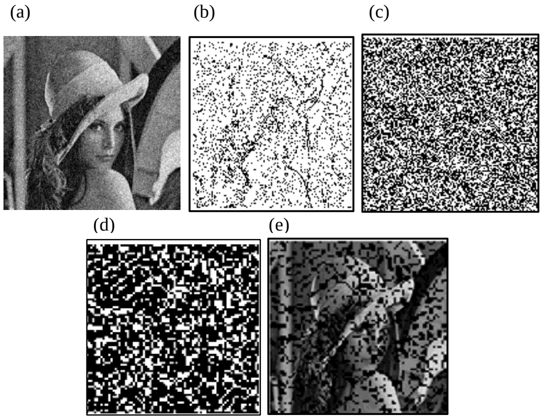


Fig. 11: Gaussian noise attack: a) unintentional attack; b) generated signature; c) extracted watermark; d) difference between two watermarks (c, b) and e) isolated error blocks

Figure 11 shows that the proposed method has the ability to detect and localize the cropped region. Another example is when watermarked image is attacked with Gaussian noise with standard deviation (0.0001). As shown in Fig. 11 the error blocks are distributed in binary image and this refers to unintentional attack.

CONCLUSION

This study has suggested a blind semi fragile watermarking scheme tor proving the authenticity of an image scheme. This technique based on wavelet edge. For ensuring the robustness of the watermark, it embedded in the low frequency domain of DWT. As shown in Table 1, the suggested method does not produced any perceptible changes in the cover image and provides a very high probability of tamper detection. The suggested method can detect whether the image is tampered or not. In addition, the proposed method can localize the tampered region within image. The suggested method can also decide whether the attack on image can be done intentionally or unintentionally. As shown in Table 1-3 and Fig. 8-11, the suggested scheme is robust against unintentional attacks (image processing operation) while fragile against intentional (malicious) attacks.

REFERENCES

Cruz, C., J.A. Mendoza, M.N. Miyatake, H.P. Meana and B. Ndrkoski, 2009. Semi-fragile watermarking based image authentication with recovery capability. Proceeding of the 2009 International Conference on Information Engineering and Computer Science, December 19-20, 2009, IEEE, Mexico City, Mexico, ISBN:978-1-4244-4994-1, pp: 1-4.

Huynh-Thu, Q. and M. Ghanbari, 2008. Scope of validity of PSNR in image/video quality assessment. Electronics Lett., 44: 800-801.

Kutter, M. and F.A.P. Petitcolas, 1999. A fair benchmark for image watermarking systems. Proceedings of the Security and Watermarking of Multimedia Contents, January 25-27, 1999, Society of Photo-Optical Instrumentation Engineers, Sans Jose, California, USA., pp: 226-239.

Lintao, L.V., H. Fan, W. Jinfeng and Y. Yuxiang, 2012. A semi-fragile watermarking scheme for image tamper localization and recovery. J. Theor. Appl. Inf. Technol., 42: 287-291.

Lu, J., M. Wang, J. Dai, Q. Huang, L. Li and C.C. Chang, 2015. Multiple watermark scheme based on DWT-DCT quantization for medical images. J. Inform. Hiding Multimedia Signal Process., 6: 458-472.

Manickam, L., S.A.K. Jilani and M.N. Prasad, 2013. A novel fragile watermarking scheme for image tamper detection using K mean clustering. Int. J. Comput. Trends Technol., 4: QA75.5-QA76.95.

Navas, K.A., M. Sasikumar and S. Sreevidya, 2007. A benchmark for medical image watermarking. Proceedings of the 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services, June 27-30, IEEE Computer Society, Maribor, pp: 237-240.

Preda, R.O., I. Marcu and A. Ciobanu, 2015. Image authentication and recovery using wavelet-based dual watermarking. U.P.B. Sci. Bull. Ser. C., 77: 200-212.

Qi, X., X. Xin and R. Chang, 2009. Image authentication and tamper detection using two complementary watermarks. Proceeding of the 2009 16th IEEE International Conference on Image Processing (ICIP), November 7-10, 2009, IEEE, Logan, Utah, ISBN:978-1-4244-5653-6, pp: 4257-4260.

Shukla, M.K.M. and M.A.K. Mehta, 2015. A review on digital watermarking techniques, applications and attacks. Int. J. Eng. Comput., 4: 11237-11245.

Sumalatha, L., V.V. Krishna and A.V. Babu, 2012. Image content authentication based on wavelet edge features. Int. J. Comput. Appl., 49: 24-24.

Tiwari, A. and M. Sharma, 2012. Semifragile watermarking schemes for image authentication-A survey. Int. J. Comput. Network Inf. Secur., 4: 43-49.

Wang, Z., A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, 2004. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process., 13: 600-612.