http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 9 (1): 20-26, 2010
ISSN 1812-5638

A Semi-Fragile Watermarking Algorithm using Adaptive Least Significant Bit Substitution

[1,2]Hengfu Yang, [1]Xingming Sun and [1,3]Guang Sun
[1]School of Computer and Communication, Hunan University, Changsha, 410082, China
[2]Department of Information Science and Engineering,
Hunan First Normal University, Changsha, 410205, China
[3]Department of Information Management,
Hunan Financial and Economic College, Changsha, 410205, China

Abstract: A novel semi-fragile watermarking scheme in image spatial domain is proposed. Each watermark bit is duplicated and embedded into sub-blocks of the host image by adaptive Least Significant Bit (LSB) substitution. The adaptive LSB substitution fully exploits the Human Visual System (HVS) masking characteristics, which ensures high visual quality of watermarked image. The watermark for authentication is extracted by taking a majority vote on the extracted bits. To differentiate attack types effectively, a classification rule for image authentication was developed. Experimental results show that the proposed scheme has good transparency and robustness against admissible signal operations, while it is sensitive to malicious attacks such as heavy noise addition, rotation by large angle, cutting and pasting. In addition, the scheme can localize the tampered region precisely.

Key words: Adaptive LSB substitution, human visual system, semi-fragile watermarking, tampering localization

INTRODUCTION

With the fast development of Internet and multimedia information processing technology, the whole world has become smaller and smaller. By the internet, nearly everything can be found and downloaded, such as audios, images and videos. Digital media can be easily redistributed, copied and modified over the Internet nowadays, which renders copyright protection issues. The integrity and authenticity of digital media can be protected by using digital watermarking. Digital watermarking directly embeds signal (watermark) containing owner identification into the host signal in such a way that the watermark cannot be removed without serious visual quality degradation of the host signal. The watermarking techniques can be broadly classified into two categories: robust watermarking and fragile watermarking. Robust watermarks (Seo and Yoo, 2006; Thirugnanam *et al*., 2009) are generally used for copyright protection and ownership proof since they are robust to nearly all kinds of image manipulations. In comparison, fragile watermarks (Zhang and Wang, 2009; Liu *et al*., 2007) are useful for purposes of content authentication and integrity verification since they are completely fragile to any modifications. Semi-fragile watermarks are tolerant to allowable modifications (content preserving operations, incidental attacks) but sensitive to malicious attacks and are more practical than fragile watermarks in image authentication, since they are robust to content preserving operations. In recent years, semi-fragile watermarking has received considerable attention and many semi-fragile watermarking algorithms have been reported (Feng *et al*., 2005; Xiao and Wang, 2008; Zhao and Sun, 2008; Woo *et al*., 2009). Feng *et al*. (2005) presented a semi-fragile watermark based on the image permutation. The method can survive the JPEG compression, but it has low tamper localization precision and embedded capacity. Xiao and Wang (2008) devised a novel semi-fragile image watermarking scheme based on the theory of Laplacian sharpening. This scheme can tolerate Laplacian sharpening very well, while it is very fragile to neighborhood averaging and median filter. Zhao and Sun (2008) utilized HVS masking characteristics to develop a semi-fragile watermarking algorithm in wavelet domain, which can accept mild JPEG compression but is fragile to filtering, noise addition. In Woo *et al*. (2009) method, the watermark is generated by taking a down-scale version of the cover image and embedded into the horizontal sub-band and the vertical sub-band in wavelet domain. This method a llows high quality JPEG

Corresponding Author: Xingming Sun, School of Computer and Communication, Hunan University, No. 252, Lushan South Road, Yuelu District, Changsha, 410082, China Tel: 86-731-88821341 Fax: 86-731-88821341

compression, minor local distortion and minimal noise insertion, while it is fragile to mean filtering. Whereas, mild salt-and-pepper noise additions, low pass filtering only do mild modifications to images, they shouldn't be classified as malicious attacks. To achieve higher robustness to allowable modifications and more accurate authentication, a novel semi-fragile watermarking algorithm was proposed by considering the HVS characteristics (such as luminance, texture and edge sensitivity). In this scheme, the watermark bits are embedded into the host image by adaptive LSB substitution.

HUMAN VISUAL SYSTEM MODEL

To balance transparency and robustness, an effective watermarking scheme should exploit HVS masking characteristics. Here, the luminance masking, texture masking and edge masking features of the cover image was used to develop a spatial HVS model in a better way.

Human eyes are more sensitive to changes in the areas with middle level luminance, while less sensitive to changes in those areas with high and low gray-scale values (Barni *et al*., 2001). The calculation of luminance masking is as follows:

$$\alpha(i,j)=\frac{|x(i,j)-128|}{128} \quad (1)$$

where, x(i, j) is the pixel value of pixel I(i, j).

As for texture masking, we can use the entropy value to approximately depict the texture masking (Yang and Sun, 2007). Let H(i, j) be the entropy value of sub-block (size (2b+1)×(2b+1), where $\{1\le b\le 4\}$) centered by the pixel I(i, j), this maximum entropy is achievable when all of the gray-levels have the same probability. For a 256 level image sub-block with size of (2b+1)×(2b+1), the maximum entropy value H_{max} is computed as follows:

$$H_{max}=-\sum_{i=-b}^{b}\sum_{j=-b}^{b}\frac{1}{(2b+1)^2}\log_2\left(\frac{1}{(2b+1)^2}\right)=2\log_2(2b+1) \quad (2)$$

Then, the normalized entropy β(i, j) of each pixel can be obtained by the equation:

$$\beta(i,j)=\frac{H(i,j)}{H_{max}} \quad (3)$$

Many proposed methods did not handle the edge masking very well (Wu *et al*., 2005; Yang *et al*., 2008). Obviously, when the texture masking is computed, it also includes the edge parts. However, if the texture masking is used directly, the edge may be easily corrupted and it causes severe distortions to the cover image. Therefore, we have to correct the edge regions. Because image areas with prominent edges have greater variance value, while smooth image areas have smaller variance value. Here, the variance V(i, j) of the image blocks will be used to indicate the edge masking. In addition, we know that the maximum variance is in the block where adjacent pixels have the maximum and minimum permissible gray-scale value. The maximum variance V_{max} is as follows:

$$V_{max}=\frac{(2b^2+2b)(2b^2+2b+1)}{(2b+1)^4}G^2<\left(\frac{\left(2b^2+2b+\frac{1}{2}\right)}{(2b+1)^2}G\right)^2=\left(\frac{G}{2}\right)^2 \quad (4)$$

where, G is the maximum permissible gray-scale value.

The normalized variance γ(i, j) can be obtained by the following Eq. 5.

$$\gamma(i,j)=\frac{V(i,j)}{V_{max}}=\frac{V(i,j)}{(G/2)^2} \quad (5)$$

It is well known that sharp edges play an important role in human spatial vision. To achieve good invisibility, a good watermarking scheme should avoid abrupt changes in edge areas. By combining the luminance masking, texture masking and edge masking together, the final HVS masking can be expressed as follows:

$$\phi(i,j)=\alpha(i,j)\times\{\beta(i,j)-\gamma(i,j)\} \quad (6)$$

The reason the edge masking is subtracted from the texture masking is because the edge has less ability to cover the watermark than high activity areas. After this combination is finished, we can compute the adaptive bit depth k(i, j) of each pixel used for adaptive LSB substitution by the following Eq. 7.

$$k(i,j)=\text{round}\left((7-r)\times\frac{\phi(i,j)-\min(\phi)}{\max(\phi)-\min(\phi)}\right) \quad (7)$$

where, function round(), max() and min() return the nearest integer, the maximum and minimum value of an expression, respectively. $1\le r\le 7$, $0\le k(i,j)\le 7-r$ and r represents the number of MSB (Most Significant Bits) of each pixel used to calculate the bit depth (note that watermark bits never be embedded into the r MSBs).

SEMI-FRAGILE WATERMARKING SCHEME BASED ON ADAPTIVE LSB SUBSTITUTION

To achieve high robustness to acceptable signal operations and good tamper localization precision, the luminance masking, texture masking and edge masking

was utilized to develop a spatial domain HVS model. Then, an image adaptive semi-fragile watermarking algorithm was proposed based on the HVS model. Figure 1 shows the watermark embedding procedure.

Adaptive LSB substitution: The watermark bits are embedded into the image sub-block by the adaptive LSB substitution. Let I be the original 256 level grayscale image with size of m×n represented as:

$$I = \{x(i,j) | 0 \le i < m, 0 \le j < n, x(i,j) \in \{0,1,...,255\}\} \quad (8)$$

Let W be the binary watermark image with size of s×t represented as:

$$W = \{w(u,v) | 0 \le u < s, 0 \le v < t, w(u,v) \in \{0,1\}\} \quad (9)$$

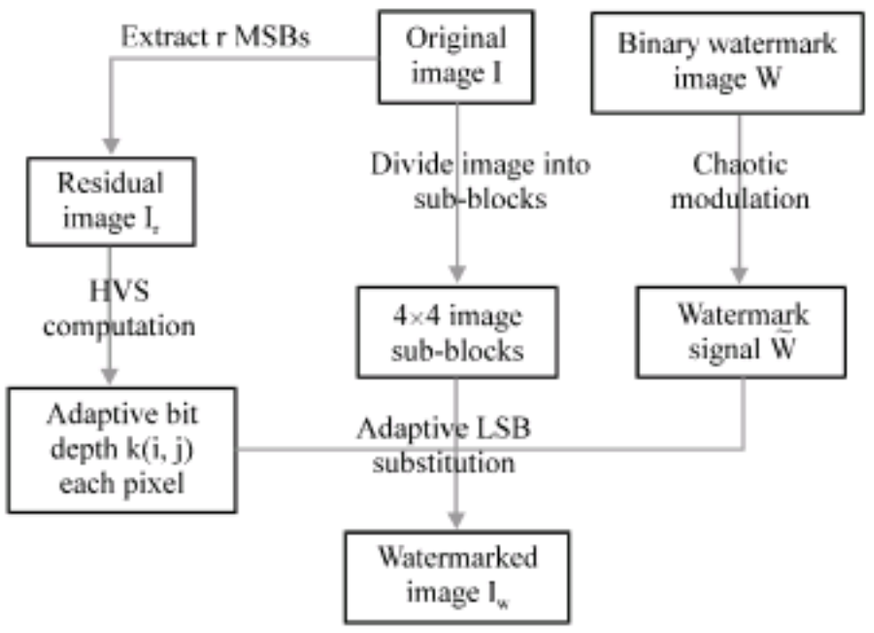

Fig. 1: Block diagram of watermark embedding

Details of watermark embedding are described as follows:

Step 1: Extract the r (say r = 4) MSBs of the original image I to get the residual image I_r.

Step 2: Compute the adaptive bit depth k(i, j) of each pixel in the original cover-image based on the residual image I_r using Eq. 7.

Step 3: For enhanced security, a pseudo random sequence generated by a key (e.g., chaotic map (Xiang *et al*., 1999) is used to modulate the watermark signal. The modulated watermark is defined as:

$$\tilde{W} = \{\tilde{w}(u,v) | 0 \le u < s, 0 \le v < t, \tilde{w}(u,v) \in \{0,1\}\} \quad (10)$$

Step 4: Divide the original image I into 4×4 sub-blocks.

Step 5: Repeatedly embed the watermark bit $\tilde{w}(u,v)$ into the k(i, j) LSBs of each pixel I(i, j) of the image sub-block B_{uv} by adaptive LSB substitution.

Step 6: Repeat step 5 until all the watermark bits are embedded into the cover image and obtain the watermarked image I_w.

Figure 2 gives an example of adaptive LSB substitution.

Watermark extraction: The extraction procedure includes the following step:

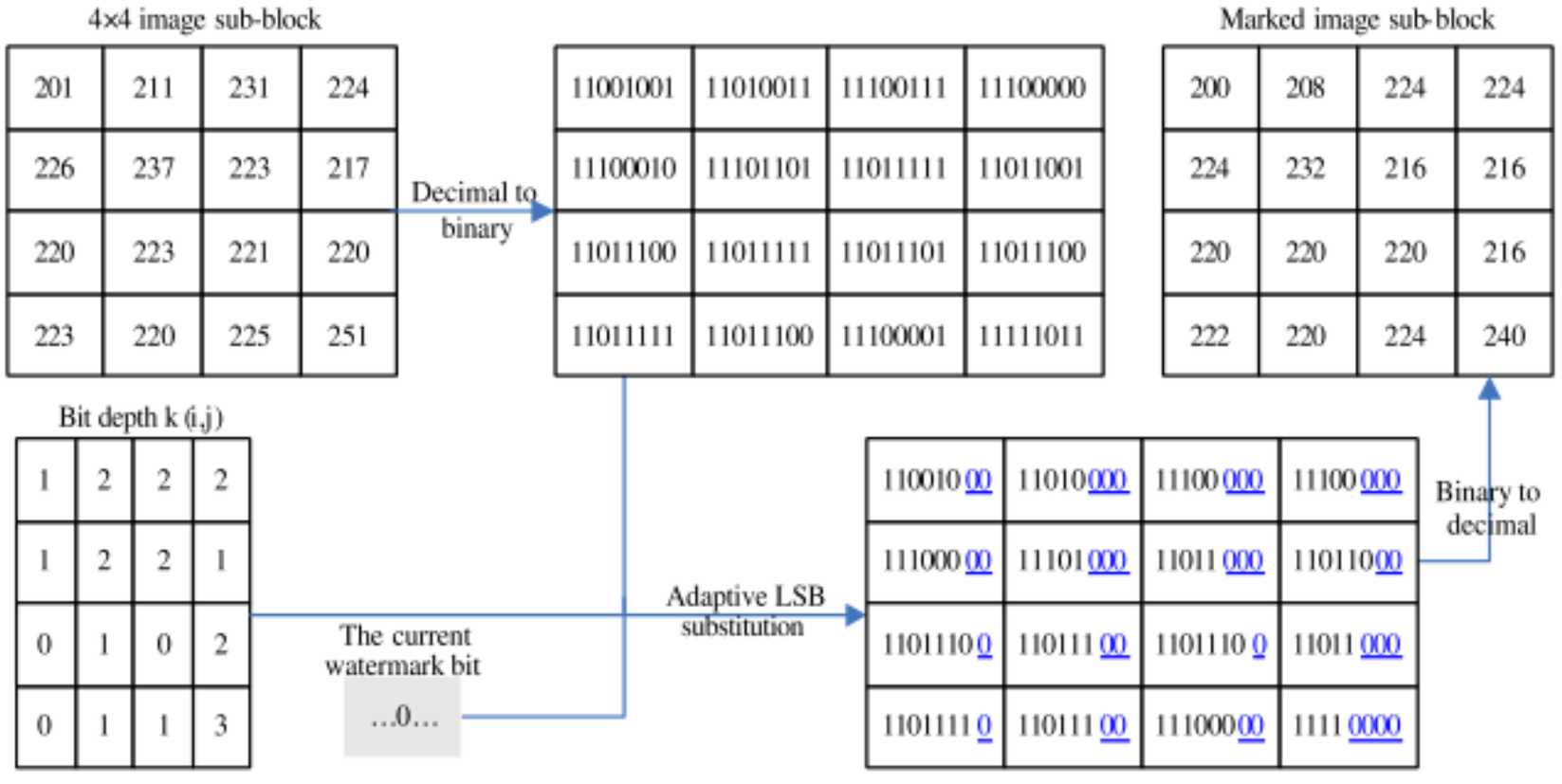

Fig. 2: An example of adaptive LSB substitution

Step 1: Extract the r MSBs of the watermarked image I_w to get the residual image I_r.

Step 2: Calculate the adaptive bit depth k(i, j) of each pixel based on the residual image I_r using Eq. 7.

Step 3: Extract k(i, j) LSBs of each pixel in the image sub-block B_{uv} and then obtain the watermark bit embedded into the sub-block B_{uv} by majority vote defined as:

$$\tilde{w}'(u,v)=\text{round}\left(\frac{\sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{l=0}^{k(4u+i,4v+j)}\text{LSB}_l\left(I_w(4u+i,4v+j)\right)}{\sum_{i=0}^{3}\sum_{j=0}^{3}\left(k(4u+i,4v+j)+1\right)}\right) \quad (11)$$

where, $\text{LSB}_l(\ )$ is the function that extracts the lth LSB of a pixel.

Step 4: Repeat step 3 until all watermark bits are extracted. Finally, demodulate the watermark signal $\tilde{W}'$ by a pseudo random sequence generated by the same key in embedding process to obtain the final watermark W' expressed as:

$$W'=\left\{w'(u,v)\middle|0\le u<s,0\le v<t,w'(u,v)\in\{0,1\}\right\} \quad (12)$$

Image authentication: Let D be the difference watermark image defined as:

$$D=\left\{d(u,v)\middle|\begin{matrix}d(u,v)=w(u,v)\otimes w'(u,v),\\0\le u<s,0\le v<t\end{matrix}\right\} \quad (13)$$

where, $\otimes$ denotes XOR operation.

The pixel in the difference watermark image has value 1 (white pixel) if the extracted watermark bit is false, 0 (black pixel) otherwise. In the case of mild modification, most of the mark error pixels are spread on the difference watermark image or the extracted watermark. On the contrary, these mark error pixels are concentrated in certain regions.

For a mark error pixel in the difference watermark image, it is a tampered pixel if at least one of its eight neighbor pixels is a mark error pixel and a maliciously tampered pixel if at least two of its surrounding eight neighbors are mark error pixels and a noise point caused by error detection otherwise. So, we can define the following parameters.

$$\phi_e=\left\{\begin{matrix}\text{The total of mark error pixels in}\\\text{the difference watermark image}\end{matrix}\right\} \quad (14)$$

$$\phi_t=\left\{\begin{matrix}\text{The total of tampered pixels in}\\\text{the difference watermark image}\end{matrix}\right\} \quad (15)$$

$$\phi_m=\left\{\begin{matrix}\text{The total of maliciously tampered pixels}\\\text{in the difference watermark image}\end{matrix}\right\} \quad (16)$$

$$\phi=\left\{\begin{matrix}\text{The total of pixels in the}\\\text{difference watermark image}\end{matrix}\right\} \quad (17)$$

$$\rho=\frac{\phi_e}{\phi}=\frac{\phi_e}{m\times n} \quad (18)$$

$$\mu=\frac{\phi_m}{E(\phi_t)} \quad (19)$$

where, E( ) returns the mathematical expectation.

Now, an objective judgment rule can be defined to decide whether the tested image is maliciously tampered or mildly modified.

If $\rho = 0$, the tested image has been neither maliciously tampered nor incidental altered.

If $\rho>0$ and $\mu<T$, then the tested image has encountered only incidental alterations, where T is the preset threshold, $T \in [0.5, 1]$. In the experiments, we set T = 0.9466. Yang and Sun's (2007) method for the selection of threshold T.

If $\rho>0$ and $\mu\ge T$, then the tested image has been maliciously tampered.

Moreover, for an unauthentic test image, noise-like pixels can be removed and a compact tamper region in the difference watermark image can be created by labeling all pixels in the sub-block centered by a maliciously tampered pixel as white pixels.

EXPERIMENT RESULTS

Some different types of standard test images with size of 256×256 are used as the cover-images. The watermark is a binary image with size 64×64. In the experiment, we set the slide window size is 3×3, i.e., b = 1 and set the number r of MSBs for the calculation of adaptive bit depth equals to 4.

Invisibility: Taking Lena image as an example, the invisibility of the proposed algorithm is examined. Figure 3a-d show the original image, watermark image, watermarked image and the 15-times magnified difference image between the original image and the watermarked image, respectively. More experiments results are shown in Table 1.

The visual quality of watermarked images is measured by PSNR (Peak Signal Noise Ratio) value. Human eyes

can not percept the existence of the watermark from Fig. 3c. From Table 1, it can be seen that the proposed scheme can obtain PSNR values over 38 dB which indicates that the watermark is invisible to human eyes. Furthermore, we notice that the embedded watermark is adaptive to the original image features by observing the difference images (Fig. 3d).

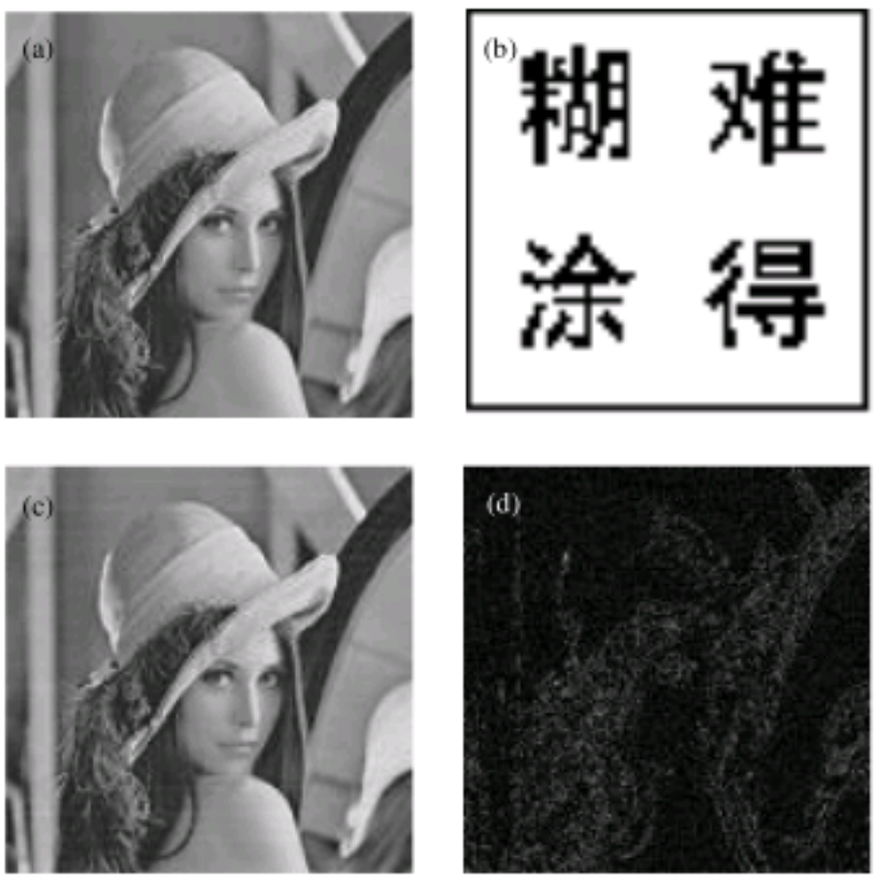

Fig. 3: Watermark invisibility, (a) original image, (b) watermark image, (c) watermarked image and (d) the difference image between a and c

Table 1: PSNR values of the watermarked images (dB)

Images	PSNR
Lena	39.60
Couple	40.51
Barbara	40.37
Boat	39.82
F-16	38.14
Peppers	39.47
Houses	38.01
Pills	38.35

Attack classification: The experimental results on the judgment rule for the attack classification and performance comparison with existing schemes was shown in Table 2. The attack is malicious if μ is greater than T (0.9446), incidental otherwise. By this rule, our scheme can distinguish effectively malicious attacks from incidental ones. Our scheme is more efficient than other algorithms (Xiao and Wang, 2008; Zhao and Sun, 2008; Woo *et al*., 2009) in considering the automatic classification rule for the type of attacks because of the automatic computation of the threshold T. Moreover, the proposed algorithm accurately classifies mild salt-and-pepper noise addition and low pass filtering as incidental attacks while some other semi-fragile algorithms (Xiao and Wang, 2008; Zhao and Sun, 2008; Woo *et al*., 2009) falsely categorize them as malicious attacks.

Robustness to incidental attacks: Taking the stego-image Fig. 3c for an example, the robustness against incidental attacks will be examined including mild JPEG compression, low-pass filtering, Laplacian sharpening, salt and pepper noise addition. The robustness is evaluated by Normalized Correlation (NC) and the results are shown in Table 3. The proposed algorithm has high NC values under some mild modifications and shows good robustness against incidental attacks.

Tamper localization: Figure 4 shows some examples of tamper localization. The unaltered stego-image is shown in Fig. 3c. In Fig. 4a, the hat and hairs in the mirror was cropped. In Fig. 4b, tampering was done by cropping the face and pasting it on the bottom right corner and the image in Fig. 4c was tampered by cropping a block with size of 16×16 on the top left corner. Figure 4d-f show the corresponding difference watermark images. Tamper localization correctly highlighted the tampered region marked by white pixels as depicted in Fig. 4g-i, which indicates the proposed scheme can accurately detect the tampered region.

Table 2: Attacks classification

			Attack type			
Attacks	ρ	μ	Present scheme	Xiao and Wang (2008)	Zhao and Sun (2008)	Woo *et al*. (2009)
No attack	0.0000	0.0000	Authentic	Authentic	Authentic	Authentic
JPEG (quality factor 100)	0.0417	0.0612	Incidental	Incidental	Incidental	Incidental
JPEG (quality factor 90)	0.4072	0.9082	Incidental	Malicious	Incidental	Incidental
Mean filtering	0.3696	0.7722	Incidental	Malicious	Malicious	Malicious
Median filtering	0.0037	0.0021	Incidental	Malicious	Malicious	Malicious
Laplacian sharpening	0.1399	0.6530	Incidental	Incidental	Malicious	Malicious
15% salt and pepper noise	0.0001	0.0000	Incidental	Malicious	Malicious	Malicious
1% Gaussian noise	0.5098	0.9648	Malicious	Malicious	Malicious	Malicious
Clockwise rotation by 3°	0.4683	0.9477	Malicious	Malicious	Malicious	Malicious
Cropping in the upper-left quarter	0.1475	0.9868	Malicious	Malicious	Malicious	Malicious
Cropping in the upper-left 16×16 block	0.0026	1.0000	Malicious	Malicious	Malicious	Malicious

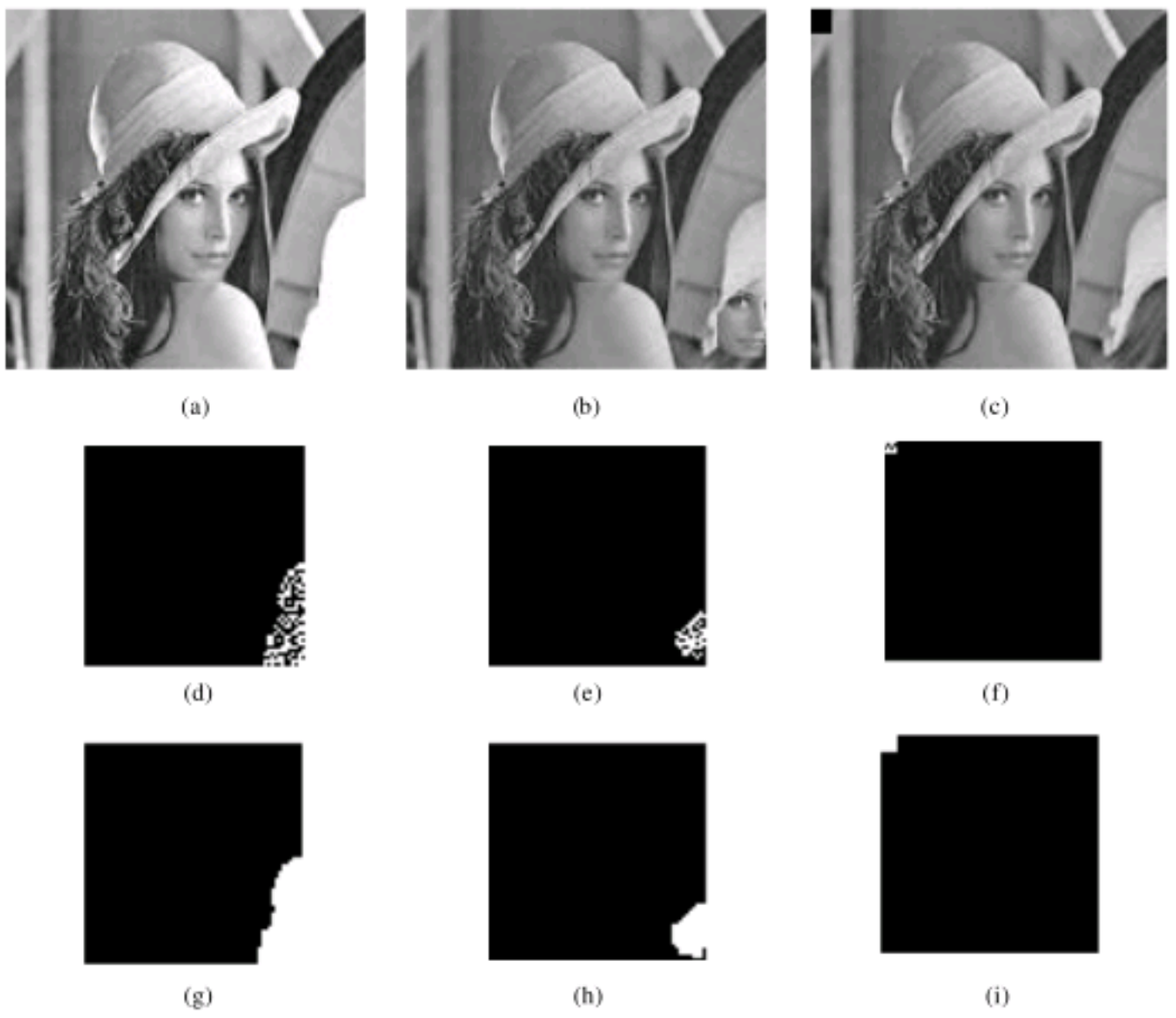

Fig. 4: Tamper detection and localization, (a-c) tampered images, (d-f) difference watermark images and (g-i) detected tampered regions

Table 3: Robustness to incidental attacks

Attacks	NC
JPEG (quality factor 100)	0.9583
JPEG (quality factor 95)	0.7659
10% Salt and pepper noise	1.0000
Mean filtering	0.7304
Median filtering	0.9963
Laplacian sharpening	0.8601

CONCLUSION

An image adaptive semi-fragile watermarking scheme for image authentication and tamper detection was proposed. The scheme embedded the duplicated watermark bits into sub-blocks of the host image by adaptive LSB substitution and had low computation cost. Due to the usage of the HVS masking mode during adaptive LSB substitution, the watermark is adaptive to the original image feature, which ensures high visual quality of watermarked image. An effective classification rule for image authentication was developed, which could differentiate effectively malicious attacks from incidental attacks. Experiment results showed that the proposed scheme had good robustness to incidental attacks, while it was very fragile to malicious attacks. Moreover, this algorithm can localize maliciously tampered regions accurately (marked by white pixels).

ACKNOWLEDGMENTS

This study was supported by National Natural Science Foundation of China (60736016, 60873198), Scientific Research Fund of Hunan Provincial Education Department of China (08C018) and National Basic Research Program of China (2006CB303000, 2009CB326202).

REFERENCES

Barni, M., F. Bartolini and A. Piva, 2001. Improved wavelet-based watermarking through pixel-wise masking. IEEE Trans. Image Process., 10: 783-791.

Feng, G.R., L.G. Jiang and C. He, 2005. Permutation-based semi-fragile watermark scheme. IEICE Trans. Fundamentals, 88: 374-377.

Liu, S.H., H.X. Yao, W. Gao and Y.L. Liu, 2007. An image fragile watermark scheme based on chaotic image pattern and pixel-pairs. Applied Math. Comput., 185: 869-882.

Seo, J.S. and C.D. Yoo, 2006. Image watermarking based on invariant regions of scale-space representation. IEEE Trans. Signal Process., 54: 1537-1549.

Thirugnanam, G., M. Natarajan, P. Mangaiyarkarasi and S. Arulselvi, 2009. Wavelet-based watermarking scheme using filter parametrisation for medical images. Int. J. Med. Eng. Informat., 1: 435-444.

Woo, C.S., D. Jiang and P. Binh, 2009. Semi fragile watermark with self authentication and self recovery. Malaysian J. Comput. Sci., 22: 64-84.

Wu, H.C., N.I. Wu, C.S. Tsai and M.S. Hwang, 2005. Image steganographic scheme based on pixel-value differencing and LSB replacement methods. IEE Proc. Vision Image Signal, 152: 611-615.

Xiang, H., L.D. Wang and H. Lin, 1999. Digital watermarking systems with chaotic sequence. Proceedings of the Electronic Imaging'99, Security and Watermarking of Multimedia Contents, (EISWMC'99), SPIE, San Jose, CA, USA., pp: 449-457.

Xiao, J. and Y. Wang, 2008. A semi-fragile watermarking tolerant of *Laplacian sharpening*. Proceedings of the International Conference on Computer Science and Software Engineering, Wuhan, China, (ICCSSE'08), IEEE Computer Society, pp: 579-582.

Yang, H.F. and X.M. Sun, 2007. Semi-fragile watermarking for image authentication and tamper detection using HVS mode. Proceedings of the International Conference on Multimedia and Ubiquitous Engineering, Seoul, Korea, (ICMUE'07), IEEE Computer Society, pp: 1112-1117.

Yang, C.H., C.Y. Weng, S.J. Wang and H.M. Sun, 2008. Adaptive data hiding in edge areas of images with spatial LSB domain systems. IEEE Trans. Inform. Forensics Security, 3: 488-497.

Zhang, X. and S. Wang, 2009. Fragile watermarking scheme using a hierarchical mechanism. Signal Process, 89: 675-679.

Zhao, Y. and X.H. Sun, 2008. A semi-fragile watermarking algorithm based on HVS model and DWT. Proceedings of the International Conference on Computer Science and Software Engineering, Wuhan, China, (ICCSSE'08), IEEE Computer Society, pp: 638-641.