Multimed Tools Appl (2011) 54:385–395
DOI 10.1007/s11042-010-0541-z

Self-embedding watermark with flexible restoration quality

Xinpeng Zhang · Shuozhong Wang · Zhenxing Qian · Guorui Feng

Published online: 25 May 2010

Abstract A novel self-embedding watermarking scheme is proposed, in which the reference data derived from the most significant bits (MSB) of host image and the localization data derived from MSB and reference data are embedded into the least significant bits (LSB) of the cover. At authentication side, while the localization data are used to detect the blocks containing substitute information, the reference data extracted from other regions and the spatial correlation are exploited to recover the principal content in tampered area by a pixel-by-pixel manner. In this scheme, the narrower the tampered area is, the higher quality of recovered content can be obtained.

Keywords Fragile watermarking · Self-embedding · Image quality

1 Introduction

Fragile watermarking aims to check integrity and authenticity of digital contents and to locate the areas replaced with fake information. While block-wise schemes have been developed to detect blocks containing fake contents [4, 5], some pixel-wise schemes can accurately locate the modified pixels when the tampered area is not too extensive [6]. If the embedded watermark data are derived both from pixels and blocks, a receiver can first identify the tampered blocks and then use the watermark hidden in the rest blocks to find the detailed modification pattern [7]. Since it takes advantages of both block-wise and

X. Zhang (✉) · S. Wang · Z. Qian · G. Feng
School of Communication and Information Engineering, Shanghai University, Shanghai 200072,
People's Republic of China
e-mail: xzhang@shu.edu.cn

S. Wang
e-mail: shuowang@shu.edu.cn

Z. Qian
e-mail: zxqian@shu.edu.cn

G. Feng
e-mail: grfeng@shu.edu.cn

pixel-wise techniques, the performance in locating tempered pixels is better than that of the method presented in [6]. Furthermore, in some self-embedding approaches, the embedded watermark is related to the host image content so that the original content in tampered area can be reconstructed. For example, the principal DCT coefficients or a low color depth version of the original contents can be embedded into the least significant bits (LSB), or into differences between pixels at different positions [1]. After locating malicious modification in a watermarked image, information extracted from the reserved regions is exploited to recover the principal content of the tampered areas. In [8], the embedded data are exclusive-OR between a pseudo-random sequence and the polarity sequence of DCT coefficients. Similarly, rough retrieval of the original content in the tampered areas can be obtained by iterative projections of the polarity information on a convex set. In [2], values of the first 11 DCT coefficients in each block are embedded into the LSB of another block and used to recover the content of tampered blocks.

In these self-embedding methods, the data representing the principal content in a region are always hidden in another region of the image. Thus restoration will fail when certain region and the region containing its original information are both tampered. We call it a *tampering coincidence* problem. If the content restoration is successful, the recovered content is derived from the data hidden in the corresponding region. That means the quality of restoration is the same no matter the tampered area is small or extensive.

We propose a new self-embedding watermarking scheme with flexible restoration quality. In this scheme, a block-wise mechanism for tampered area localization and a pixel-wise mechanism for original content recovery are integrated, and the reference data produced by exclusive-OR operation on the original MSB are embedded into the LSB planes. At authentication side, after identifying the tampered blocks, reference data extracted from other regions and the spatial correlation are exploited to recover the principal content in the tampered area in a pixel-by-pixel manner. Since the data representing the content of a block are scattered and embedded into the entire image, the tampering coincidence problem is avoided. Using this method, the smaller the tampered area, the more reference data can be extracted, leading to a higher the quality of the recovered content.

2 Watermark embedding procedure

In the scheme, the 5 most significant bits (MSB) of all pixels in the host image are kept unchanged, and the 3 least significant bits (LSB) of all pixels are replaced with watermark data. Here, the watermark data are determined by the MSBs and made up of two parts, which are respectively used to locate tampered blocks and to recover the original content.

The detailed steps are as follows:

1. Denote the numbers of rows and columns in an original image as N_1 and N_2, the total number of pixels as N ($N=N_1\times N_2$), and the gray pixel-values $p_n \in [0, 255]$, $n=1, 2, \ldots$, N. Here, we assume that both N_1 and N_2 are multiples of 8. Each p_n can be represented by 8 bits, $b_{n,7}$, $b_{n,6}$, ..., $b_{n,0}$, where

$$b_{n,m} = \lfloor p_n / 2^m \rfloor \bmod 2 , \quad m = 0, 1, \ldots , 7 \tag{1}$$

   We permute the N pixels, and the way of permutation is determined by a secret key. The permuted pixels are then divided into a series of pixel-pairs, each of which containing two pixels. Therefore the number of pixel-pairs is $N/2$. For each pixel-pair, denote its two pixels as p_i and p_j, and obtain 5 bits by executing exclusive-OR operations between the higher MSB of a pixel and the lower MSB of another one,

$$r_m = b_{i,7-m} \oplus b_{j,m+3}\ , \quad m = 0,\ 1, \ldots, 4 \tag{2}$$

We call the calculated bits as reference-bits. So, we get a total of $5N/2$ reference-bits from $N/2$ pixel-pairs. Then, we segment the $N/2$ pixel-pairs into $N/64$ sets, each of which containing 32 pixel-pairs. For each pixel-pair set, there are 160 produced reference-bits, and we call them as a reference-bit group. The number of reference-bit groups is also $N/64$.

2. Divide the original image into $N/64$ non-overlapped blocks sized 8×8. In each block, we pseudo-randomly select 160 positions from the 192 bits in the 3 LSB-layers according to the secret key. The total number of selected LSB is also $5N/2$. In order to enhance security, the LSB selections in different blocks should be mutually different. After mapping the reference-bit-groups to the blocks in a one-to-one manner, replace the original bits at the 160 selected positions in each block with the reference-bits in the corresponding group.
3. Randomly generate 32 bits $I(1)$, $I(2)$, ..., $I(32)$ for the host image and call them as image-ID-bits. For each block, we collect the 320 original bits in the 5 MSB-layers, and the 160 reference-bits used to replace the selected LSBs. Also, we convert the row index of the block i and the column index of the block j into 64 bits and call them position-bits ($1 \leq i \leq N_1/8$, $1 \leq j \leq N_2/8$). Then, feed the 320 MSBs, 160 reference-bits and 64 position-bits into a hash function to generate 32 hash-bits, and denote them as $h_{i,j}(1)$, $h_{i,j}(2)$, ..., $h_{i,j}(32)$. Here, the hash function must have the property that any change on an input would result in a completely different output. Calculate 32 localization-bits for the block,

$$l_{i,j}(m) = h_{i,j}(m) \oplus I(m)\ , \quad m = 1,\ 2,\ \ldots,\ 32 \tag{3}$$

Put the localization-bits into the 32 remaining LSB positions in the block. At last, combine the original MSBs and the substituted LSBs to produce a watermarked image.

In watermark embedding procedure, the pixel pairs are pseudo-randomly generated and the embedding positions of the reference bits of pixel pairs are also pseudo-randomly determined. That means the two pixels in a same pair perhaps come from different areas and their reference bits may be embedded into another area. To ensure security, a number of operations are dependent on the secret key. Actually, we may use a primary secret key to generate a pseudo-random sequence, and regard it as a series of pseudo-keys to directly control pixel permutation and position selection. For permuting N pixels, we take a portion of the pseudo-random sequence with length N, and sort it. Then, the sorting order can be used as a way of pixel permutation, and the numbers of possible ways of pixel permutation is 2^N. With sufficiently large N, it is virtually impossible to perform a successful attack. Similarly, for selecting 160 positions from 192 LSB, we may take different portions with a length of 192 from the pseudo-random sequence for different blocks. The 160 smallest values indicate the selected positions, and the selected positions in different blocks are mutually different.

In the watermark embedding procedure, 5 MSBs of the cover image are preserved and 3 LSBs replaced with the reference-bits and localization-bits. Assuming that the original distribution of the 3 LSBs is uniform, the average energy of distortion caused by watermarking on each pixel is

$$E_{\mathrm{D}} = \frac{1}{64} \cdot \sum_{u=0}^{7} \sum_{v=0}^{7} (u - v)^2 \tag{4}$$

So, the approximate PSNR is

$$\mathrm{PSNR} \approx 10 \cdot \log_{10}\left(255^2 / E_{\mathrm{D}}\right) = 37.9 \text{ dB} \tag{5}$$

In general, the distortion caused by watermark embedding can not be detected by human visual system. Actually, the more the data in original image is kept, the more imperceptibility can be achieved, but the watermark capacity is lower. To make a reasonable tradeoff, we let the 5 MSB be preserved in the proposed scheme.

3 Procedure of content restoration

Suppose an adversary alters some content of the watermarked image without changing the image size. Having received a suspicious image, we first identify the tampered blocks, and then recover the MSBs in the tampered area by using the reference data extracted from other blocks and the spatial correlation of the natural image.

The first stage is to locate the tampered blocks. After dividing the received image into non-overlapped 8×8 blocks, we select 160 positions from the 3 LSB-layers in each block using the same secret key. For each block, feed its 320 bits in the 5 MSB-layers, 160 bits in 3 LSB-layers at the selected positions and 64 position-bits of the block into the hash function to obtain 32 hash-bits $h_{i,j}(1)$, $h_{i,j}(2)$, ..., $h_{i,j}(32)$, and extract the 32 bits in 3 LSB-layers at the rest positions $l_{i,j}(1)$, $l_{i,j}(2)$, ..., $l_{i,j}(32)$. Calculate the image-ID-bits for each block

$$I_{i,j}(m) = h_{i,j}(m) \oplus l_{i,j}(m)\ , \quad m = 1, 2, \ldots, 32 \tag{6}$$

If the image has not been tampered, the calculated image-ID-bits of all blocks should be identical with the original image-ID-bits. Otherwise, if some blocks are modified or moved from another position/image as in [3], the calculated image-ID-bits will be changed. Although the original image-ID-bits is unknown at authentication side, we can compare the image-ID-bits calculated from different blocks, and judge the blocks possessing the identical calculated image-ID-bits as “not tampered”. For the other blocks, a “tampered” decision is made. Here, a block without any tampering must be judged as “not tampered”. Probability with which a block containing modified contents or moved from another position/image is falsely judged as “not tampered” is 2^{-32}. False judgment is therefore virtually impossible. That means any modification on MSB or LSB will result in a “tampered” decision.

Denoting the ratio between the numbers of tampered blocks in an image and that of all blocks as α, we can recover the MSB of pixels in tampered blocks. We consider the following three cases.

Case 1: For a pixel in the tampered area, if another pixel belonging to the same pair is located in “not tampered” area and the five reference-bits derived from the pair are also embedded into “not tampered” area, we can retrieve the original MSBs of the pixel without any error. Actually, probability for thus a case to occur is

$(1-\alpha)^2$. In this case, the 5 MSBs of another pixel and the 5 reference-bits can be obtained from the received image. Denoting the two pixels in the pair as p_i and p_j, if p_i is in tampered area and p_j is another pixel, we calculate 5 MSBs of p_i

$$b_{i,7-m} = r_m \oplus b_{j,m+3}, \quad m = 0, 1, \ldots, 4 \tag{7}$$

where r_m is the 5 reference-bits, or calculate 5 MSBs of p_j if it is in tampered area,

$$b_{j,m+3} = r_m \oplus b_{i,7-m} , \quad m = 0, 1, \ldots, 4 \tag{8}$$

Case 2: When both pixels in a pair are in the tampered area and the corresponding reference-bits are hidden in “not tampered” area, we can estimate their MSBs in the following way. For a pixel in tampered area, probability for this case to occur is $(1-\alpha)\cdot\alpha$. After extracting the 5 reference-bits from the received image, a restriction to the MSBs of the two pixels is given. For example, if $(r_0\ r_1\ r_2\ r_3\ r_4)=(10110)$, the MSBs of the two pixels, p_i and p_j, must satisfy

$$b_{i,7} \neq b_{j,3}, \quad b_{i,6} = b_{j,4}, \quad b_{i,5} \neq b_{j,5}, \quad b_{i,4} \neq b_{j,6}, \quad b_{i,3} = b_{j,7} \tag{9}$$

Decimal values of the MSBs can be written as

$$v_i = \sum_{m=3}^{7} b_{i,m} \cdot 2^m, \quad v_j = \sum_{m=3}^{7} b_{j,m} \cdot 2^m \tag{10}$$

There are all together 32 possible combinations of (v_i, v_j) according to the restriction. Denote a set including the “not tampered” area and the pixels recovered in Case 1 as Set 1, the pixel closest to p_i in Set 1 as q_i, and the pixel closest to p_j in Set 1 as q_j. For each possible (v_i, v_j), calculate

$$D = (v_i - u_i)^2 + (v_j - u_j)^2 \tag{11}$$

where u_i and u_j are decimal values of the 5 MSBs of q_i and q_j. Because of the spatial correlation in a natural image, the value of D corresponding to the original (v_i, v_j) would be small. Since the relationship between the higher MSB of a pixel and the lower MSB of another one is restricted by the reference-bits, for the other candidates of (v_i, v_j), there must be some higher MSB different from the original, leading to the large values of D. Then, we find the minimal one among the 32 values of D, and use the corresponding (v_i, v_j) as the recovered result of MSB of p_i and p_j. In other words, both the restriction condition derived from the reference-bits and the spatial correlation in the natural image are used to recover the original content in this case.

Case 3: The reference-bits of a pair containing one or two pixels in the tampered area are stored in the tampered area. For a pixel in tampered area, probability of this case is α. The reference-bit cannot be extracted from the received image in this case. So, we exploit only the spatial correlation to recover the MSBs of missing pixels. Denote the missing pixel in the pair as p_i or p_j, a set including the “not tampered” area and the pixels recovered in Cases 1 and 2 as Set 2, the pixel closest to p_i in Set 2 as g_i, and the pixel closest to p_j in Set 2 as g_j. Then, the MSBs of g_i and g_j are used as the recovered MSBs of p_i and p_j, respectively.

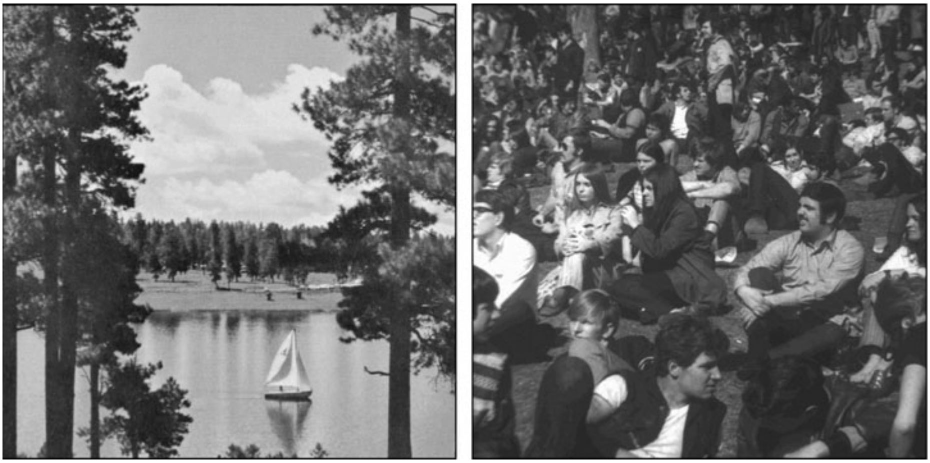

Fig. 1 Watermarked images Lake and Crowd

In this scheme, if the tampered area is small, most pixels in the tampered area belong to the first case, and their MSBs can be correctly restored according to the extracted reference data. Moreover, the correctly-restored pixels can provide sufficient information to estimate the original MSBs of the other missing pixels by exploiting the spatial correlation. So, the recovered result is very close to the original content. On the other hand, if the tampered area is extensive, image recovery can still be performed with compromised quality.

4 Experimental results

Using two test images Lake and Crowd sized 512×512 as the covers, Fig. 1 gives their watermarked versions. PSNR values due to watermark embedding are respectively 37.9 dB and 37.8 dB, confirming the theoretical result in (5), and the distortion is imperceptible. We modified the watermarked images by replacing the original content with fake information.

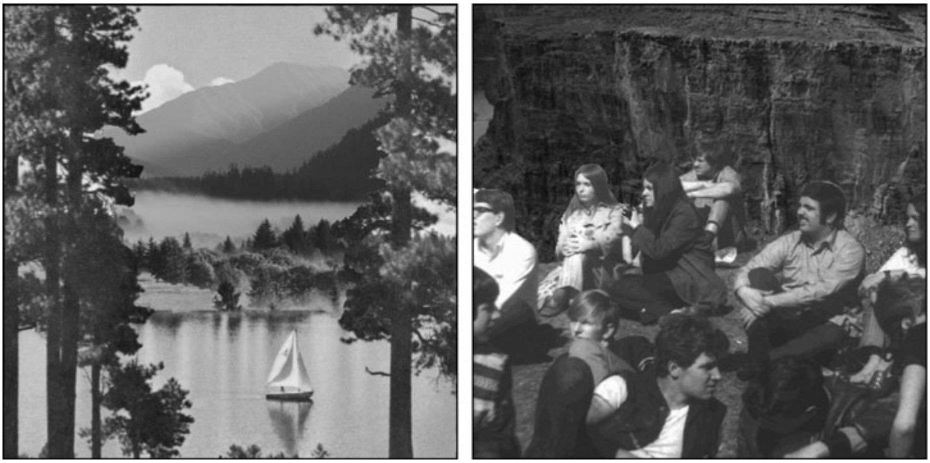

Fig. 2 Tampered versions

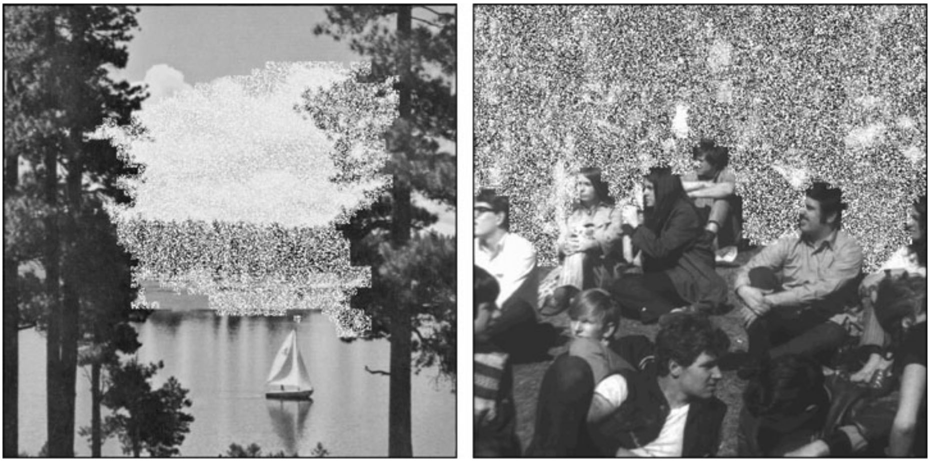

Fig. 3 “Not tampered” content and recovered pixels in the first case

The tampered images are shown in Fig. 2, and the tampering rates are 28.7% and 45.4%, respectively. In tampered-block identification, all the blocks containing fake content were found. Figure 3 gives the results when the MSB of the pixels in the first case were recovered, and the other pixels in the tampered area are represented by extreme white. It can be seen the pixels in this case distribute in the entire tampered area. With the values of the retrieved pixels, the spatial correlation can be employed to recover their surrounding tampered pixels. The final restored images are shown in Fig. 4, and PSNR of the recovered content calculated only in the tampered area are 30.2 dB and 27.1 dB, respectively.

If the adversary alters the original content of a watermarked image with different tampering rates, qualities of the recovered results in the tampered area vary. As mentioned above, the probabilities of the tampered pixels belonging to cases 1, 2 and 3 are $(1-\alpha)^2$, $(1-\alpha)\cdot\alpha$ and α, respectively. In fact, the differences between the actual proportions and the theoretical probabilities are less than 4%. Table 1 shows PSNR of the recovered content in

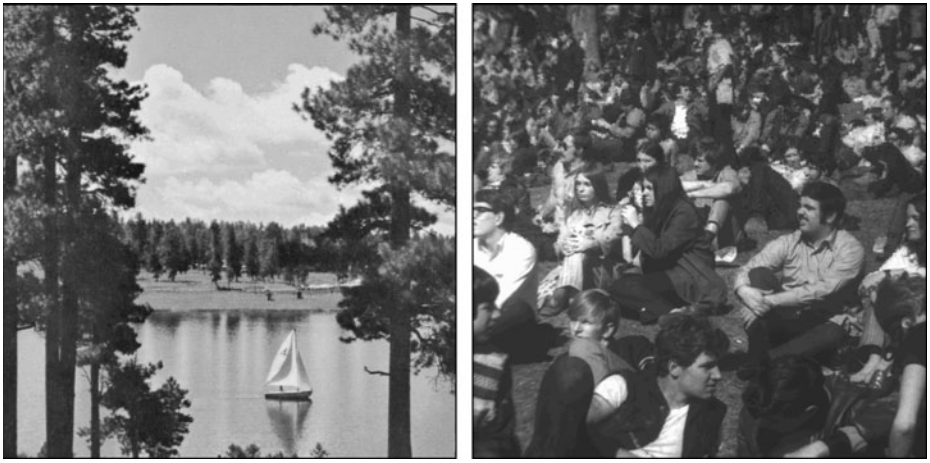

Fig. 4 Restored images

Table 1 PSNR of recovered content in the tampered area with different tampering rates

		Tampering rate							
		5%	11%	19%	26%	33%	40%	48%	54%
Cover image	Lena	37.68	35.39	33.35	32.06	31.03	29.44	28.09	26.70
	Crowd	38.07	35.81	33.15	31.40	29.78	28.23	26.69	25.56
	Lake	36.02	34.49	32.87	31.35	30.05	28.12	26.32	24.88
	Baboon	33.42	32.35	30.83	29.36	27.92	25.70	23.68	22.55

the tampered area with respect to the tampering rates. Four gray images sized 512×512 were used as the covers. Experiment on other images provides similar results. It has been observed that, even if the tampering rate is greater than 50%, the recovered content has reasonable quality. As described previously, the spatial correlation of natural image is exploited to estimate the original MSB of missing pixels, so that the smoother content can be restored with better quality. The rank of PSNR values in Table 1 is consistent with the fluctuation magnitudes of the four test images.

Table 2 compares several fragile watermarking schemes with restoration capability. The methods in [1] and [8] suffer from the tampering coincidence problem and the method in [7] does not work with a large tampering rate. By using the proposed scheme, the original content in an extensive area can be recovered and the tampering coincidence problem is avoided. Also, the lower the tampering rate is, the better quality of restored content can be obtained. That means the proposed scheme is more flexible than the previous methods.

5 Conclusions

In the scheme proposed here, the LSBs of a host image are replaced with the reference data and localization data derived from the cover. In the authentication, the localization data are used to locate the blocks containing substitute information, and the reference data and the spatial correlation are used to recover the original content of tampered area. The smaller the tampered area, the MSBs of more pixels in the tampered area can be correctly recovered. For the other pixels in the tampered area, the MSBs are estimated according to their neighbors. In this way, since the reference data are scattered in the entire image, the tampering coincidence problem is overcome. And the principal content can still be

Table 2 Comparison of restoration capability among several fragile watermarking schemes

Watermarking scheme	PSNR due to watermarking	PSNR of recovered content in the tampered area	Condition of restoration
Method 1 in [1]	43.8 dB	21.5 dB	Regions storing the original information of tampered areas must be reserved.
Method 2 in [1]	33.1 dB	28.8 dB	
Method in [8]	36.7 dB	22.8 dB	
Method in [7]	37.9 dB	37.9 dB	Tampering rate<6.6%
Proposed scheme	37.9 dB	[22, 38] dB	Tampering rate<54%

recovered even if the tampered region is extensive, with the quality of recovered content decreasing with the increasing extent of tampering.

Acknowledgments This work was supported by the Natural Science Foundation of China (60872116, 60832010, and 60773079), the Shanghai Rising-Star Program (10QH14011), and the Key Scientific Research Project of Shanghai Education Committee (10ZZ59).

References

1. Fridrich J, Goljan M (1999) Images with self-correcting capabilities. Proceeding of IEEE International Conference on Image Processing, 792–796
2. He H, Zhang J, Chen F (2009) Adjacent-block based statistical detection method for self-embedding watermarking techniques. Signal Process 89:1557–1566
3. Holliman M, Memon N (2000) Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. IEEE Trans Image Process 9:432–441
4. Suthaharan S (2004) Fragile image watermarking using a gradient image for improved localization and security. Pattern Recognit Lett 25:1893–1903
5. Wong PW, Memon N (2001) Secret and public key image watermarking schemes for image authentication and ownership verification. IEEE Trans Image Process 10:1593–1601
6. Zhang X, Wang S (2007) Statistical fragile watermarking capable of locating individual tampered pixels. IEEE Signal Process Lett 14:727–730
7. Zhang X, Wang S (2009) Fragile watermarking scheme using a hierarchical mechanism. Signal Processing 89:675–679
8. Zhu X, Ho A, Marziliano P (2007) A new semi-fragile image watermarking with robust tampering restoration using irregular sampling. Signal Process Image Commun 22:515–528

Xinpeng Zhang received the B.S. degree in computational mathematics from Jilin University, China, in 1995, and the M.E. and Ph.D. degrees in communication and information system from Shanghai University, China, in 2001 and 2004, respectively. Since 2004, he has been with the faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. His research interests include information hiding, image processing and digital forensics. He has published more than 100 papers in these areas.

Shuozhong Wang received B.S. degree in 1966 from Peking University, P.R. China, and Ph.D. degree in 1982 from University of Birmingham, England. He was with Institute of Acoustics, Chinese Academy of Sciences, from January 1983 to October 1985 as research fellow, and joined Shanghai University of Technology in October 1985 as associate professor. He is now professor of School of Communication and Information Engineering, Shanghai University. He was associate scientist at Department of EECS, University of Michigan, USA, from March 1993 to August 1994, and a research fellow at Department of Information Systems, City University of Hong Kong, in 1998 and 2002. His research interests include underwater acoustics, image processing, and multimedia security. He has published more than 150 papers in these areas. Many of his research projects are supported by the Natural Science Foundation of China.

Zhenxing Qian received the B.S. degree in 2003 and the Ph.D. degree in 2007 from University of Science & Technology of China (USTC). Since 2009, he has been with the faculty of the School of Communication and Information Engineering, Shanghai University. His research interests include data hiding, image processing, and digital forensics.

Guorui Feng received the B.S. and M.S. degree in computational mathematics from Jilin University, China, in 1998 and 2001 respectively. He received Ph.D. degree in Electronic Engineering from Shanghai Jiaotong University, China, 2005. From January 2006 to December 2006, he was an assistant professor in East China Normal University, China. During 2007, he was a research fellow in Nanyang Technological University, Singapore. Now he is with the School of Communication and Information Engineering, Shanghai University, China. His current research interests include image processing, hiding information and computational intelligence.