International Journal of Computer Sciences and Engineering Systems
Vol. 11 No. 2 (December, 2017)

Difference Image Watermarking Based Reversible Image Authentication with Tampering Localization Capability

De-Wen XUE[1,2] and Zhe-Ming LU[1,2,3]

[1]Guangdong Electronics Industry Institute, NO.10 Building, SongShan Lake Productivity Promotion BaseDongguan 518055, P. R. China, *E-mail: xdewen@yahoo.com.cn*

[2]Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School Shenzhen, 518055, P. R. China, *E-mail: xdewen@yahoo.com.cn*

[3]Media Processing and Communication Lab, Sun-Yat-Sen Univ. East CampusGuangzhou, 510006, P. R. China *E-mail: zhemingl@yahoo.com*

Abstract: *In this paper, we present a new reversible image authentication scheme based on watermarking with the capability of tampering localization. The proposed technique shifts slightly the histogram of the difference image generated from the original cover image, thus modifies pixel values tenderly to obtain a high embedding capacity. Watermark data is blockwisely embedded so the tampered positions in a forgery image can be blockwisely detected wherever any disturbance is imposed. The distortion caused by embedding can be completely removed after the hidden data has been extracted if the marked image is authentic. Experimental results demonstrate that the PSNR values of watermarked images are very high (>51.14 dB) while the computation complexity is low.*

Keywords: Reversible Watermarking, Difference Image, Histogram Modification, Tampering Localization

1. INTRODUCTION

Increasing interest is drawn in reversible watermarking based image authentication schemes in recent years, especially for some crucial applications such as digital watermarking systems involved with image authentication in the fields of medical and military imaging, remote sensing, as well as law enforcement, where bit-by-bit exactness of the recovered cover image in comparison with the original content is desired or even mandatory after watermark is extracted successfully. Watermarking schemes satisfying this reversibility requirement are usually referred to as reversible, lossless, distortion-free, or invertible watermarking schemes.

Research on reversible watermarking stems from an application of image authentication in the patent by Honsinger *et al.* [1], in which modulo addition and a robust spatial additive watermark are utilized to realize the reversibility of the watermarking process. Soon another reversible data embedding scheme was proposed by Fridrich, J. *et al.* [2], in which the saved space of some losslessly compressed bit-planes in the spatial domain is used to embed the hiding data. Several other reversible watermarking schemes with comparatively higher capacity were proposed in [3] and [4]. Ref. [3] tags different pixel groups individually according to the watermark bits to be embedded. Ref. [4] embeds data in the middle bit-planes of the integer wavelet transform coefficients and a pre-processing scheme is designed to prevent the pixel overflow and underflow caused by the modification of wavelet coefficients. Tian [5] proposed a reversible embedding approach based on difference expanding technique. This approach exerts the high redundancy characteristics among neighboring pixels by Haar wavelet transform and obtains much improvement on embedding capacity. Furthermore, in Tian's approach, due to the simple form of Haar wavelet, the pixel error caused by the difference value expansion is easier to be estimated than in [4]. Later, histogram modification stepped on watermarking stage, attracting increasing attention for its natural dominance in the performance with respect to visual quality of the marked image. Ref. [6] selects peak points and shifts values in image's histogram to embed the data in the peak points and neighboring points. Based on [6], Ni et al. [7] utilizes the minimum points of the histogram to reach a better quality (PSNR > 48 dB) and a high capacity. However, these schemes are still not satisfactory for some cases which need even higher capacity and desirable quality for watermarked images. Based on [7], a new reversible watermarking approach with better PSNR performance was proposed by Lee et al. [8], where the concept of difference image is brought out, and the visual quality is improved to a higher level(PSNR > 51 dB).

As abovementioned, it is no doubt research work on reversible watermarking techniques has progressed through a long way, yet further applications' requirements move

faster, of which one general requirement is tampering localization. i.e., under certain circumstances, apart from the reversibility, watermarking scheme is also demanded to detect the attack position, so that even the marked image has been partially tampered, the rest of the image is still authentically useful if exact tampering position is detected by the watermarking system. Though performance has reached to quite high level in reversible watermarking schemes above mentioned, they are far from satisfactory with respect to tampering localization. Baušys et al. [9] proposed a reversible watermarking scheme for image authentication in frequency domain, where the localization capability is considerable, yet the visual quality is comparably lower (PSNR<38dB). Wu [10] proposed another tampering localization capable reversible watermarking scheme based on histogram shifting of integer wavelet coefficients, of which the performance is better. Nevertheless, its quality is still undesirable in some strict scenarios.

In this paper, we propose a new reversible image watermarking scheme with the capability of tampering localization based on the difference image approach proposed in [8]. Watermark data is blockwisely embedded so as to localize tampering action made on the marked image, which is not available in [8], whereas the performance is nearly the same as that in [8](PSNR> 51dB), which is much better than those in [9] and [10]. It is also noted that the computational execution of our scheme in spatial domain is much simpler compared with those in frequency domain in [9] and [10]. We review some relevant knowledge on the approach proposed in [8] in section 2, and then propose our scheme, theoretical analysis and the simulation results in section 3, 4 and 5, respectively. Conclusion is drawn in Section 6.

2. RELEVANT KNOWLEDGE

The difference image based reversible watermarking technique in [8] is shown in Fig.1, and briefly described as follows.

2.1 Embedding Scheme

Subheadings should be in a bold font and in lower case with initial capitals. They should start at the left-hand margin on a separate line.

For a grayscale image $I(i, j)$ of size $M \times N$ pixels, the difference image $D(i, j)$ of size $M \times N / 2$ is generated from the original image. For $0 \le i \le M - 1$ and $0 \le j \le N / 2 - 1$,

$$D(i, j) = I(i, 2j + 1) - I(i, 2j) \quad (1)$$

where $I(i, 2j + 1)$ and $I(i, 2j)$ are the odd-line field and the even-line field, respectively. For watermark embedding, the histogram bins of –2 and 2 are emptied by shifting some pixel values in the difference image. If the difference value is greater than or equal to 2, one is added to the odd-line pixel. If the difference value is less than or equal to -2, one is subtracted from the odd-line pixel. Then, the modified difference image $\tilde{D}(i, j)$ can be represented as

$$\tilde{D}(i, j) = \tilde{I}(i,2j+1) - I(i,2j) \quad (2)$$

where

$$\tilde{I}(i,2j+1) = \begin{cases} I(i,2j+1)+1 & \text{if } D(i,j) \ge 2 \\ I(i,2j+1)-1 & \text{if } D(i,j) \le -2 \\ I(i,2j+1) & \text{Otherwise} \end{cases}$$

In the histogram modification process, the watermark is embedded into the modified difference image. The modified difference image is scanned. Once a pixel with the difference value of -1 or 1 is encountered, the watermark to be embedded is checked. If the bit to be embedded is 1, the difference value of -1 is moved to -2 by subtracting one from the odd-line pixel, or 1 to 2 by adding one to the odd-line pixel. If the bit to be embedded is 0, the pixel of the difference image is skipped until a pixel with the difference value –1 or 1 is encountered. Thus, the watermarked old-line field $Iw(i, 2j + 1)$ is obtained as follows: If $W(u, v) = 1$ and $\tilde{D}(i, j) = 1$ or -1,

$$Iw(i,2j+1) = \begin{cases} \tilde{I}(i,2j+1)+1 & \text{if } \tilde{D}(i,j)=1 \\ \tilde{I}(i,2j+1)-1 & \text{if } \tilde{D}(i,j)=-1 \end{cases} \quad (4)$$

and in all other cases,

$$Iw(i,2j+1) = \tilde{I}(i,2j+1) \quad (5)$$

and the watermarked even-line fields $Iw(i, 2j)$ is given by

$$Iw(i, 2j) = I(i, 2j) \quad (6)$$

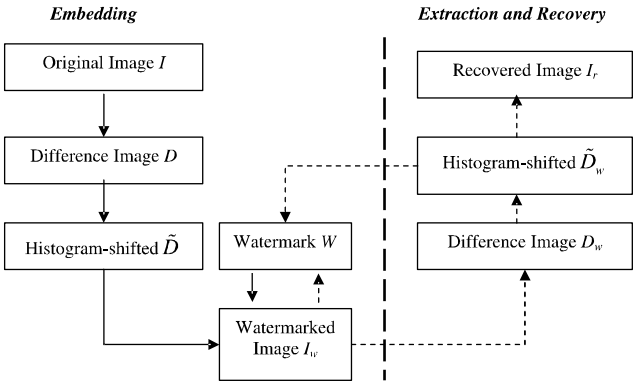


Figure 1: Reversible Watermarking Scheme based on Difference Image

2.2 Extraction and Recovery Scheme

As depicted in Fig.1, watermark extraction and recovery scheme is just the inverted version of that in embedding stage.

For the received watermarked image $Iw(i, j)$ of size $Mw \times Nw$ pixels, the difference image $Dw(i, j)$ of size $Mw \times Nw/2$ pixels is calculated. The whole difference image is scanned. The authentication watermark $We(m, n)$ is extracted as follows:

$$We(m,n)=\begin{cases}0 \text{ if } De(i,j)=\text{-1 or 1}\\ 1 \text{ if } De(i,j)=\text{-2 or 2}\end{cases} \quad (7)$$

Simultaneously, the watermarked image is reversed back to the original image by shifting some pixel values in the difference image. The whole difference image is scanned once again. The recovered odd-line field $Ir(i,2j+1)$ is expressed as

$$Ir(i,2j+1)=\begin{cases}Iw(i,2j+1)+1 & \text{if } Dw(i,j)\le -2\\ Iw(i,2j+1)-1 & \text{if } Dw(i,j)\ge 2\\ Iw(i,2j+1) & \text{Otherwise}\end{cases} \quad (8)$$

Since only pixel values of the odd-line field are manipulated in the watermark embedding process, the recovered even-line field $Ir(i, 2j)$ is obtained by

$$Ir(i,2j)=Iw(i,2j) \quad (9)$$

Above stated is the difference image based reversible watermarking scheme mentioned in [8]. We can see, effective as the scheme is, it is nowhere for the attack localization function.

3. PROPOSED SCHEME

The proposed reversible watermarking scheme capable of blockwise tampering localization is depicted in Fig. 2, and described in detail as follows.

3.1 Pre-Processing

For a grayscale image I of size $M \times N$ pixels, we first arbitrarily partition I into b sub block images $I_1, I_2,\ldots I_b$,which are supposed to be individual ROI (regions of interest),and thus bear individual meanings for tampering localization.

Ideally speaking, even to the same one image, according to the particular focus on different ROI, block number b and the size of the sub block images vary from user to user.

Yet practically speaking, without loss of generality, here we default that the partitioning process is equally operated on I from upper left to bottom right, i.e., the sizes of the b sub block images are equal, and I_1 corresponds to the upper left part of I, whereas I_b the bottom right part, just for the convenience of testing experiment.

3.2 Embedding Scheme

With the b sub block images $I_1, I_2,\ldots I_b$ of size $m \times n$ ($m \times n \times b = M \times N$) obtained in step A, we proceed to generate their corresponding difference images $D_1, D_2\ldots D_b$, of size $m \times n/2$ respectively, according to the scheme mentioned in section 2. For $0 \le i \le m-1$ and $0 \le j \le n/2-1$, $D_1, D_2\ldots D_b$ can be obtained by,

$$\begin{aligned}D_1(i,j)&=I_1(i,2j+1)-I_1(i,2j)\\ D_2(i,j)&=I_2(i,2j+1)-I_2(i,2j)\\ &\cdots\\ D_b(i,j)&=I_b(i,2j+1)-I_b(i,2j)\end{aligned} \quad (10)$$

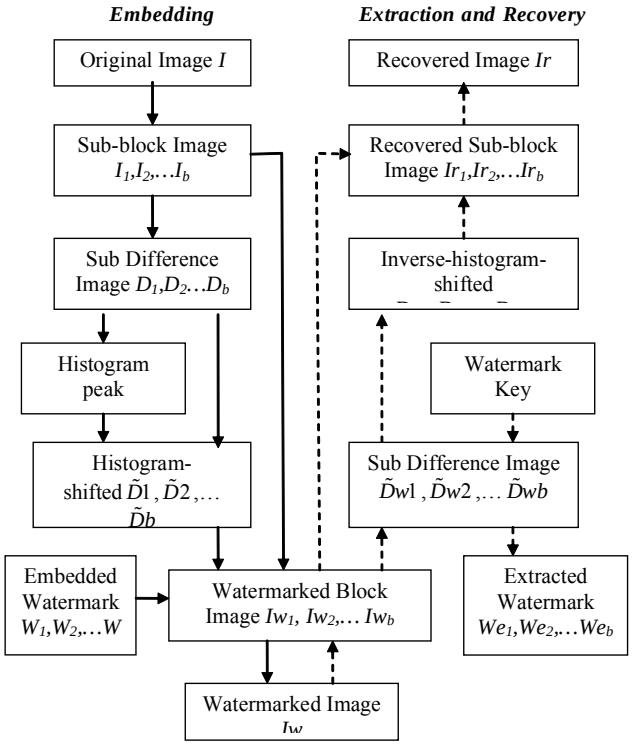


Figure 2: Proposed Watermark Embedding and Extracting Scheme

where $I_t(i, 2j+1)$ and $I_t(i, 2j)$ are the odd-line field and the even-line field of $I_t(i, j)$, respectively ($1 \le t \le b$).

For the obtained $D_t(i, j)$, we proceed to find the pixel point p_t, where the histogram of $D_t(i, j)$ appears to be the peak value.

Next we begin to consider blockwisely embedding the b candidate watermarks $W_1, W_2,\ldots W_b$, into the b corresponding sub block cover images $I_1, I_2,\ldots I_b$.

Consider embedding watermark W_t into sub block cover image $I_t(i, j)$, the histogram bins of $D_t(i, j)$ at $-(p_t+1)$ and (p_t+1) are emptied by shifting some pixel values in the difference image. If the difference value is greater than or equal to (p_t+1), one is added to the odd-line pixel. If the difference value is less than or equal to $-(p_t+1)$, one is subtracted from the odd-line pixel. Then, the modified difference image $\tilde{D}_t(i, j)$ can be represented as

$$\tilde{D}_t(i,j)=\tilde{I}_t(i,2j+1)-I_t(i,2j) \quad (11)$$

where

$$\tilde{I}_t(i,2j+1)=\begin{cases}I_t(i,2j+1)+1 & \text{if } D_t(i,j)\ge p_t+1\\ I_t(i,2j+1)-1 & \text{if } D_t(i,j)\le -(p_t+1)\\ I_t(i,2j+1) & \text{Otherwise}\end{cases} \quad (12)$$

In the histogram modification process, the watermark is embedded into the modified difference image. The modified difference image is scanned. Once a pixel with the difference value of $-p_t$ or p_t is encountered, the watermark

to be embedded is checked. If the bit to be embedded is 1, the difference value of $-p_t$ is moved to $-(p_t + 1)$ by subtracting one from the odd-line pixel or p_t to $p_t + 1$ by adding one to the odd-line pixel. If the bit to be embedded is 0, the pixel of the difference image is skipped until a pixel with the difference value $-p_t$ or p_t is encountered. Thus, the watermarked old-line field $Iw_t(i, 2j+1)$ is obtained as follows: If $W_t(u, v) = 1$ and $\tilde{D}_t(i, j) = p_t$ or $-p_t$,

$$Iw_t(i,2j+1) = \begin{cases} \tilde{I}_t(i,2j+1)+1 & \text{if } \tilde{D}_t(i,j) = p_t \\ \tilde{I}_t(i,2j+1)-1 & \text{if } \tilde{D}_t(i,j) = -p_t \end{cases} \quad (13)$$

and in all other cases,

$$Iw_t(i,2j+1) = \tilde{I}_t(i,2j+1) \quad (14)$$

and the watermarked even-line fields $Iw_t(i, 2j)$ is given by

$$Iw_t(i,2j) = I_t(i,2j) \quad (15)$$

After all Iw_t $(1 \le t \le b)$ are generated by embedding W_t into I_t according to above stated process, we then construct the whole watermarked image *Iw* by combining the obtained marked sub block images Iw_1, Iw_2,… Iw_b, according to the reverse partitioning order.

It is noted that the peak points $p_1, p_2,\ldots p_b$ together with the partitioning order, are saved as the watermarking keys.

3.3 Extraction and Recovery Scheme

As is depicted in Fig. 2, watermark extraction and recovery scheme is just the inverted version of that in embedding stage.

For the received watermarked image $Iw(i, j)$ of size $M \times N$ pixels, we first repartition $Iw(i, j)$ into b sub marked block images Iw_1, Iw_2,…Iw_b, following the same order as embedding process.

Next we proceed to extract all watermarks $We_1, We_2,\ldots We_b$ out of the sub marked block images Iw_1, Iw_2,…Iw_b, under the watermarking keys $p_1, p_2,\ldots p_b$.

For the sub watermarked block image $Iw_t(i, j)$ of size $m_w \times n_w$ pixels, the difference image $Dw_t(i, j)$ of size $m_w \times n_w / 2$ pixels is calculated. The whole difference image is scanned. The authentication watermark $We_t(u, v)$ is extracted as follows:

$$We_t(m,n) = \begin{cases} 0 & \text{if } Dw_t(i,j) = -p_t \text{ or } p_t \\ 1 & \text{if } Dw_t(i,j) = -(p_t+1) \text{ or } p_t+1 \end{cases} \quad (16)$$

while the watermarked pixel points is reversely modified, and transitional recovered image $\tilde{I}r_t(i, j)$ is obtained as

$$\tilde{I}r_t(i,2j+1) \begin{cases} Iw_t(i,2j+1)+1 & \text{if } Dw_t(i,j) = -(p_t+1) \\ Iw_t(i,2j+1)-1 & \text{if } Dw_t(i,j) = p_t+1 \\ Iw_t(i,2j+1) & \text{otherwise} \end{cases} \quad (17)$$

$$\tilde{I}r_t(i,2j) = Iw_t(i,2j) \quad (18)$$

Correspondingly, $\tilde{D}r_t(i, j)$ is obtained as the difference image of $\tilde{I}r_t(i, j)$.

Finally, the sub watermarked image is reversed back to the original sub block image by shifting some pixel values in the difference image $\tilde{D}r_t(i, j)$. The whole difference image is scanned. The recovered odd-line field $Ir_t(i, 2j+1)$ is expressed as

$$Ir_t(i,2j+1) = \begin{cases} \tilde{I}r_t(i,2j+1)+1 & \text{if } \tilde{D}r_t(i,j) \le -(p_t+2) \\ \tilde{I}r_t(i,2j+1)-1 & \text{if } \tilde{D}r_t(i,j) \ge p_t+2 \\ \tilde{I}r_t(i,2j+1) & \text{Otherwise} \end{cases}$$

Since only pixel values of the odd-line field are manipulated in the watermark embedding process, the recovered even-line field $Ir_t(i, 2j)$ is obtained by

$$Ir_t(i,2j) = \tilde{I}r_t(i,2j) \quad (20)$$

After all Ir_t $(1 \le t \le b)$ are recovered by extracting We_t out of Iw_t according to above stated process, we then construct the whole recovered image *Ir* by combining the obtained extracted sub block images Ir_1, Ir_2,… Ir_b, according to the reverse partitioning order.

If the watermarked image *Iw* is authentic and suffers no tampering, then the extracted watermarks $We_1, We_2,\ldots We_b$ are just the same as the original watermarks W_1, W_2,…W_b, while the restored cover image *Ir* is identical to the original cover image *I*.

3.4 Tampering Localization

Blockwisely, the proposed scheme can effectively facilitate the localization of tampering operations via merely inspecting the extracted watermarks.

If the received watermarked image *Iw* is authentic and no alteration happens during communication process, then the watermarks extracted should be just the some ones as before embedding.

If we observe part of the extracted watermarks are different from their counterparts of original watermarks, then the corresponding blocks of the cover image are asserted to have suffered tampering operations, yet the rest parts of the cover image are still authentically useful to the receiver.

If all of the extracted watermarks are different from their counterparts of original watermarks, then the whole cover image are asserted to have suffered tampering operations globally, or even, the received image is merely a forgery.

4. THEORETICAL ANALYSIS

4.1 Visual Quality

We assume that there is no pixel with overflow and underflow in the original image (as is the common case for normal images). Worst speaking, all pixels of the odd-line field are added or subtracted by 1, when, correspondingly, the mean squared error of this case is only 0.5. Thereby, the PSNR of the watermarked image is calculated as

$$PSNR = 10\log_{10}\left(\frac{255^2}{0.5}\right) \approx 51.141\,(\text{dB}) \qquad (21)$$

In words, the bottom bound of the PSNR of the watermarked image versus the original one is about 51.141 dB, which is way higher than that of other lossless data hiding schemes.

4.2 Embedding Capacity

As we embed watermarks into the peak pixel points p_1, p_2,…$p_{b,}$ of the histogram of difference images D_1, D_2…D_b of sub cover images I_1, I_2,…I_b, the embedding capacity can be calculated as

$$Capacity = mp_1 + mp_2 + \ldots + mp_b \qquad (22)$$

Where mp_1, mp_2…mp_b are the histogram counts corresponding pixel value p_1, p_2,…p_b in the difference image $D(i,j)$. The capacity is desirable compared to other reversible watermarking schemes with the capability of tampering localization.

4.3 Tampering Localization Precision

As has been discussed in section 3, the proposed scheme can blockwisely detect the attacks operated on the watermarked image. The localization precision is dependent on the size of the sub blocks utilized in the scheme, which relies on the perspective on ROI (regions of interest) of the user. In the utmost case, every two point can be deemed as the sub block cover image. Thus, the utmost localization precision can reach as high as double pixels.

5. EXPERIMENTAL RESULTS AND COMPARISON

To observe the performance of the proposed authentication scheme, we perform computer simulations on standard test images of size 512 × 512 pixels. Without loss of generality, we choose $b = 4$, just for the convenience of experiment.

Fig. 3 display the 4 candidate watermark images, 4 binary logo images of size 64×64 pixels, 128×46 pixels, 58×58 pixels, 128×46 pixels, together they add up to as equivalent to a binary sequence of 19,236 bits.

It is noted that the real capacity is much higher than the addition value of the sizes of the candidate watermark images. Yet the candidate images' sizes are depended on the choice of the users, if only they are within the bound of the peak point p_t.

Fig. 4 shows the watermarked test Lena image with the 4 candidate authentication watermarks added into its 4 corresponding parts via the proposed scheme.

Figure 3: Candidate Watermark Images

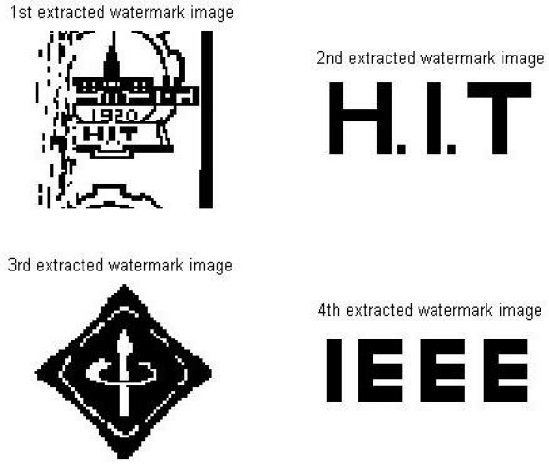
Figure 4: Watermarked Lena Image

It can be seen from the figure that since pixel values are only gently modified in watermark embedding process, there is barely any visible degradation caused in the watermarked image. The PSNR value is 51.827 dB.

Fig. 5 (a), (b) shows respectively a tampered watermarked Lena image with its upper left 50□~50 region attacked by random noise, and the 4 candidate watermarks extracted. It is clearly displayed that the slight alteration of the watermarked image, gives rise to a severe destruction of the counterpart watermark image. Experiments of more precisely partitioning can be operated by the similar process.

Table. 1 shows the PSNR comparison on Lena among the proposed and other methods. Compared with the proposed frequency transform-based methods in [9] and [10], the gains of PSNR are around 14 dB and 8 dB, respectively. This is quite a break-through in visual quality, recalling the considerable capacity of our proposed scheme.

(a) Upper left 50×50 Random Noise Attacked Watermarked Lena Image



(b) Watermarks Extracted from the Attacked Watermarked Lena Image

Figure 5: Attacked Image and Corresponding Watermarks Extracted

Table 1
PSNR Comparison Tested on Lena

Schemes	*Proposed*	*Baušys's*	*Wu's*
PSNR (dB)	51.827	37.780	43.420

6. CONCLUSION

In this paper, we propose a reversible framework for image authentication and tamper localization. The scheme enables us to ensure image authenticity. Blockwisely watermark embedding allows locating damaged regions with the resolution of up to two pixels. Reversible watermark embedding guarantees restoration of the original image after watermark embedding. Moreover, the visual quality of the marked image reaches a rather high altitude.

REFERENCES

[1] C. Honsinger, P. Jones, M. Rabbani, and J. Stoffel, "Lossless Recovery of an Original Image Containing Embedded Data", *US Patent*, No. US6278791, 1999.

[2] J. Fridrich, M. Goljan, and R. Du, "Invertible Authentication", Proc. SPIE, Security and Watermarking of Multimedia Contents, San Jose, California, January 23-26, 2001.

[3] M. Goljan, J. Fridrich, and R. Du, "Distortion-free data embedding for images", *4th Information Hiding Workshop*, LNCS **2137**, Springer-Verlag, New York, 2001 pp. 27-41.

[4] G. Xuan, J. Zhu, J. Chen, Y. Shi, Z. Ni, and W. Su, "Distortionless Data Hiding Based on Integer Wavelet Transform", *IEE Electronics Letters*, **38**(25), 2002, 1646-1648.

[5] J. Tian, "Reversible Watermarking by Difference Expansion", *Proceedings of Workshop on Multimedia and Security*, 19-22, 2002.

[6] Z. Ni, Y. Shi, N. Ansari, and W. Su, "Reversible Data Hiding", *IEEE Proceedings of ISCAS'03*, **2**, 2003, II-912-II-915.

[7] Z. Ni, Y. Shi, N. Ansari, and W. Su, "Reversible Data Hiding", *IEEE Transactions on Circuts and Systems for Video Technology*, **16**(3), 2006, 354-362.

[8] S. Lee,Y. Suh, and Y. Ho, "Reversible Image Authentication Based on Watermarking", *IEEE Proc. Of ICME'06*, July 2006, 1321-1324.

[9] R. Baušys, A. Kriukovas, "Reversible Watermarking Scheme for Image Authentication in Frequency Domain", *International Symposium ELMAR-2006*, Zadar, Croatia, June 2006.

[10] X. Wu, "Reversible Semi-fragile Watermarking based on Histogram Shifting of Integer Wavelet Coefficients", *IEEE Proc. of DEST* '07, Inaugural IEEE-IES, Feb. 2007, 501-505.