A Novel Robust Reversible Watermarking Method Against JPEG Compression

Hongya Wang[1,2], Xiaolong Li[1,2(✉)], Mengyao Xiao[1,2], and Yao Zhao[1,2]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China
lixl@bjtu.edu.cn
[2] Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China

Abstract. Robustness limits the application of reversible watermarking. To overcome this weakness, several robust reversible watermarking (RRW) techniques have been proposed so far. However, most existing RRW methods are unstable in terms of robustness and reversibility. Then, in this paper, to ameliorate these disadvantages, a new RRW algorithm is presented. A cover image is first divided into non-overlapping blocks, and a high pass filter is applied to each block to generate a histogram which is a Laplacien-like distribution. Then, the watermark is embedded into the blocks by shifting the generated histogram. Specifically, the histogram shifting is conducted by modifying each pixel in the block, and the embedding distortion is minimized based on a new modification mechanism. Moreover, for blind data extraction and image recovery, a strategy for determining the parameters used in histogram shifting is also proposed. In this way, more than 4,096 bits can be reversibly embedded into a cover image with a good visual quality and sufficient robustness against JPEG compression. The superiority of the proposed method is verified through extensive experiments.

Keywords: Robust reversible watermarking · Histogram shifting · JPEG compression · Blind extraction and recovery · Embedding distortion

1 Introduction

Reversible watermarking [1] is a special type of digital watermarking where both the watermark extraction and the host content recovery can be accomplished without loss at the decoder side. In the past two decades, many reversible watermarking methods have been proposed for digital images [2–9]. However, the transmission channel is supposed to be lossless in these methods, and the watermark extraction becomes a challenge in case of attack. To overcome this weakness, a new type of reversible watermarking, namely, robust reversible watermarking (RRW) is proposed [10]. By RRW, not only the embedded watermark but also the original host image can be restored without distortion in a lossless environment. Moreover, if the marked image is lossily compressed, although the host image may not be restored exactly, the embedded watermark should be recovered in this case. Existing RRW methods can be classified into two categories: redundant histogram

X. Sun et al. (Eds.): ICAIS 2021, CCIS 1424, pp. 312–322, 2021.
https://doi.org/10.1007/978-3-030-78621-2_25

shifting (RHS) based methods and multi-layer watermarking (MLW) based methods, and representative achievements of RHS techniques mainly include histogram rotation (HR) and generalized histogram shifting (GHS).

In [10], Vleeschouwer *et al.* first proposed the HR technique, in which each embedding block is randomly divided into two groups with equal number of pixels, and the watermark is embedded by modifying the centroid vectors of the two groups, while the salt and pepper noise is introduced. To deal with this issue, Zou *et al.* [11] proposed to modify the average values of the intermediate frequency sub-band in integer wavelet domain. In [12], Ni *et al.* improved the method [10] and avoided the salt and pepper distortion of HR while error bits are introduced for the blocks with extreme values 0 or 255. And thus, the error correction coding has been exploited in this work. Later on, Zeng *et al.* [13] proposed to divide the cover image into blocks to calculate the arithmetic difference of each block, and introduced two thresholds to embed the watermark by shifting the arithmetic differences. The performance of this method is better than some previous works, but the side information for data extraction and image recovery should be send to the receiver by using an additional communication channel, and thus this method is not blind. Then, in [14–17], An and Gao proposed several RRW methods to improve the robustness performance. Especially, in [15], they proposed a new robust reversible embedding framework based on clustering and wavelet transformation. However, this method is not blind as well, and the side information is still necessary for the decoder. In [18] and [19], Coltuc *et al.* first proposed the MLW technique, and there are two layers of watermark to be embedded into the cover image: a robust watermark is first embedded into the cover image to derive an intermediate image, and then a reversible watermark is embedded into the intermediate image. Most traditional reversible watermarking and robust watermarking techniques are available for this framework. However, for MLW, the noise is inevitably introduced to the intermediate image because both the robust and reversible embedding are applied on the same embedding domain. Then, in [20], Wang *et al.* proposed a new RRW technique to separately embed the robust and reversible watermark into the independent embedding domains to avoid the noise.

Existing RRW methods, including Zeng *et al.*'s work [13], have the disadvantages of insufficient embedding capacity and weak robustness due to the large embedding distortion. Based on this consideration, to improve the previous work [13], we propose a new RRW method in this paper. First, the cover image is divided into non-overlapping blocks and a histogram is generated by applying a high pass filter to each block. The generated histogram follows a Laplacien-like distribution. Then, the watermark is embedded into the blocks by shifting this histogram. Specifically, the method [13] is improved by minimizing the embedding distortion using a new modification mechanism. In addition, for blind data extraction and image recovery, a strategy for determining the parameters used in histogram shifting is also proposed. In this way, the proposed method is reversible without side information at the decoder side, and it provides a significant performance improvement in terms of both visual quality and embedding capacity. Experimental results show that the proposed scheme has sufficient robustness against JPEG compression compared with Zeng *et al.*'s method [13].

The rest of the paper is organized as follows. In Sect. 2, Zeng *et al.*'s RRW algorithm [13] is briefly introduced, and followed by the proposed scheme described in detail in

Sect. 3. Then, the experimental results and the comparison with Zeng *et al.*'s method [13] are reported in Sect. 4. Finally, the conclusions are presented in Sect. 5.

2 Related Work

In [13], Zeng *et al.* proposed a new RRW method by using two thresholds. The watermark is embedded by shifting the arithmetic differences of the divided image blocks.

First, the cover image is divided into non-overlapping blocks of size $m \times n$. Then, a matrix M sized $m \times n$ is introduced to calculate the arithmetic difference of each block. Specifically, the matrix M is given by

$$M(i,j) = \begin{cases} 1, & \text{if } i \text{ and } j \text{ have the same parity} \\ -1, & \text{otherwise} \end{cases}. \tag{1}$$

As an example, a matrix sized 8×8 is shown in Fig. 1. After that, the arithmetic difference of each block, is given by

$$\alpha = \sum_{i=1}^{m} \sum_{j=1}^{n} C(i,j) M(i,j), \tag{2}$$

where $C(i,j)$ means the pixel value of a given block C in the location (i,j). For example, the distribution of α for the Barbara image with block size of 8×8 is shown in Fig. 2, where the vertical axis represents the values of α and the horizontal axis is the occurrence of α.

1	-1	1	-1	1	-1	1	-1
-1	1	-1	1	-1	1	-1	1
1	-1	1	-1	1	-1	1	-1
-1	1	-1	1	-1	1	-1	1
1	-1	1	-1	1	-1	1	-1
-1	1	-1	1	-1	1	-1	1
1	-1	1	-1	1	-1	1	-1
-1	1	-1	1	-1	1	-1	1

Fig. 1. Matrix M sized 8×8.

For the data embedding of a given block C, two thresholds $T, G > 0$ are utilized to shift the value of α. If $\alpha > T$, it is shifted to the right side by $2G + T$ to create vacancy. Specifically, the cover pixel $C(i,j)$ is modified as

$$C^*(i,j) = \begin{cases} C(i,j) + \beta_1, & \text{if } M(i,j) = 1 \\ C(i,j), & \text{otherwise} \end{cases}, \tag{3}$$

where

$$\beta_1 = \left\lceil \frac{4G + 2T}{mn} \right\rceil. \tag{4}$$

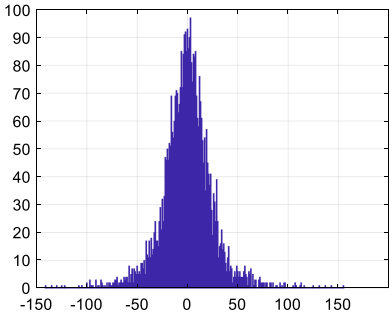

Fig. 2. The distribution of α of the Barbara image with the 8×8 matrix shown in Fig. 1.

In this way, one can verify that the value of α will be changed as

$$\sum_{i=1}^{m}\sum_{j=1}^{n} C^*(i,j)M(i,j) = \alpha + \frac{mn}{2}\beta_1 > 2(G+T). \tag{5}$$

That is to say, there is no value inside the range of $(T, 2(G+T)]$ after shifting. Similarly, if $\alpha < -T$, it is shifted to the left side by modifying the cover pixel $C(i,j)$ as

$$C^*(i,j) = \begin{cases} C(i,j) + \beta_1, & \text{if } M(i,j) = -1 \\ C(i,j), & \text{otherwise} \end{cases}. \tag{6}$$

If $\alpha \in [0, T]$, the cover pixel $C(i,j)$ is modified as

$$C^*(i,j) = \begin{cases} C(i,j) + \omega\beta_2, & \text{if } M(i,j) = 1 \\ C(i,j), & \text{otherwise} \end{cases}, \tag{7}$$

where $\omega \in \{0, 1\}$, represents the watermark bit to be embedded and

$$\beta_2 = \left\lceil \frac{2G + 2T}{mn} \right\rceil. \tag{8}$$

Similarly, if $\alpha \in [-T, 0)$, the cover pixel $C(i,j)$ is modified as

$$C^*(i,j) = \begin{cases} C(i,j) + \omega\beta_2, & \text{if } M(i,j) = -1 \\ C(i,j), & \text{otherwise} \end{cases}. \tag{9}$$

As shown in Fig. 3, the value of α falls into the range of $[-T, T]$ due to the embedding bit '0', called bit-0-zone. And the value of α is within the range of $[T + G, 2T + G]$ or $[-(2T + G), -(T + G))$ due to the embedding bit '1', called bit-1-zone.

In case of no attack, the extraction is a reverse process, where a bit '0' is extracted when $\alpha \in [-T, T]$, and a bit '1' is extracted when $\alpha \in (T, 2T + G]$ or $\alpha \in [-(2T + G), -T)$. Moreover, the cover image can be recovered completely. If $\alpha \in (T, 2T + G]$, the original pixel $C(i,j)$ can be recovered by

$$C(i,j) = \begin{cases} C^*(i,j) - \omega\beta_2, & \text{if } M(i,j) = 1 \\ C^*(i,j), & \text{otherwise} \end{cases}, \tag{10}$$

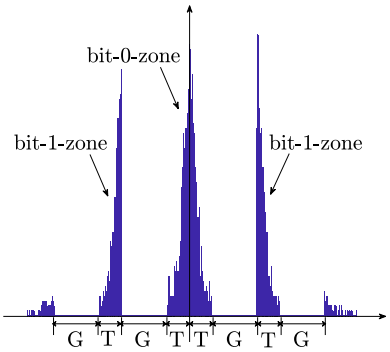

Fig. 3. The distribution of α after watermark embedding.

where $\omega \in \{0, 1\}$, represents the extracted watermark bit. Similarly, if $\alpha \in [-(2T + G), -T)$, the original pixel $C(i, j)$ can be recovered and the details are omitted. If $\alpha > 2T + G$ or $\alpha < -(2T + G)$, the original pixel $C(i, j)$ can be recovered as well according to Eq. (3) and Eq. (6).

As shown in Fig. 4, when the marked image has been attacked by JPEG compression, Zeng *et al.* [13] first find two ranges of $[-Adj\_0, Adj\_0]$ and $[-Adj\_1, Adj\_1]$. Then, the values of T and G can be calculated by Eq. (11). Finally, the watermark can be extracted correctly even if the marked image has been attacked by JPEG compression to some extent.

$$\begin{cases} T = Adj\_0 \\ G = Adj\_1 - 2Adj\_0 \end{cases} \tag{11}$$

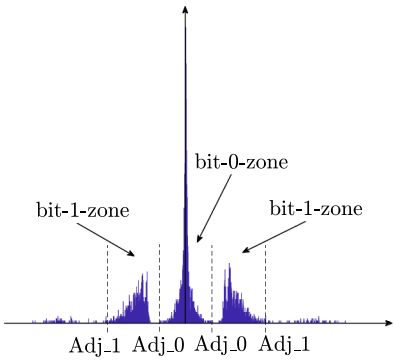

Fig. 4. The distribution of α after JPEG compression.

In general, since the value of T is supposed to be sufficiently large to ensure the robustness. In the experiments of [13], all the pixels are just expanded. Clearly, for this method, the MSE is about $\frac{2(T+G)^2}{mn}$. Moreover, in terms of reversibility, the values of T and G are necessary for the decoder to correctly extract the watermark.

3 Proposed Method

In this section, our proposed method is introduced in detail, which is an improvement of Zeng *et al.*'s work [13]. Here, we adopt the notations used in Sect. 2. We first introduce the proposed data embedding process. The cover image C is firstly divided into a number of non-overlapping blocks sized $m \times n$. Then, as shown in Eq. (1) and Eq. (2), the arithmetic difference α of each block is calculated using the matrix M. Next, two thresholds $T > 0$ and $G > 0$ are selected as parameters for data embedding. Notice that, in our method, the threshold T is selected such that it is exactly larger than the maximum of α for all divided image blocks. Finally, the watermark is embedded by shifting the histogram of α. Specifically, if $\alpha \geq 0$, the marked pixel $C^*(i,j)$ is modified as

$$C^*(i,j) = \begin{cases} C(i,j) + \omega\beta, \text{ if } M(i,j) = 1 \\ C(i,j) - \omega\beta, \text{ if } M(i,j) = -1 \end{cases}, \tag{12}$$

where $\omega \in \{0, 1\}$, represents the watermark bit to be embedded and

$$\beta = \left\lceil \frac{G+T}{mn} \right\rceil. \tag{13}$$

And if $\alpha < 0$, the marked pixel $C^*(i,j)$ is modified as

$$C^*(i,j) = \begin{cases} C(i,j) - \omega\beta, \text{ if } M(i,j) = 1 \\ C(i,j) + \omega\beta, \text{ if } M(i,j) = -1 \end{cases}. \tag{14}$$

Figure 5 shows the distribution histogram after embedding the watermark. The value of α falls into the range of $[-T,T]$ due to the embedding bit '0', called bit-0-zone. While the value of α is within the range of $[T+G, 2T+G]$ or $[-(2T+G), -(T+G))$ due to the embedding bit '1', called bit-1-zone. Bit-0-zone and bit-1-zone are separated by the length G. In this way, after a non-malicious attack such as JPEG compression, the two areas will not overlap each other so that the watermark can be correctly extracted. In other words, the proposed algorithm is robust against non-malicious attacks.

Here, by minimizing the embedding distortion, we optimize Zeng *et al.*'s embedding method [13] and then propose the embedding algorithm as described above. For $T > 0$ and $G > 0$, the two methods both shift α by $T + G$ to embed bit '1' into a divided image block. In [13], only half of the pixels in the block have changed by $\left\lceil \frac{2(T+G)}{mn} \right\rceil$. After calculation, the mean square error (MSE) is about $\frac{2(T+G)^2}{mn}$. While in this paper, we have modified all pixels in the block by $\left\lceil \frac{(T+G)}{mn} \right\rceil$, and then conclude that the MSE is approximate to $\frac{(T+G)^2}{mn}$, only half of the MSE in [13]. In other words, our proposed algorithm has superior performance in terms of visual quality.

If the marked image C^* is not distorted, as shown in Fig. 5, we can find two ranges of $[-Lim\_0, Lim\_0]$ and $[-Lim\_1, Lim\_1]$ to get the values of T and G. In particular, T and G are initialized to the integer multiple of $m \times n/2$ in our proposed algorithm. According to that, we can first calculate the value of Lim_1 using the maximum of α for all blocks. Then, we can obtain the value of Lim_0 according to the number of bits '1' embedded

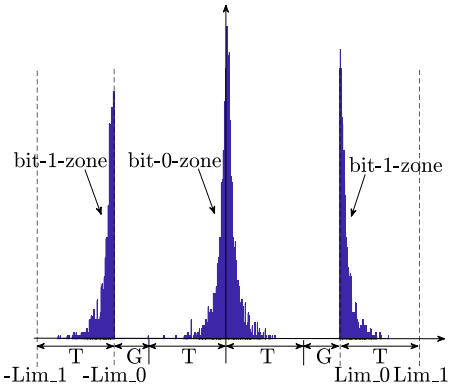

Fig. 5. The distribution of α after data embedding.

into the image. Finally, the values of T and G can be calculated by Eq. (15). Similarly, when the watermarked image C^* is attacked by JPEG compression, we can also find two ranges and calculate T and G according to Eq. (11). In this case, although the cover image cannot be restored completely, the watermark can be extracted using T and G. In particular, if $\alpha \in [-T, T]$, a bit '0' is extracted. If $\alpha \in [-(2T+G), -T) \cup (T, 2T+G]$, a bit '1' is extracted.

$$\begin{cases} T = Lim\_1 - Lim\_0 \\ G = 2Lim\_0 - Lim\_1 \end{cases} \tag{15}$$

Now consider the case that the marked image C^* is not distorted. Not only we can extract watermark correctly, but also the original cover image C can be recovered without loss. Specifically, if $\alpha \in (T, 2T + G]$, the original pixel $C(i, j)$ is given by

$$C(i,j) = \begin{cases} C^*(i,j) - \omega\beta, \text{ if } M(i,j) = 1 \\ C^*(i,j) + \omega\beta, \text{ if } M(i,j) = -1 \end{cases}, \tag{16}$$

where $\omega \in \{0, 1\}$, represents the extracted watermark bit. And if $\alpha \in [-(2T+G), -T)$, the original pixel $C(i, j)$ is given by

$$C(i,j) = \begin{cases} C^*(i,j) + \omega\beta, \text{ if } M(i,j) = 1 \\ C^*(i,j) - \omega\beta, \text{ if } (i,j) = -1 \end{cases}. \tag{17}$$

4 Experimental Results

In this section, three commonly used images including Lena, Airplane and Barbara are utilized to evaluate the robustness of our algorithm compared with Zeng *et al.*'s method [13]. For each image, 100 groups of watermarks are embedded to calculate PSNR and bit error rate (BER). The BER is given by

$$BER = \frac{\omega_e}{\omega_b}. \tag{18}$$

where ω_e indicates the number of bits extracted incorrectly, and ω_b denotes the number of bits embedded into the image.

Figure 6 shows the relationship between PSNR and embedding level β, where $\beta = \left\lceil \frac{T+G}{mn} \right\rceil$. To ensure the watermark invisible to human eyes, PSNR should be greater than 38 dB [21], i.e., $\beta < 5$. Here, the threshold T is decided such that it is exactly larger than the maximum of α for all divided image blocks, where T and G are integer multiples of $m \times n/2$. For the Lena image with the block sized 8×8, we decide that $T = 128$ and $G = 64$ such that β is equal to 6, not less than 5. In the same way, thresholds T and G are selected for the Airplane image with the block of size 8×8, i.e. $T = 160, G = 32$.

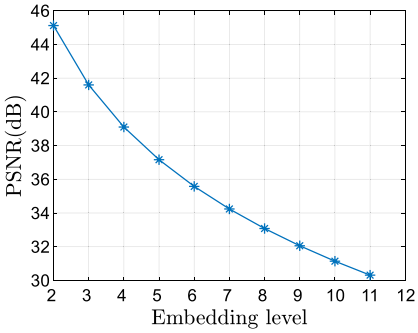

Fig. 6. The relationship between β and PSNR.

Specifically, we evaluate the robustness according to the relationship between BER and q with the same value of PSNR, and the relationship between 1-BER and PSNR with the same compression strength. With the same value of q and PSNR, the lower the BER, the stronger the robustness against JPEG compression.

Table 1 shows the BER with compression quality factors of 90, 85 and 80 respectively. Meanwhile, the PSNR of our algorithm is slightly higher than that of Zeng *et al.* [13]. Moreover, the values of BER are obviously lower under the same compression quality factor. That is to say, the proposed algorithm is more robust to JPEG compression than Zeng *et al.*'s method [13].

Table 1. Comparison for block size 8×8.

Method	Zeng *et al.* PSNR = 38.59 dB			Proposed PSNR = 39.10 dB		
q	90	85	80	90	85	80
Lena	0.05	0.13	19.65	0	0	0.01
Airplane	0	0.05	24.94	0	0	0.03
Barbara	0.23	0.62	19.89	0	0	0.16

As shown in Fig. 7, we initialize the JPEG quality to 90%, 80% and 70%. Notice that when the PSNR value is less than a certain value, the BER is quite low. While the PSNR becomes lager than this value, the BER increases sharply. Obviously, if the PSNR value is the same, the 1-BER of the proposed algorithm is smaller. That is to say, our proposed algorithm outperforms [13] in terms of the robustness.

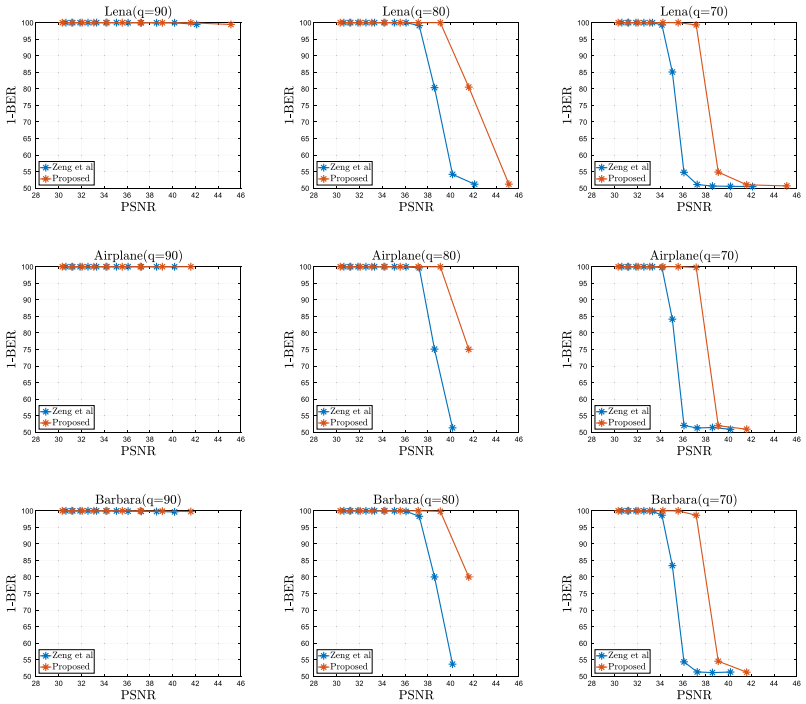

Fig. 7. Comparison for $EC = 4,096$ bits and different PSNRs.

Figure 8 shows the relationship between BER and q with the same value of PSNR. In [13], the BER increases sharply in case of $q < 85\%$. The proposed algorithm has the same phenomenon in case of $q < 80\%$. Therefore, the proposed algorithm is more robust to JPEG compression.

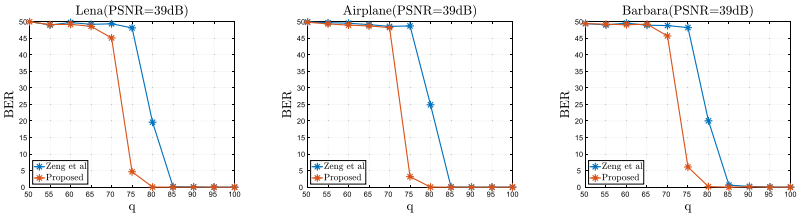

Fig. 8. Comparison for $EC = 4,096$ bits and quality factors.

5 Conclusions

To overcome the disadvantages of insufficient embedding capacity and weak robustness of the previous RRW methods, in this paper, a new embedding mechanism for RRW is proposed. First, the cover image is divided into non-overlapping blocks and a histogram is generated by applying a high pass filter to each block. Then, the watermark is embedded into the blocks by shifting this histogram. Specifically, instead of using half pixels in the block, each pixel is modified to minimize the embedding distortion. Moreover, a strategy for determining the parameters is proposed to make the method reversible without side information. Experimental results show that compared with the previous methods, the proposed scheme has sufficient robustness against JPEG compression.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (Nos. 61972031 and U1736213).

References

1. Barton, J. M.: Method and apparatus for embedding authentication information within digital data. U.S.Patent 5646997 (1997)
2. Fridrich, J., Goljan, M., Du, R.: Lossless data embedding new paradigm in digital watermarking. EURASIP J. Appl. Signal Process. **2002**(2), 185–196 (2002)
3. Tian, J.: Reversible data embedding using a difference expansion. IEEE Trans. Circ. Syst. Video Technol. **13**(8), 890–896 (2003)
4. Ni, Z., Shi, Y.-Q., Ansari, N., Su, W.: Reversible data hiding. IEEE Trans. Circ. Syst. Video Technol. **16**(3), 354–362 (2006)
5. Sachnev, V., Kim, H.J., Nam, J., Suresh, S., Shi, Y.-Q.: Reversible watermarking algorithm using sorting and prediction. IEEE Trans. Circ. Syst. Video Technol. **19**(7), 989–999 (2009)
6. Shi, Y.-Q., Li, X., Zhang, X., Wu, H.-T., Ma, B.: Reversible data hiding: advances in the past two decades. IEEE Access **4**, 3210–3237 (2016)
7. Xiang, L., Yang, S., Liu, Y., Li, Q., Zhu, C.: Novel linguistic steganography based on character-level text generation. Mathematics **8**, 1558 (2020)
8. Xiang, L., Guo, G., Li, Q., Zhu, C., Chen, J.: Spam detection in reviews using lstm-based multi-entity temporal features. Intell. Autom. Soft Comput. **26**(6), 1375–1390 (2020)
9. Yang, Z., Zhang, S., Hu, Y., Hu, Z., Huang, Y.: VAE-stega: linguistic steganography based on variational auto-encoder. IEEE Trans. Inf. Forensics Secur.ity **16**, 880–895 (2021)
10. De Vleeschouwer, C., Delaigle, J.F., Macq, B.: Circular interpretation of bijective transformations in lossless watermarking for media asset management. IEEE Trans. Multimed. **5**(1), 97–105 (2003)
11. Zou, D., Shi, Y., Ni, Z., Su, W.: A semi-fragile lossless digital watermarking scheme based on integer wavelet transform. IEEE Trans. Circ. Syst. Video Technol. **16**(10), 1294–1300 (2006)
12. Ni, Z., Shi, Y., Ansari, N., Su, W., Sun, Q., Lin, X.: Robust lossless image data hiding designed for semi-fragile image authentication. IEEE Trans. Circ. Syst. Video Technol. **18**(4), 497–509 (2008)
13. Zeng, X.-T., Ping, L.-D., Pan, X.-Z.: A lossless robust data hiding scheme. Pattern Recogn. **43**(4), 1656–1667 (2010)
14. Gao, X., An, L., Yuan, Y., Tao, D., Li, X.: Lossless data embedding using generalized statistical quantity histogram. IEEE Trans. Circ. Syst. Video Technol. **21**(8), 1061–1070 (2011)

15. An, L., Gao, X., Li, X., Tao, D., Deng, C., Li, J.: Robust reversible watermarking via clustering and enhanced pixel-wise masking. IEEE Trans. Image Process. **21**(8), 3598–3611 (2012)
16. An, L., Gao, X., Yuan, Y., Tao, D.: Robust lossless data hiding using clustering and statistical quantity histogram. Neurocomputing **77**(1), 1–11 (2012)
17. An, L., Gao, X., Yuan, Y., Tao, D., Deng, C., Ji, F.: Content-adaptive reliable robust lossless data embedding. Neurocomputing **79**, 1–11 (2012)
18. Coltuc, D.: Towards distortion-free robust image authentication. J. Phys.: Conf. Ser. **77**(1), 012005 (2007)
19. Coltuc, D., Chassery, J.: Distortion-free robust watermarking: a case study, vol. 6505. International Society for Optics and Photonics (2007).
20. Wang, X., Li, X., Pei, Q.: Independent embedding domain based two-stage robust reversible watermarking. IEEE Trans. Circ. Syst. Video Technol. **PP**(99), 1 (2019)
21. Katzenbeisser, S., Petitcolas, A.P.: Information hiding techniques for stegano-graphy and digital watermarking. Artech House Inc. **28**(6), 1–2 (1999)