International Journal of Digital Crime and Forensics
Volume 11 • Issue 1 • January-March 2019

Asymmetric Distortion Function for JPEG Steganography Using Block Artifact Compensation

Zichi Wang, Shanghai Institute for Advanced Communication and Data Science, Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, China

Zhaoxia Yin, Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Anhui, China

Xinpeng Zhang, Shanghai Institute for Advanced Communication and Data Science, Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, China

ABSTRACT

This article describes how the existing distortion functions for JPEG steganography allot same cost for ± 1 embedding changes. Because of the correlation of natural image, however, changes with different polarities make different influences on image. Therefore, the embedding costs for ± 1 embedding changes should not be equivalent. This article proposes a general method to distinguish the embedding costs for different polarities of embedding changes for JPEG images with the help of reference images constructed by block artifact compensation. The original JPEG image is decompressed into spatial domain firstly, and then the block artifact is compensated by smoothing filtering implemented on border pixels of each 8×8 block. After that, the compensated image which is more similar to the original uncompressed image is recompressed into DCT domain and adopted as side information to guide the adjusting of the given distortion function. Experiment results show that after the proposed method is employed, the security performance of current popular JPEG steganographic methods is observably increased.

KEYWORDS

Distortion Function, JPEG Images, Polarity, Steganography

INTRODUCTION

Steganography aims to transmit data secretly through digital media without drawing suspicion by slightly modifying cover data (Zhang, 2016). The early steganographic methods try to increase the undetectability by decreasing the quantity of embedding changes (Fridrich & Soukal, 2006; Zhang & Wang, 2006; Zhang, Zhang & Wang, 2008), or by maintaining a kind of statistical model (Westfeld, 2001; Phil, 2003; Fridrich, Pevný & Kodovský, 2007). But these methods cannot achieve

DOI: 10.4018/IJDCF.2019010107

satisfactory undetectability due to the lack of consideration about image content. With the emergence of the syndrome trellis coding (STC) (Filler, Judas & Fridrich, 2011) which can minimize additive distortion between the cover and the stego image under a user-defined distortion function with a given payload, the security performance of steganography has been greatly improved and the direction of steganography is turn to the design of distortion function.

A distortion function allots an embedding cost for each cover element and the embedding cost quantifies the effect for modifying the cover element. The distortion between cover and the corresponding stego object is expressed as a sum of costs of modified elements. There are a mountain of excellent distortion functions for spatial images (Pevný, Filler & Bas, 2010; Holub & Fridrich, 2012, Holub & Fridrich, 2013; Li, Wang, Huang & Li, 2014, Sedighi, Fridrich & Cogranne, 2015) or JPEG images (Holub & Fridrich, 2013; Guo, Ni & Shi, 2014; Guo, Ni, Su, Tang & Shi, 2015; Wang, Zhang & Yin, 2016). Most of these distortion functions allot a same embedding cost for ± 1 embedding changes. But because of the correlation of natural images, changes with different polarities make different influences on an image. Therefore, the embedding cost for ± 1 embedding changes should not be equivalent. So, these existing distortion functions can be optimized by distinguish the embedding costs for ± 1 embedding changes.

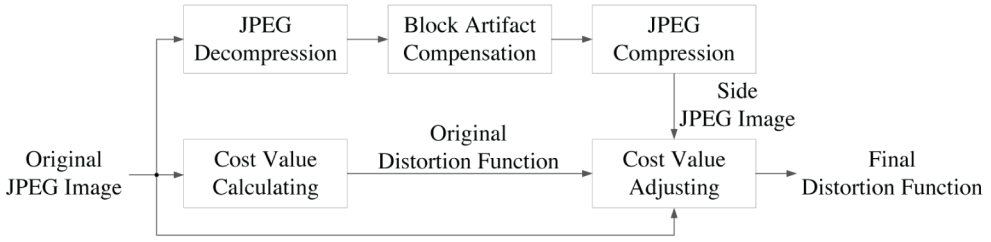
A general method to distinguish the embedding costs for different polarities of embedding changes for spatial images is proposed in (Wang, Lv, Wei & Zhang, 2016). Both the fluctuation after pixels being +1 or -1 and the texture of cover image are employed to adjust a given distortion function. This approach makes the fluctuation around modified pixels become more similar to that of their neighbourhoods. Thus, less detectable artifacts achieved. As JPEG is widely used, it is meaningful to design asymmetric distortion function for JPEG steganography. Some schemes are designed for distinguish the embedding cost for different polarities of embedding changes for JPEG steganography with the help of additional information such as spatial precover (Denemark & Fridrich, 2015) or multiple JPEG images of the same scene (Denemark & Fridrich, 2017). However, the need of additional information makes these schemes impracticable in real world. Up to now, there is no asymmetric JPEG distortion function without any help of additional information.

This paper firstly proposes a general method to distinguish the embedding costs for different polarities of embedding changes for JPEG images by compensating the block artifact. The original JPEG image is decompressed into spatial domain, and then the block artifact caused by JPEG compression is compensated. After that, the image is recompressed into DCT domain and adopted as side information to adjust the given distortion function. Note that the side information is produced from the given JPEG image, not from any additional information.

PROPOSED METHOD

The sketch of the proposed method is shown in Figure 1. The original JPEG image is decompressed into spatial domain firstly, and then the block artifact is compensated by smoothing filtering implemented

Figure 1. Sketch of the proposed method



on border pixels of each 8 × 8 block. After that, the image is recompressed into DCT domain and adopted as side information to adjust the original distortion function calculated by existing methods. The details are as follows.

For a JPEG image sized $M \times N$, the (i, j)th quantized DCT coefficient is denoted as $c(i, j)$, where $i \in \{1, 2, \ldots, M\}$, $j \in \{1, 2, \ldots, N\}$. The cost of making an embedding change by +1 or -1 of $c(i, j)$ are denoted as $\rho_+(i, j)$, $\rho_-(i, j)$, respectively. In current JPEG steganographic methods based on STC, $\rho_+(i, j)$ is always equal to $\rho_-(i, j)$. It means $\rho_+(i, j) = \rho_-(i, j) = \rho(i, j)$.

After JPEG decompression, denote the generated spatial image as $\mathbf{X}$ and the (i, j)th pixel $x(i, j)$. Then the block artifact in $\mathbf{X}$ caused by JPEG compression is compensated by smoothing filtering implemented on border pixels of each 8 × 8 block using Equations (1) and (2). During JPEG compression, the blocking processing introduces horizontal and vertical breaks into images, which is known as block artifact. These breaks are particularly evident on border pixels of each block. To avoid introducing new breaks, the smoothing filtering is only implemented on border pixels of each 8 × 8 block to compensate the block artifact.

$$\mathbf{X}' = \mathbf{X} \otimes \mathbf{F} \tag{1}$$

$$\hat{x}(i, j) = \begin{cases} x'(i, j), & (i, j) \in D \\ x(i, j), & \text{otherwise} \end{cases} \tag{2}$$

where $\mathbf{F}$ is a 3 × 3 smoothing filter defined in Equation (3), which is used to compensate the block artifact in $\mathbf{X}$.

$$\mathbf{F} = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3}$$

$D = \{ (8k_1, u), (1+8k_1, u), (v, 8k_2), (v, 1+8k_2) \mid u = 2, 3, \ldots, N\text{-}1, v = 2, 3, \ldots, M\text{-}1, k_1 = 1, 2, \ldots, \lfloor M/8 \rfloor - 1, k_2 = 1, 2, \ldots, \lfloor N/8 \rfloor - 1 \}$ is a set containing the border locations of each 8 × 8 block. As shown in Figure 2, the locations marked with red circle belong to D.

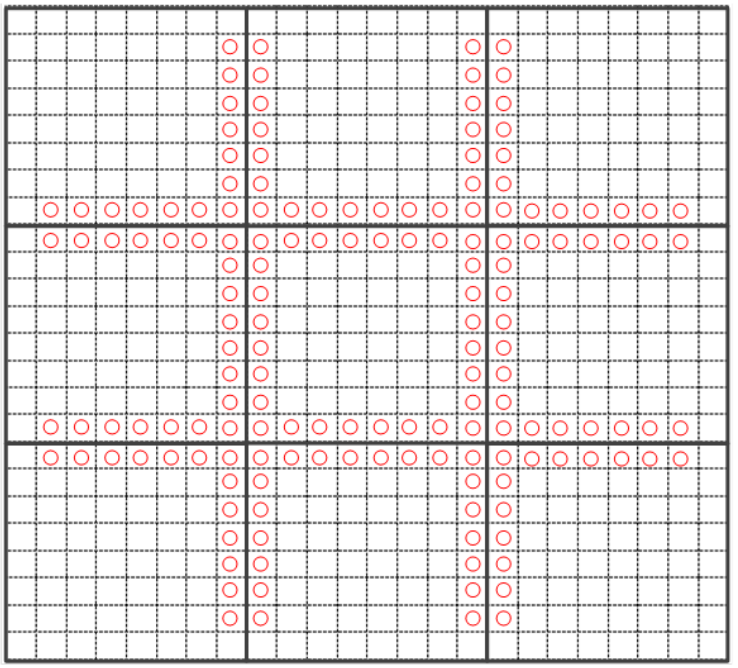
Then the compensated spatial image $\hat{\mathbf{X}}$ is recompressed into DCT domain with the same quality factor and adopted as side information to adjust the original distortion function calculated by existing methods. Denote the recompressed (i, j)th quantized DCT coefficients as $c_{\mathrm{SI}}(i, j)$. The approach to adjust the given embedding cost $\rho(i, j)$ is described in Equations (4) and (5),

$$\rho_+(i, j) = \begin{cases} a \cdot \rho(i, j) & , \quad c(i, j) < c_{\mathrm{SI}}(i, j) \\ \rho(i, j) & , \quad c(i, j) \geq c_{\mathrm{SI}}(i, j) \end{cases} \tag{4}$$

$$\rho_-(i, j) = \begin{cases} a \cdot \rho(i, j) & , \quad c(i, j) > c_{\mathrm{SI}}(i, j) \\ \rho(i, j) & , \quad c(i, j) \leq c_{\mathrm{SI}}(i, j) \end{cases} \tag{5}$$

where parameter $a \in [0,1]$ being used to adjust the extent of modification made on $\rho(i, j)$. The logic behind Equations (4) and (5) is that the JPEG image with block artifact compensated is more similar

Figure 2. Border locations of each 8 × 8 block



to original uncompressed spatial image. It is advisable to change $c(i,j)$ toward $c_{SI}(i,j)$ if there is any modification needed. To find the optimal value of a, some experiments are carried out, and the details are as follows.

The image dataset UCID (Schaefer, G., & Stich, M. 2004) which contains 1338 uncompressed color images sized 512 × 384 is employed. These images are transformed into grayscale image firstly and then compressed into JPEG with quality factor QF = 75 and QF = 95. The error probability P_E tested on JUNIWARD (Holub & Fridrich, 2013) and UED (Guo, Ni & Shi, 2014) with the steganalysis feature DCTR-8000D (Holub, V., & Fridrich, J. 2014) and ensemble classifier (Kodovsky, Fridrich, & Holub, 2012) is shown in Figure 3.

It can be seen from Figure 3 that the variation between P_E and the value of a performs similar tendency for different cases, and the optimal value of a is around 0.7 and independent of image quality factors and steganographic methods. Therefore, the value of a is determined as 0.7 for all JPEG image.

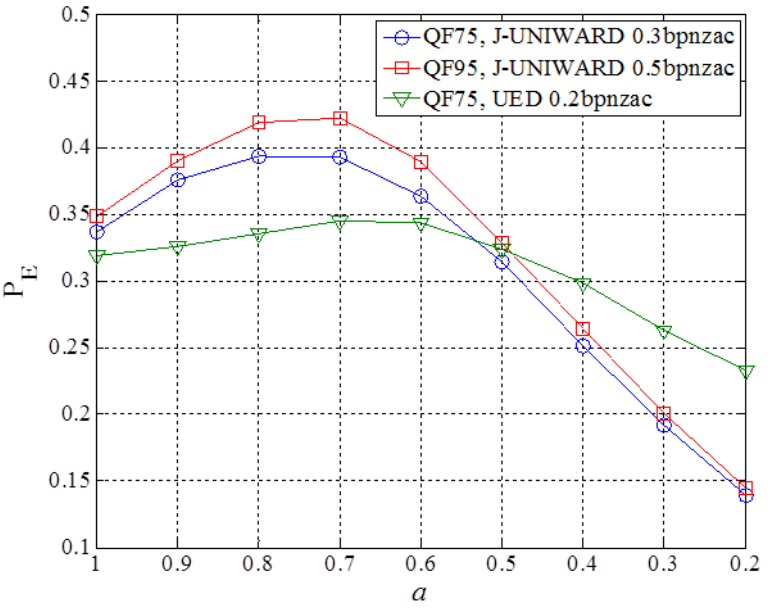
EXPERIMENT RESULTS

To verify the effectiveness of the proposed distortion function optimization method, several experiments are designed in this Section.

Experiment Setup

The image datasets employed in our experiments are UCID which contains 1338 uncompressed color images sized 512 × 384, and BOSSbass ver. 1.01 (Bas, Filler & Pevný, 2011) which contains 10000 uncompressed grayscale images sized 512 × 512. The images in UCID are transformed into grayscale images, and then all the images in UCID and BOSSbass ver. 1.01 are compressed into JPEG with quality factor QF = 75 and QF = 95 to be adopted as covers for experiment comparison.

Figure 3. Parameter Optimization



Popular JPEG steganographic methods JUNIWARD (Holub & Fridrich, 2013), UED (Guo, Ni & Shi, 2014), and UERD (Guo, Ni, Su, Tang & Shi, 2015) are employed to test the effectiveness of the proposed method. All the methods are worked via STC. The payloads range from 0.05 to 0.5 bit per non-zero AC coefficient (bpnzac).

In the phase of steganalysis, the most representative and optimal feature sets against JPEG steganography, DCTR-8000D (Holub & Fridrich, 2014) and GRF-17000D (Song, Liu, Yang, Luo & Zhang, 2015) are employed in our experiments. And the ensemble classifier (Kodovsky, Fridrich & Holub, 2012) is used to measure the property of feature sets. In detail, half of the cover and stego feature sets are used as the training sets while the remaining half are used as testing sets. The criterion to evaluate the performance of feature sets is the minimal total error P_{E} under equal priors achieved on the testing sets:

$$P_E = \min_{P_{FA}} \left(\frac{P_{\mathrm{FA}} + P_{MD}}{2} \right) \tag{6}$$

where P_{FA} is the false alarm rate and P_{MD} is the missed detection rate. The performance is evaluated using the average value of P_{E} over 10 random tests.

Security Performance

When the proposed method is incorporated in JUNIWARD, UED, and UERD, the updated methods JUNIWARD-P, UED-P, and UERD-P are obtained respectively. Figure 4 and Figure 5 show the comparisons of JUNIWARD with JUNIWARD-P, UED with UED-P, and UERD with UERD-P on UCID. And the corresponding comparisons on BOSSbass ver. 1.01 are shown in Figure 6 and Figure 7. It is clear that the security performance of steganography is improved by employing the proposed method. This observation attests that the approach to adjust distortion function proposed in this paper is effective.

Figure 4. Testing error for JUNIWARD, UED, and UERD on image dataset UCID with DCTR-8000D and ensemble classifier (a) QF = 75, (b) QF = 95

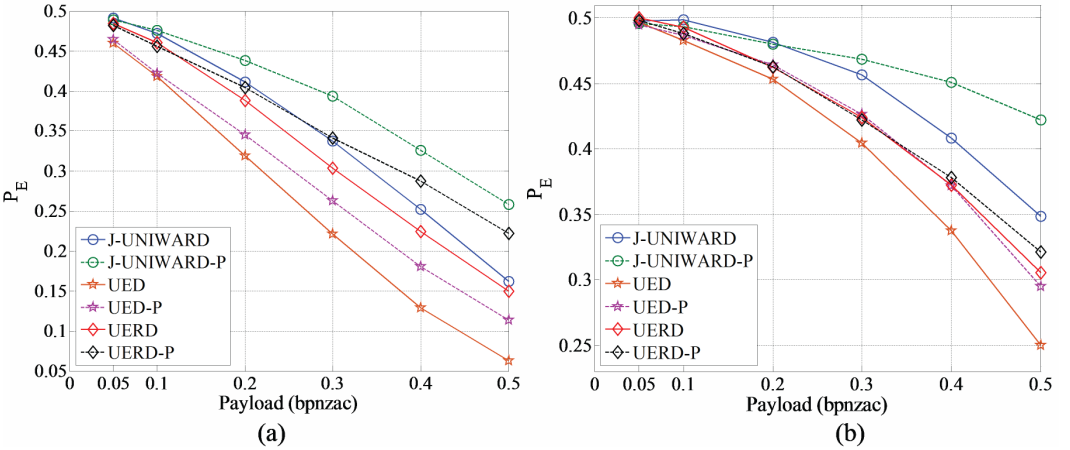
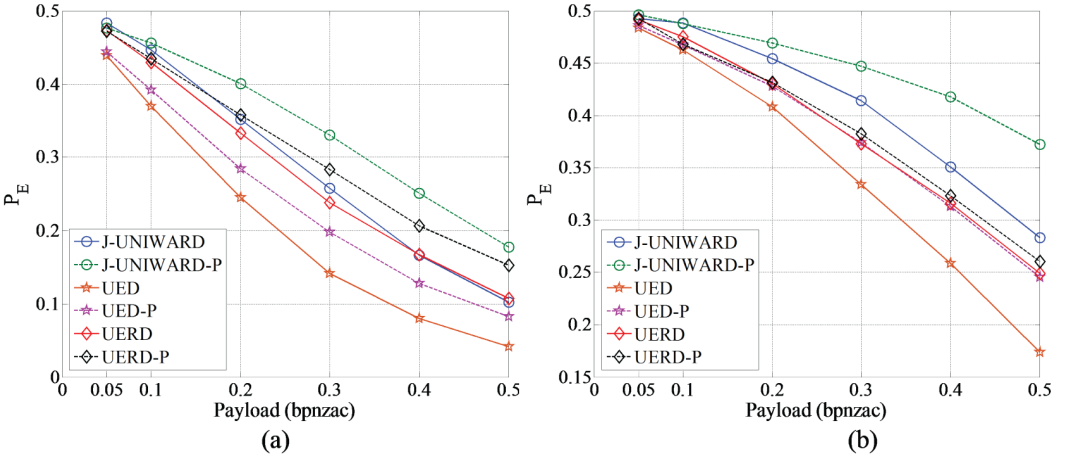


Figure 5. Testing error for JUNIWARD, UED, and UERD on image dataset UCID with GFR-17000D and ensemble classifier (a) QF = 75, (b) QF = 95



Numerically, for image dataset UCID, the P_E for JUNIWARD can be improved by 9.58% after employing the proposed method when payload is 0.5 bpnzac and QF = 75 with DCTR-8000D, 8.84% when payload is 0.4 bpnzac and QF = 75 with GFR-17000D, and 8.95% when payload is 0.2 bpnzac and QF = 95 with GFR-17000D. For UED, the improvement is 5.16% when payload is 0.4 bpnzac and QF = 75 with DCTR-8000D, 5.60% when payload is 0.3 bpnzac and QF = 75 with GFR-17000D, and 7.18% when payload is 0.5 bpnzac and QF = 95 with GFR-17000D. On the whole, the average improvement on P_E of JUNIWARD tested on image dataset UCID is 3.60%, and 2.91% for UED.

For image dataset BOSSbass ver. 1.01, the P_E for JUNIWARD can be improved by 9.45% after employing the proposed method when payload is 0.4 bpnzac and QF = 75 with DCTR-8000D, 7.64% when payload is 0.3 bpnzac and QF = 75 with GFR-17000D, and 10.79% when payload is 0.5 bpnzac and QF = 95 with DCTR-8000D. For UED, the improvement is 4.96% when payload is 0.2 bpnzac and QF = 75 with DCTR-8000D, 4.50% when payload is 0.2 bpnzac and QF = 75 with GFR-17000D, and 5.50% when payload is 0.5 bpnzac and QF = 95 with DCTR-8000D. On the

Figure 6. Testing error for JUNIWARD, UED, and UERD on image dataset BOSSbass ver. 1.01 with DCTR-8000D and ensemble classifier (a) QF = 75, (b) QF = 95

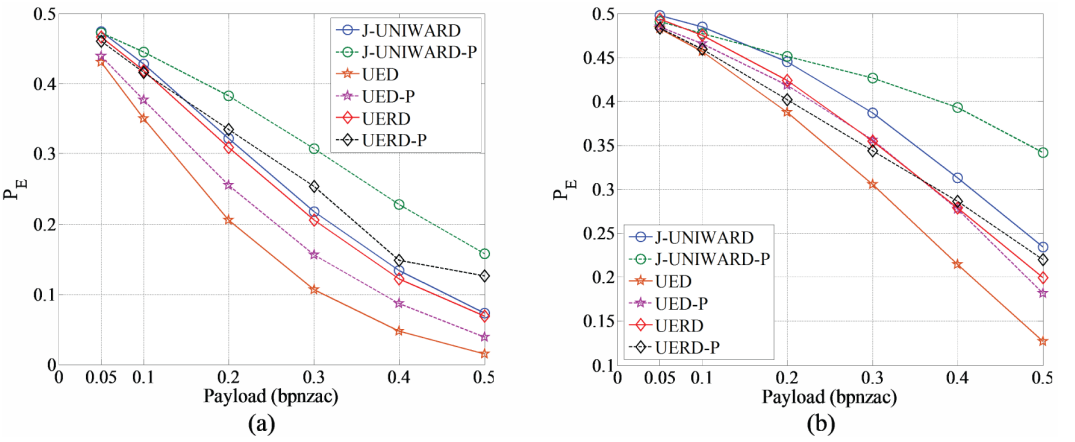
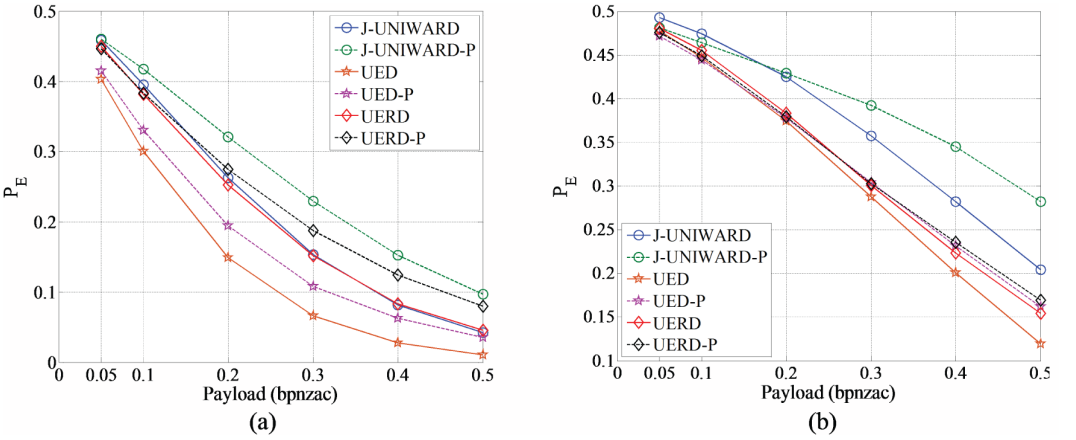


Figure 7. Testing error for JUNIWARD, UED, and UERD on image dataset BOSSbass ver. 1.01 with GFR-17000D and ensemble classifier (a) QF = 75, (b) QF = 95



whole, the average improvement on P_E of JUNIWARD tested on image dataset BOSSbass ver. 1.01 is 4.20%, and 2.83% for UED.

In addition, it can be seen from Figures 4-7 that the improvement on P_E under high payload is larger than those under low payload. The reason of this observation is related to the detection error P_E of steganalysis tools which is around 50% under low payload. Since 50% is the highest value of P_E for binary classification, it is difficult to increase P_E further under low payload.

CONCLUSION

This paper firstly proposes a general method to distinguish the embedding cost for different polarity of embedding change for JPEG images with the help of a reference image constructed by block artifact compensation. The undetectability of current popular JPEG steganographic methods is observably

increased after incorporating the proposed distortion cost adjustment method. For further study, it is significant to design universal asymmetric distortion function which can be used for both spatial and JPEG steganography.

ACKNOWLEDGMENT

This work was supported in the Natural Science Foundation of China (U1636206, 61525203, 61502009, 61602295, and 61472235), the Shanghai Dawn Scholar Plan (14SG36) and the Shanghai Excellent Academic Leader Plan (16XD1401200).

REFERENCES

Bas, P., Filler, T., & Pevný, T. (2011). Break Our Steganographic System: The Ins and Outs of Organizing BOSS. In *Proc. 13th International Conference on Information Hiding*, Prague, Czech Republic (pp. 59-70). doi:10.1007/978-3-642-24178-9_5

Denemark, T., & Fridrich, J. (2015). Side-informed steganography with additive distortion. In *Proc. IEEE International Workshop on Information Forensics and Security*, Rome, Italy (pp. 16-19).

Denemark, T., & Fridrich, J. (2017). Steganography with Multiple JPEG Images of the Same Scene. *IEEE Transactions on Information Forensics and Security*, *12*(10), 2308–2319. doi:10.1109/TIFS.2017.2705625

Filler, T., Judas, J., & Fridrich, J. (2011). Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes. *IEEE Transactions on Information Forensics and Security*, *6*(3), 920–935. doi:10.1109/TIFS.2011.2134094

Frdrich, J., & Soukal, D. (2006). Matrix Embedding for Large Payloads. In Proc. International Society for Optics and Photonics, San Jose, CA.

Fridrich, J., Pevný, T., & Kodovský, J. (2007). Statistically Undetectable JPEG Steganography: Dead Ends Challenges, and Opportunities. In *Proc. the 9th workshop on Multimedia and security*, New York, NY (pp. 3-14). doi:10.1145/1288869.1288872

Guo, L. J., Ni, J. Q., & Shi, Y. Q. (2014). Uniform Embedding for Efficient JPEG Steganography. *IEEE Transactions on Information Forensics and Security*, *9*(5), 814–825. doi:10.1109/TIFS.2014.2312817

Guo, L. J., Ni, J. Q., Su, W. K., Tang, C. P., & Shi, Y. Q. (2015). Using Statistical Image Model for JPEG Steganography: Uniform Embedding Revisited. IEEE Trans. *Information Forensics and Security*, *10*(12), 2669–2680. doi:10.1109/TIFS.2015.2473815

Holub, V., & Fridrich, J. (2012). Designing Steganographic Distortion Using Directional Filters. In *Proc. IEEE International Workshop on Information Forensics and Security*, Binghamton, NY (pp. 234-239).

Holub, V., & Fridrich, J. (2013). Digital Image Steganography Using Universal Distortion. In *Proc. the first ACM workshop on Information Hiding and Multimedia Security,* New York, NY (pp. 59-68). doi:10.1145/2482513.2482514

Holub, V., & Fridrich, J. (2014). Low Complexity Features for JPEG Steganalysis Using Undecimated DCT. *IEEE Transactions on Information Forensics and Security*, *10*(2), 219–228. doi:10.1109/TIFS.2014.2364918

Kodovsky, J., Fridrich, J., & Holub, V. (2012). Ensemble Classifiers for Steganalysis of Digital Media. *IEEE Transactions on Information Forensics and Security*, *7*(2), 432–444. doi:10.1109/TIFS.2011.2175919

Li, B., Wang, M., Huang, J. W., & Li, X. (2014). A New Cost Function for Spatial Image Steganography. In *Proc. IEEE International Conference on Image Processing,* Paris, France (pp. 4206-4210). doi:10.1109/ICIP.2014.7025854

Pevný, T., Filler, T., & Bas, P. (2010). Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. In *Proc. 12th International Conference on Information Hiding,* Calgary, Canada (pp. 161-177). doi:10.1007/978-3-642-16435-4_13

Phil, S. (2003). Model-Based Steganography. In *Proc. the Second International Workshop on Digital-forensics and Watermarking,* Seoul, Korea (pp. 154-167).

Schaefer, G., & Stich, M. (2004). UCID - An uncompressed colour image database. In *Proc. Conference on Storage and Retrieval Methods and Applications for Multimedia,* San Jose, CA (pp. 472-480).

Sedighi, V., Fridrich, J., & Cogranne, R. (2015). Content-Adaptive Pentary Steganography Using the Multivariate Generalized Gaussian Cover Model. In Proc. International Society for Optics and Photonics, San Francisco, CA.

Song, X. F., Liu, F. L., Yang, C. F., Luo, X. Y., & Zhang, Y. (2015). Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters. In *Proc. the 3rd ACM Workshop on Information Hiding and Multimedia Security,* New York, NY (pp. 15-23). doi:10.1145/2756601.2756608

Wang, Z. C., Lv, J. P., Wei, Q. D., & Zhang, X. P. (2016). Distortion Function for Spatial Image Steganography Based on the Polarity of Embedding Change. In *The 15th International Workshop on Digital-forensics and Watermarking*, Beijing, China (pp. 487-493).

Wang, Z. C., Zhang, X. P., & Yin, Z. X. (2016). Hybrid Distortion Function for JPEG Steganography. *Journal of Electronic Imaging*, *25*(5), 050501. doi:10.1117/1.JEI.25.5.050501

Westfeld, A. (2001). F5—A Steganographic Algorithm. In *Proc. 4th International Workshop on Information hiding*, Pittsburgh, PA (pp. 289-302). doi:10.1007/3-540-45496-9_21

Zhang, W. M., Zhang, X. P., & Wang, S. Z. (2008). Maximizing Steganographic Embedding Efficiency by Combining Hamming Codes and Wet Paper Codes. In *Proc. 10th International Workshop on Information Hiding,* Santa Barbara, CA (pp. 60-71). doi:10.1007/978-3-540-88961-8_5

Zhang, X. P. (2016). Behavior Steganography in Social Network. In *Proc. the 12th International Conference on Intelligent Information Hiding and Multimedia Signal Processing,* Kaohsiung, Taiwan (pp. 21-23).

Zhang, X. P., & Wang, S. Z. (2006). Efficient Steganographic Embedding by Exploiting Modification Direction. *IEEE Communications Letters*, *10*(11), 781–783. doi:10.1109/LCOMM.2006.060863

Zichi Wang received the BS degree in electronics and information engineering from Shanghai University, China, in 2014, and the MS degree in signal and information processing from the same university in 2017. He is currently pursuing the PhD degree at Shanghai University, China. His research interests include steganography, steganalysis and reversible data hiding. He has published about 10 papers in these areas.

Zhaoxia Yin received the BS degree in computer science and technology from Anhui University, China, in 2005, and the MS and PhD degrees in computer application technology from the same university in 2010 and 2014, respectively. Her research interests include information hiding and image processing. She has published about 20 papers in these areas.

Xinpeng Zhang received the BS degree in computational mathematics from Jilin University, China, in 1995, and the ME and PhD degrees in communication and information system from Shanghai Univer-sity, China, in 2001 and 2004, respectively. Since 2004, he has been with the faculty of the School of Communication and Information Engineering, Shanghai Univer-sity, where he is currently a professor. His research interests include information hiding, image processing, and digital forensics. He has published over 200 papers in these areas.