Adaptive Steganalysis Against WOW Embedding Algorithm

Weixuan Tang
School of Information Science and Technology
Sun Yat-sen University
Guangzhou, P.R. China
tweix@mail2.sysu.edu.cn

Haodong Li
School of Information Science and Technology
Sun Yat-sen University
Guangzhou, P.R. China
lihaod@mail2.sysu.edu.cn

Weiqi Luo*
School of Software
Sun Yat-sen University
Guangzhou, P.R. China
weiqi.luo@yahoo.com

Jiwu Huang
College of Information Engineering
Shenzhen University
Shenzhen, P.R. China
jwhuang@szu.edu.cn

ABSTRACT

WOW (Wavelet Obtained Weights) [5] is one of the advanced steganographic methods in spatial domain, which can adaptively embed secret message into cover image according to textural complexity. Usually, the more complex of an image region, the more pixel values within it would be modified. In such a way, it can achieve good visual quality of the resulting stegos and high security against typical steganalytic detectors. Based on our analysis, however, we point out one of the limitations in the WOW embedding algorithm, namely, it is easy to narrow down those possible modified regions for a given stego image based on the embedding costs used in WOW. If we just extract features from such regions and perform analysis on them, it is expected that the detection performance would be improved compared with that of extracting steganalytic features from the whole image. In this paper, we first proposed an adaptive steganalytic scheme for the WOW method, and use the spatial rich model (SRM) based features [4] to model those possible modified regions in our experiments. The experimental results evaluated on 10,000 images have shown the effectiveness of our scheme. It is also noted that our steganalytic strategy can be combined with other steganalytic features to detect the WOW and/or other adaptive steganographic methods both in the spatial and JPEG domains.

Categories and Subject Descriptors

I.4 [**Image Processing and computer vision**]

*Corresponding author

IH&MMSec'14, June 11–13, 2014, Salzburg, Austria.

http://dx.doi.org/10.1145/2600918.2600935.

Keywords

Adaptive Steganography; WOW; Adaptive Steganalysis; SRM; Texture Complexity

1. INTRODUCTION

Steganaography has improved greatly with the development of steganalysis. LSB (Least Significant Bit) replacement is the simplest steganography. Though it can easily cheat our human eyes, it brings some artifacts into resulting images, and thus it is easy to be detected even at a low embedding rate using some steganalytic methods, such as the Chi-squared attack [13] and regular/singular groups (RS) analysis [3]. A minor modification to LSB replacement called LSB matching is proposed to avoid such obvious artifacts, and has been proved to be a more secure method compared to the LSB replacement. Subsequently, some steganographic methods have been proposed to improve the embedding efficiency and/or capacity, for instance LSB matching revisited [9] and PVD (pixel value difference) based method [14].

All the above-mentioned methods can be regarded as non-adaptive methods, since the locations of modified pixels with these methods are mainly dependent on a pseudorandom number generator. Their security performances against some advanced steganalytic features such as SPAM [10] , SRM [4], PSRM [6] or LBP based method [12] are still far from satisfactory, especially when the embedding rate is high. To improve the security, some adaptive embedding methods have been proposed. The basic idea of such adaptive methods is that preferentially modifying those complex textural regions that are hard to model, while keeping those smooth and flat regions as they are when performing data hiding. For instance, Luo *et al.* proposed an edge adaptive image steganography [8] based on LSB matching revisited. Recently, Pevny *et al.* proposed a novel strategy for adaptive steganography [11]. The strategy firstly builds a distortion function in the spatial domain to assign pixel costs by measuring the impact of changing each pixel in a feature space under investigation, and then combines with the advanced Syndrome-Trellis Codes (STCs) codding tech-

nique [2] to minimize the expected distortion for all pixels in an image. Based on this strategy, two modern steganographic methods *i.e.* HUGO [11] and WOW [5] have been proposed. The main difference of the two methods is the design of distortion function. The SPAM feature [10] model is used in HUGO, while for WOW, three directional wavelet filters have been used for obtaining the embedding costs for each pixel. Based on the results in [5], the WOW embedding algorithm achieves the best security performance in spatial domain evaluated by the modern steganalytic high-dimensional rich models SRM [4].

In this paper, we propose an adaptive steganalytic strategy for the WOW embedding algorithm. Based on our experiments, we found that like other adaptive steganograhic methods, most embedding changes would be highly concentrated on those complex textural or "noisy" regions using the WOW. In such a way, the visual quality of the resulting stego images and the security against some typical steganalytic methods are improved significantly compared with those non-adaptive ones, since embedding changes in the "noisy" regions are insensitive to our eyes and those regions are relatively difficult to model. However, this advantage would inevitably lead to a loophole in the WOW embedding algorithm. That is, it is possible to narrow down those suspicious regions (*i.e.* in those "noisy" regions) for a given questionable image. Even though those suspicious regions may be difficult to model, the relative modification rates (*i.e.* the number of modified pixels after data hiding over the number of pixels in the suspicious regions) would be increased significantly. Therefore, it is expected that the detection performance may be improved if the steganalytic features are extracted from those possible modified regions rather than the whole image like the typical steganalytic methods, and this is the main idea of the proposed steganalytic strategy.

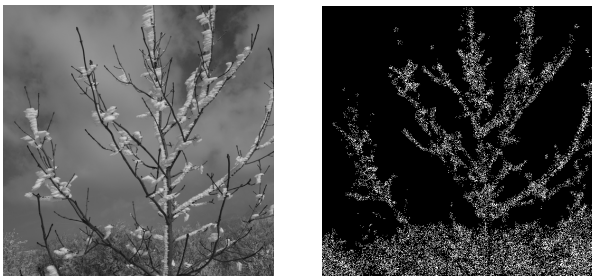
The rest of this paper is arranged as follows. Section 2 describes how to narrow down the suspicious regions, Section 3 describes the proposed adaptive steganalytic strategy; Section 4 shows the experimental results. Finally, the concluding remarks and future works are given in Section 5.

2. LOCATION OF SUSPICIOUS REGIONS

In this Section, we firstly give a brief overview of the WOW embedding algorithm, and then propose a method to locate those suspicious regions based on the embedding costs used in WOW.

The WOW embedding algorithm works as follows. Firstly, three directional filters (denoted $K^{(k)}, k = 1, 2, 3$ using Daubechies 8 wavelets) are performed on the cover image X to obtain the LH, HL and HH directional residuals $R^{(k)} = K^{(k)} * X$, respectively. Here, the $*$ denotes the convolution mirror-padded operation. And then the embedding suitability $\xi_{ij}^{(k)}$ for each pixel can be obtained by measuring the difference between $R^{(k)}$ and the same residual after changing only one pixel at ij (denoted $R_{[ij]}^{(k)}$) by the wavelet coefficient itself.

$$\xi_{ij}^{(k)} = |R^{(k)}| * |R^{(k)} - R_{[ij]}^{(k)}| \tag{1}$$



(a) Cover Image (b) Modifications

Figure 1: Illustration of cover image and the modifications after using the WOW with 0.4bpp

Then the embedding costs ρ_{ij} are computed by aggregating three suitability $\xi_{ij}^{(k)}$, $k = 1, 2, 3$ by

$$\rho_{ij}^{(p)} = (\sum_{i=1}^{3} |\xi_{ij}^{(k)}|^p)^{-1/p}), p = -1 \tag{2}$$

Finally, the STCs is applied to minimize the following distortion function and get the resulting stego image Y.

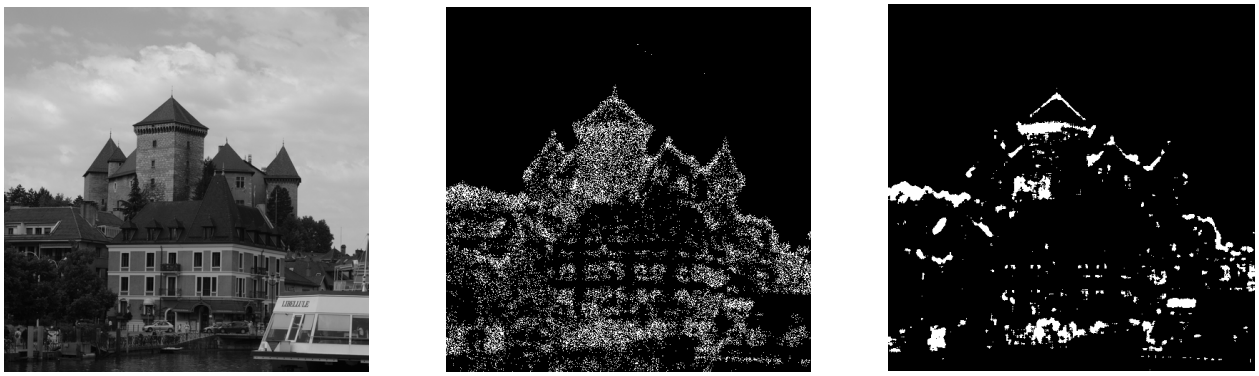
$$D(X, Y) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \rho_{i,j}(X, Y_{i,j}) |X_{i,j} - Y_{i,j}| \tag{3}$$

where $n_1 \times n_2$ denotes the dimension of cover image X; $\rho_{i,j}$ denote the costs of changing pixel X_{ij} to Y_{ij}, the WOW limits the embedding changes to ± 1.

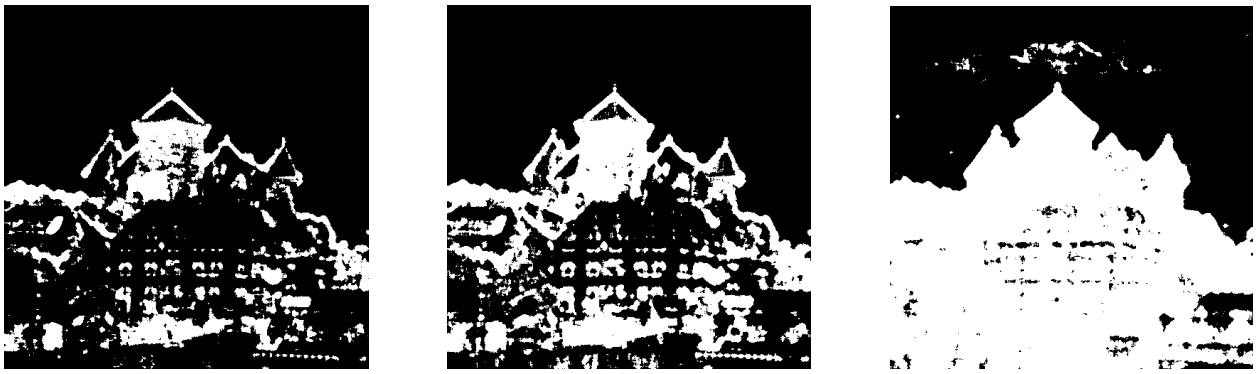
Usually, the $\rho_{i,j}^{(-1)}$ is smaller for those pixels located at the regions with more textural complexity, and thus the modifications after minimizing the above distortion function with STCs would be concentrated on textural regions, just as illustrated in Fig. 1. Since the textural regions are relatively difficult to model compared with smooth/flat regions, and the use of STCs can significantly reduce the embedding changes compared with typical methods, such as LSB Matching, the WOW is currently the most secure steganography in spatial domain.

Please note that in all universal steganalytic methods such as [4] and [12], features are extracted from the whole image, meaning that the contribution for every pixel in an image is assumed the same. Based on above analysis, however, we found that most embedding changes with the WOW embedding algorithm would be highly located at the textural regions, while lots of smooth and flat regions would not change at all. The typical universal methods may not be suitable for such adaptive methods. If we can firstly remove those smooth and flat regions, just consider those suspicious regions that are probably changed with WOW embedding algorithm, it is expected that the steganalysis performance would be better. Therefore, how to narrow down the suspicious regions is one of the key issues in the proposed steganalytic strategy. Fortunately, the embedding costs $\rho_{ij}^{(-1)}$ used in the WOW can help us deal with the problem effectively.

For a given stego image Y, we try to locate those possible modified pixels based on the embedding costs calculated



(a) Cover Image (b) Modifications (0.4bpp) (c) $p = 5\%$



(d) $p = 15\%$ (e) $p = 25\%$ (f) $p = 55\%$

Figure 2: Illustrations of modifications and the locations of those selected pixels at different parameter p

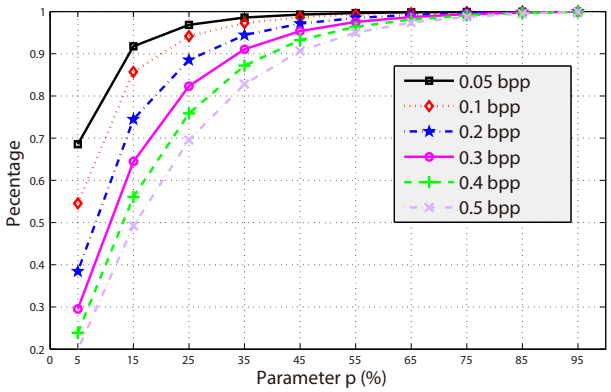
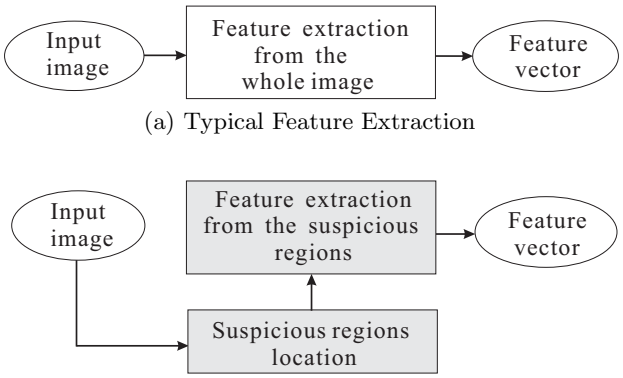


Figure 3: The percentages of modified pixels located at selected pixels with increasing the parameter p

from Y [1]. According to the adaptive rule of WOW, the smaller the ρ_{ij} is, the higher probability the pixel (i, j) would be modified. To test the effectiveness of using ρ_{ij} for locating those modified pixels, we firstly sort the ρ_{ij} from the smallest to largest, and then we just select a small proportion p of pixels with smaller embedding costs, where $p \in (0, 1]$ denotes the proportion of selected pixels in an image; and calculate the percentage that the number of modified pixels located at the selected pixels over all modified pixels. An example is shown in Fig. 2. It is observed that the selected pixels (see Fig. 2(c)-(f)) can estimate the location of modified pixels (see Fig. 2(b)) effectively. In this example, the corresponding percentages are 23.85%, 59.78%, 83.94% and 100% for $p = 5\%, 15\%, 25\%$ and 55%, respectively, which means that $45\% = 1 - 55\%$ of pixels in this image would not change at all when the embedding rate is 0.4bpp.

Obviously, the percentage depends on the embedding rate, the parameter p, and image content itself. To provide more convincing results, we show the average results evaluated on 10,000 images from BOSSBase ver. 1.01 [1]. In this experiment, the embedding rates are ranging from 0.05bpp to 0.5bpp, and p is ranging from 5% to 95% with a step 10%. The results are shown in Fig. 3. It is observed that the percentages would increase with increasing the proportion p for the six embedding rates. When p becomes 35%, all percentages are larger than 82%, and they are larger than 95% when p increases to 55%, which means that most modified pixels are concentrated on those pixels with small embedding costs. Therefore, it is possible to remove lots of unchanged pixels based on the embedding cost.

[1] Although the cover image X is not available, we can obtain the approximative ρ_{ij} from the stego image Y, since the difference between X and Y is very small.



(a) Typical Feature Extraction

(b) Adaptive Feature Extraction

Figure 4: Comparison of the typical steganalytic feature extraction and the proposed adaptive strategy

3. THE PROPOSED ADAPTIVE STEGANALYTIC STRATEGY

As described in previous Section 2, we found that lots of pixels with larger embedding costs would not changes at all with the WOW embedding algorithm. It is expected that their contributions to steganalysis would be low. Thus the proposed adaptive steganalytic strategy for WOW is illustrated in Fig. 4. It is observed that the main difference between the proposed strategy and the typical one is that we just focus on those probably modified pixels (with a proportion p) according to the embedding costs for feature extraction. In such a case, the typical strategy can be regarded as a special case of our strategy when $p = 100\%$.

Please note that there is an important parameter p (the proportion of image pixels that we selected for analysis, where $p \in (0, 1]$) in the proposed strategy. Based on our experiments, it would affect the detection performances significantly. The reason is that when the p is smaller, the selected pixels may not be sufficient for extracting effective features. However, the relative modified rate (*i.e.* the number of modified pixels over the number of selected pixels) would become larger. Please refer to the average relative modified rates evaluated on BOSSBase [1] in Table 1. It is observed that for a given embedding rate, the relative modified rates will decrease with increasing the parameter p. For instance, when embedding rate is 0.05 bpp, the relative modified rate for $p = 5\%$ is over 14 times ($\approx 11.73/0.83$) of that for $p = 100\%$ (*i.e.* original method), while the number of selected pixels is just 5/100 of the original one in this case. Since both the number of selected pixels and the relative modified rates would affect the detection performances, we should carefully select the parameter p. In the following Section 4, some experimental results would be given to show how the parameter p affects the detection performances.

4. EXPERIMENTAL RESULTS

In the experiments, 10,000 original images with size of 512×512 are from BOSSBase ver. 1.01 [1]. The spatial rich model (SRM) based features [4] is used for feature extraction, and the ensemble classifier [7] is used in the training and testing stages, and the detection performance is quantified using the ensemble's "out-of-bag" error E_{OOB} as it did in [5]. Two following experiments have been carried out.

Table 1: The relative modified rates for different parameters p and embedding rates

Embedding rate	$p =$ 5%	$p =$ 25%	$p =$ 55%	$p =$ 100%
0.05 bpp	11.73%	3.22%	1.50%	0.83%
0.10 bpp	20.33%	6.85%	3.26%	1.80%
0.20 bpp	31.20%	14.26%	7.13%	3.97%
0.30 bpp	38.04%	21.27%	11.31%	6.36%
0.40 bpp	42.95%	27.52%	15.72%	8.93%
0.50 bpp	46.71%	32.93%	20.92%	11.67%

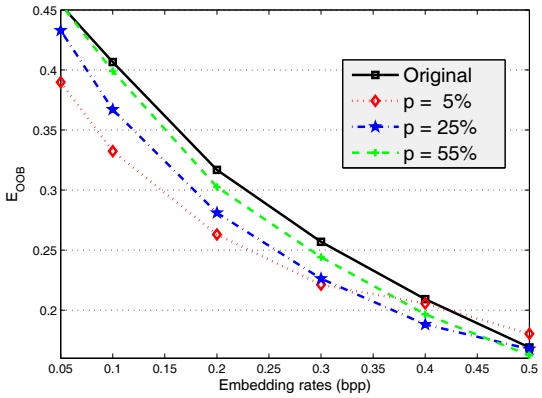


Figure 5: The detection error E_{OOB} with increasing the embedding rates from 0.05 to 0.50 bpp

4.1 Experiment #1

In this experiment, we fix the parameter p with three different values, *i.e.* 5%, 25%, 55%, and we try to compare the average detection errors with the original SRM method [4] (*i.e.* $p = 100\%$). The embedding rates of 0.05, 0.10, 0.20, 0.30, 0.40 and 0.50 bpp, have been evaluated, respectively. The experimental results are shown in Fig. 5.

From Fig. 5, it is observed that the proposed strategy with the three different parameters outperforms the original SRM in almost all cases (except for a case of $p = 5\%$ and the embedding rate is 0.50bpp), especially when the embedding rate is low, such as lower than 0.30 bpp. On average, we obtain an improvement of around 3.7%, 2.5% and 0.9% for $p = 5\%, 35\%$ and 55%, respectively.

4.2 Experiment #2

In this experiment, we fix the embedding rate with 0.05, 0.10, 0.20, 0.30, 0.40 and 0.50 bpp, respectively, and compare the average detection error with different parameter p, which ranging from 5% to 95% with a step 10%. We try to analyze the best parameter p for a given embedding rate. The experimental results are shown in Fig. 6.

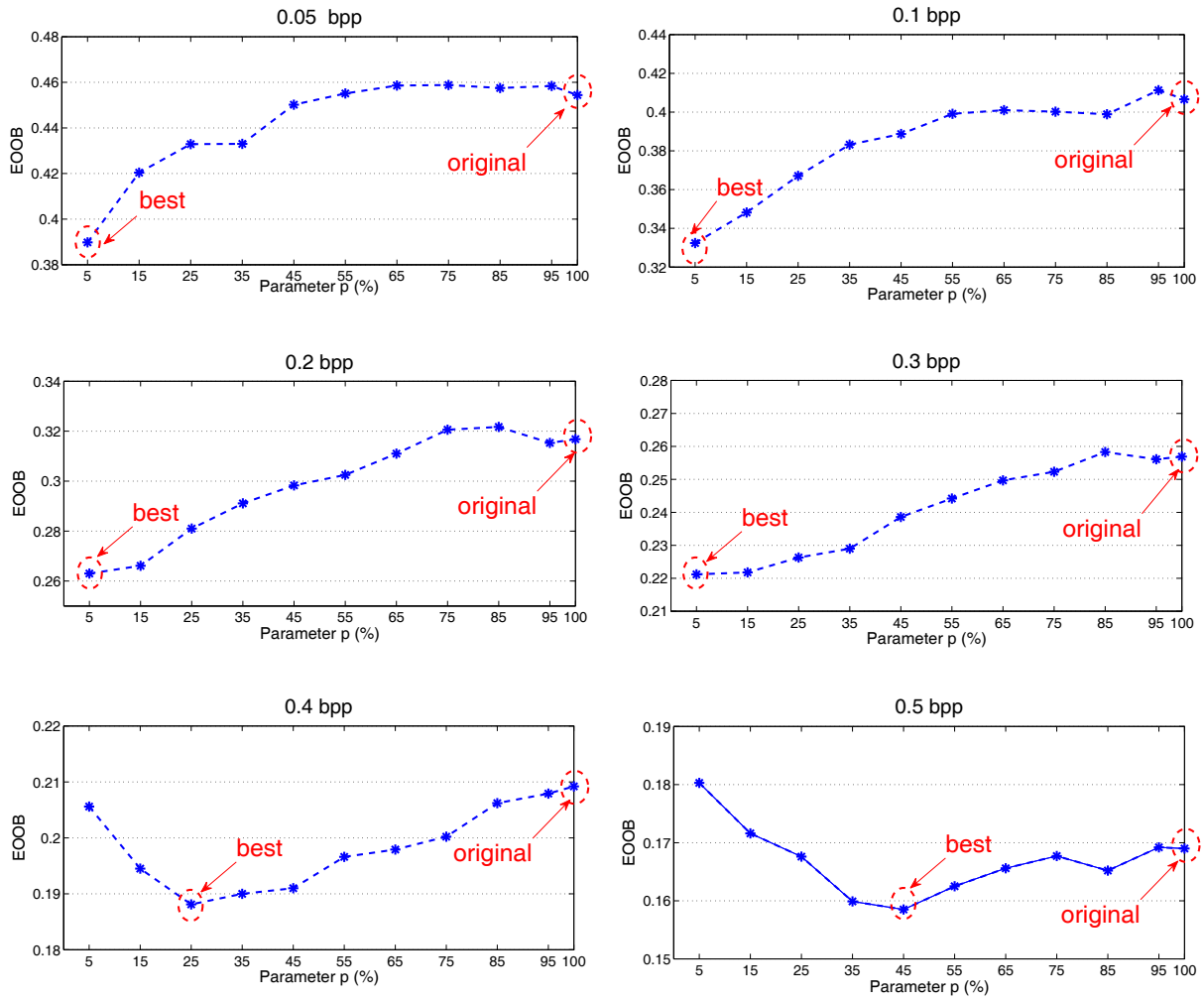


Figure 6: The detection error E_{OOB} with increasing the parameter p from 5% to 95%

Table 2: Average detection improvements with different parameters p

Embedding rate	$p \in \{5\%, \ldots 55\%\}$	$p \in \{5\%, \ldots 95\%\}$	$p \in$ Best one
0.05 bpp	2.42%	1.29%	6.46%
0.10 bpp	3.68%	2.36%	7.42%
0.20 bpp	3.31%	1.97%	5.38%
0.30 bpp	2.67%	1.72%	3.57%
0.40 bpp	1.49%	1.14%	2.11%
0.50 bpp	0.23%	0.22%	1.05%

From Fig. 6, it is also observed that the proposed strategy works better than the original method in most cases. As highlighted in the Fig. 6, however, the best parameter p is different for different embedding rates. For instance, the best p is 5% when the embedding rate is 0.05, 0.10, 0.20 and 0.30 bpp. While it increases to 25% and 45% when the embedding rate increases to 0.40 and 0.50 bpp, respectively.

For a given stego image, the best parameter p is not available since the embedding rate is unknown. Thus, we need to restrict the range of selected p. Two different ranges are considered here: Range #1: from 5% to 55% (in this case, over 95% modified pixels would be located at the selected pixels based on the results shown in Fig.3) with a step 10%; Range #2: from 5% to 95% with a step 10%. Then the average improvements and the best improvements are shown in Table 2. Both Fig. 6 and Table 2 show that we can obtain better detection results with smaller p (*e.g.* in range # 1) when the embedding rate is less than 0.5 bpp.

5. CONCLUDING REMARKS

Based on our experiments and analysis, we found that the modified pixels after using the adaptive WOW steganography would highly located at those textural/noisy regions that are difficult to model and insensitive to our human eyes. This embedding property can significantly improve the security performance against typical steganalysis as well as the visual quality of the resulting stego images compared

with the non-adaptive steganography. However, it may be a loophole for the detector, since it is possible to narrow down the suspicious regions for steganalysis. In this paper, we propose an adaptive steganalytic strategy for the WOW embedding algorithm based on this loophole. The proposed strategy tries to restrict the feature extraction on those pixels with low embedding costs rather than the whole image. The experimental results evaluated on 10,000 images show that the proposed strategy can improve the effectiveness of the typical steganalysis, such as SRM, especially when the embedding rate is low than 0.40 bpp.

Please note that the proposed strategy is flexible. In the next step, we would extend other steganalytic features such as PSRM [6] and LBP based features [12] in our strategy, and evaluate whether or not the proposed strategy still works for other steganography, such as edge adaptive method [8] and HUGO [11]. Besides, the relationship between the parameter p and the detection performance needs further analysis. Furthermore, we would further analyze the common loophole of the adaptive steganogrpahy, and propose an improved adaptive steganalytic strategy.

6. ACKNOWLEDGMENTS

This work is supported by National Science & Technology Pillar Program (2012BAK16B06), NSFC (U1135001, 61332012, 61272191), the funding of Zhujiang Science and technology (2011J2200091), and the Guangdong NSF (S2013010012039).

7. REFERENCES

[1] P. Bas, T. Filler, and T. Pevny. Break our steganographic system. In *Information Hiding*, volume 6958 of *Lecture Notes in Computer Science*, pages 59–70. Springer Berlin Heidelberg, 2011.

[2] T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Trans. on Information Forensics and Security*, 6(3):920–935, 2011.

[3] J. Fridrich, M. Goljan, and R. Du. Detecting LSB steganography in color, and gray-scale images. *IEEE Multimedia*, 8(4):22–28, Oct. 2001.

[4] J. Fridrich and J. Kodovsky. Rich models for steganalysis of digital images. *IEEE Trans. on Information Forensics and Security*, 7(3):868–882, Jun. 2011.

[5] V. Holub and J. Fridrich. Designing steganographic distortion using directional filters. In *2012 IEEE International Workshop on Information Forensics and Security(WIFS)*, pages 234–239, 2012.

[6] V. Holub, J. Fridrich, and T. Denemark. Random projections of residuals as an alternative to co-occurrences in steganalysis. *IEEE Trans. on Information Forensics and Security*, 8(12):1996–2006, 2013.

[7] J. Kodovsky, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE Trans. on Information Forensics and Security*, 7(2):432–444, 2012.

[8] W. Luo, F. Huang, and J. Huang. Edge adaptive image steganography based on LSB matching revisited. *IEEE Trans. on Information Forensics and Security*, 5(2):201–214, Jun. 2010.

[9] J. Mielikainen. LSB matching revisited. *IEEE Signal Processing Letters*, 13(5):285–287, May 2006.

[10] T. Pevny, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. *IEEE Trans. on Information Forensics and Security*, 5(2):215–224, 2010.

[11] T. Pevny, T. Filler, and P. Bas. Using high-dimensional image models to perform highly undetectable steganography. In *Information Hiding*, volume 6387 of *Lecture Notes in Computer Science*, pages 161–177. Springer Berlin Heidelberg, 2010.

[12] Y. Q. Shi, P. Sutthiwan, and L. Chen. Textural features for steganalysis. In *Information Hiding*, volume 7692 of *Lecture Notes in Computer Science*, pages 63–77. Springer Berlin Heidelberg, 2013.

[13] A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In *Information Hiding*, volume 1768 of *Lecture Notes in Computer Science*, pages 61–76. Springer Berlin Heidelberg, 2000.

[14] D.-C. Wu and W.-H. Tsai. A steganographic method for images by pixel-value differencine. *Pattern Recognition Letters*, 24:1613–1626, Jun. 2003.