

A NEW SPATIAL STEGANOGRAPHIC SCHEME BY MODELING IMAGE RESIDUALS WITH MULTIVARIATE GAUSSIAN MODEL

Xinghong Qin¹, Bin Li^{1,2*}, Jiwu Huang¹

¹Guangdong Key Laboratory of Intelligent Information Processing
and Shenzhen Key Laboratory of Media Security, College of Information Engineering,
Shenzhen University, Shenzhen 518060, China.

²Peng Cheng Laboratory, Shenzhen 518024, China.

ABSTRACT

Embedding costs used in content-adaptive image steganographic schemes can be defined in a heuristic way or with a statistical model. Inspired by previous steganographic methods, i.e., MG (multivariate Gaussian model) and MiPOD (minimizing the power of optimal detector), we propose a model-driven scheme in this paper. Firstly, we model image residuals obtained by high-pass filtering with quantized multivariate Gaussian distribution. Then, we derive the approximated Fisher Information (FI). We show that FI is related to both Gaussian variance and filter coefficients. Lastly, by selecting the maximum FI value derived with various filters as the final FI, we obtain embedding costs. Experimental results show that the proposed scheme is comparable to existing steganographic methods in resisting steganalysis equipped with rich models and selection-channel-aware rich models. It is also computational efficient when compared to MiPOD, which is the state-of-the-art model-driven method.

Index Terms— Steganography, steganalysis, multivariate Gaussian model, spatial images

1. INTRODUCTION

Steganography is the technique of secretly conveying messages through digital media, and it receives challenges from the technique of steganalysis which aims to reveal its presence [1–5]. Most of existing content-adaptive image steganographic methods are based on a distortion-minimization framework [6, 7], in which the distortion function can be designed to associate cost with data embedding impact. In most schemes [8–10], data embedding changes are distributed in the noisy/complex regions of an image through heuristically defining low embedding costs in such regions. On the other hand, some schemes try to define costs using a statistical model. In [11], a model-driven approach named MG

(multivariate Gaussian model) was proposed to model the image elements as a sequence of independent Gaussian variables and try to minimize the Kullback-Leibler (KL) divergence between the cover and the stego objects with the corresponding Fisher information (FI) [1]. The performance is comparable to HUGO (highly undetectable steGO) [12]. As an extension of MG, a multivariate generalized Gaussian (MVGG) model was adopted in [13] and its performance is better than MG. In [14], a more advanced method called MiPOD (minimizing the power of optimal detector) was proposed to model the image noises and minimize the detectability of the optimal detector. Its performance is comparable to HILL (high-low-low) [10] against steganalytic schemes such as the spatial rich model (SRM) [15] and maxSRMd2 [16]. However, it is required to estimate noise variance and it is rather complicated and less efficient to perform the estimation.

Inspired by MG and MiPOD, we propose a steganographic scheme in this paper by modeling image residuals, which are obtained by filtering an image with high-pass filters, with multivariate Gaussian distribution. Since steganalysis benefits from extracting effective features from image residuals, our method, abbreviated as MGR (multivariate Gaussian for residuals), aims to better preserve the statistical model of an image by directly associating embedding costs with image residuals. The distribution of stego image residuals can be approximately derived from the embedding change probabilities associated with image pixels. In this way, FI can be efficiently obtained by using the estimated local variance of residuals and the corresponding high-pass filter coefficients. We use several high-pass filters to obtain the residuals, so that the maximum FI value can be obtained to further enhance security performance. This manner can be considered as a countermeasure to resist steganalysis utilizing multiple filters, such as SRM. Experiments show that the proposed method performs well and has low computation complexity.

The rest of this paper is organized as follows. In Section 2, the preliminary knowledge regarding the related work of MG is briefly reviewed. The proposed method is presented in Section 3 and experiments are given in Section 4. This paper

*B. Li is the correspondence author. This work was supported in part by NSFC (Grant 61572329, 61872244, U1636202, 61772349) and in part by the Shenzhen R&D Program (Grant JCYJ20160328144421330).

is concluded in Section 5.

2. PRELIMINARIES

In MG [11], a cover image is modeled as a sequence of n -independent random variables $\mathbf{X} = (X_1, \dots, X_n)$, each of which is distributed as quantized zero-mean Gaussian with variance ν_i , denoted by $Q_\Delta(\mathcal{N}(0, \nu_i))$, where Q_Δ is a uniform scalar quantizer with quantization step Δ . Denote $p^{(i)} = \{p_j^{(i)}\}$ and $q^{(i)} = \{q_j^{(i)}\}$ ($j \in \mathcal{M} = \{k\Delta | k \in \mathbb{Z}\}$) the probability mass function (PMF) of cover X_i and stego Y_i , respectively. For a large n and small embedding change probabilities β_i , the total KL divergence between the cover and the stego can be approximated by:

$$\sum_{i=1}^n D_{KL}(p^{(i)} || q^{(i)}) = \frac{1}{2} \sum_{i=1}^n \beta_i^2 I_i(0), \quad (1)$$

where

$$I_i(0) = \sum_j \frac{1}{p_j^{(i)}} \left(\frac{\partial q_j^{(i)}}{\partial \beta_i} \Big|_{\beta_i=0} \right)^2 \quad (2)$$

is the FI [1]. We limit our discussion to ternary embedding, where the embedding change $N_i = Y_i - X_i \in \{+1, -1, 0\}$. We assume $\beta_i^+ = \beta_i^- = \beta_i$ and $\beta_i^0 = 1 - 2\beta_i$. Under such an assumption, the PMF of the stego is

$$q_j = (1 - 2\beta_i)p_j + \beta_i(p_{j+1} + p_{j-1}). \quad (3)$$

Denote $f_\nu(x)$ the zero-mean Gaussian probability density function with the variance ν . The quantization performs as

$$F_\Delta(x) = \int_{x-\Delta/2}^{x+\Delta/2} f_\nu(x) dx. \quad (4)$$

Using the mean value theory (MVT), we have

$$p_j = F_\Delta(j\Delta) = \Delta f_v(j'\Delta) \quad (5)$$

for some $j' \in (x - \Delta/2, x + \Delta/2)$. Using Taylor expansion of $F_\Delta(x)$ at $x = j\Delta$, we have

$$p_{j\pm 1} = \sum_{l=0}^{\infty} F_\Delta^{(l)}(j\Delta) \frac{(\pm\Delta)^l}{l!}. \quad (6)$$

Therefore, we have

$$\frac{\partial q_j^{(i)}}{\partial \beta_i} \Big|_{\beta_i=0} = -2p_j + p_{j-1} + p_{j+1} = \Delta^3 f_v''(j\Delta) + \mathcal{O}(\Delta^4), \quad (7)$$

and

$$I_i(0) \approx \sum_j \frac{\Delta^6 (f_v''(j\Delta))^2}{\Delta f_v(j'\Delta)} \approx \frac{\Delta^4}{\nu^2} \quad (8)$$

The minimization problem of (1) is subjected to the payload constraint as

$$\alpha n = \sum_{i=1}^n h(\beta_i), \quad (9)$$

where $h(\beta_i) = -2\beta_i \ln \beta_i - (1 - 2\beta_i) \ln(1 - 2\beta_i)$ is the entropy function and α is the relative payload. β_i and λ can be numerically obtained using Lagrange multipliers, and embedding costs can be computed as

$$\xi_i = \frac{1}{\lambda} \ln\left(\frac{1}{\beta_i} - 2\right). \quad (10)$$

With the costs, practical embedding codes such as STC (syndrome trellis code) [7], or embedding simulator can be used.

3. RESIDUAL MODEL-BASED STEGANOGRAPHY

3.1. Image Filtered Residuals

It is common in steganalysis to use high-pass filters to obtain image residuals for extracting effective features [15, 17]. Therefore, we aim to minimize embedding distortion by better preserving the statistical model of image residuals. Assume $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where \mathbf{N} is the embedding changes. When applying a high-pass filter \mathbf{H} on the stego image, we arrive

$$\boldsymbol{\eta}_Y = \mathbf{Y} \otimes \mathbf{H} = (\mathbf{X} + \mathbf{N}) \otimes \mathbf{H} = \boldsymbol{\eta}_X + \mathbf{N} \otimes \mathbf{H}, \quad (11)$$

where $\boldsymbol{\eta}_X = (\eta_{X_1}, \dots, \eta_{X_n})$ and $\boldsymbol{\eta}_Y = (\eta_{Y_1}, \dots, \eta_{Y_n})$ are the filtered residuals of cover and stego, respectively. The 2-D high-pass filter is formed as:

$$\mathbf{H} = \begin{bmatrix} a_{11} & \cdots & a_{1S} \\ \cdots & \cdots & \cdots \\ a_{R1} & \cdots & a_{RS} \end{bmatrix}, a_{uv} \in \mathbb{Z}, \sum_{u=1}^R \sum_{v=1}^S a_{uv} = 0. \quad (12)$$

3.2. Obtaining Embedding Costs from Residual Model

Different from MG [11] in which image pixels were modeled by Gaussian distribution, we model the filtered residuals as zero-mean quantized Gaussian distributions $\eta_{X_i} \sim Q_\Delta(\mathcal{N}(0, \nu_i))$. Without loss of generality, we reuse the symbols $p^{(i)} = \{p_j^{(i)}\}$ and $q^{(i)} = \{q_j^{(i)}\}$ ($j \in \mathcal{M}$) defined in Section 2 to denote the PMF of η_{X_i} and that of η_{Y_i} , respectively.

Assume the probabilities of embedding changes are almost the same in local region, *i.e.*, $\beta_t \approx \beta_i$ for $t \in \mathcal{H}$, where \mathcal{H} is the neighborhood of η_{X_i} defined by the support of the high-pass filter \mathbf{H} . This is reasonable since by applying the spreading rule [5] with a low-pass filter to smooth the costs such as HILL, or smooth the FI such as MiPOD, the embedding probabilities will become uniform in local region. Similar to (3), the distribution of η_{Y_i} is derived as

$$\begin{aligned} q_j &= \beta_i^0 p_j + \frac{1}{R \times S} \sum_{u=1}^R \sum_{v=1}^S (\beta_{uv}^+ p_{j+a_{uv}} + \beta_{uv}^- p_{j-a_{uv}}) \\ &\approx (1 - 2\beta_i) p_j + \frac{\beta_i}{R \times S} \sum_{u=1}^R \sum_{v=1}^S (p_{j+a_{uv}} + p_{j-a_{uv}}). \end{aligned} \quad (13)$$

Following the process from (4) to (8), we have

$$p_{j\pm k} = \sum_{l=0}^{\infty} F_{\Delta}^{(l)}(j\Delta) \frac{(\pm k\Delta)^l}{l!} \quad (14)$$

$$= p_j + \frac{k^2 \Delta^3 f_{\nu}''(j\Delta)}{2} + \mathcal{O}((k\Delta)^4).$$

By taking the partial derivative of q_j , we have

$$\frac{\partial q_j}{\partial \beta_i} \Big|_{\beta_i=0} = -2p_j + \frac{1}{R \times S} \sum_{u=1}^R \sum_{v=1}^S (p_{j+a_{uv}} + p_{j-a_{uv}})$$

$$= \frac{\Delta^3 f_{\nu}''(j\Delta)}{R \times S} \sum_{u=1}^R \sum_{v=1}^S a_{uv}^2 + \mathcal{O}(\Delta^4). \quad (15)$$

Consequently, the steganographic FI associated with a high-pass filter \mathbf{H} is

$$I_i^{\mathbf{H}}(0) \approx \sum_j \frac{\Delta^6 (f_{\nu}''(j\Delta))^2 (\sum_{u=1}^R \sum_{v=1}^S a_{uv}^2)^2}{(R \times S)^2 \Delta f_{\nu}(j'\Delta)}$$

$$\approx \frac{\Delta^4 (\sum_{u=1}^R \sum_{v=1}^S a_{uv}^2)^2}{(R \times S)^2 \nu_i^2}. \quad (16)$$

Please note that we can consider the obtained FI in (16) as a generalized version of (8). When the filter \mathbf{H} is replaced by an impulse function where $R = S = 1$ and $a_{1,1} = 1$, the obtained expression of $I_i^{\mathbf{H}}(0)$ is reduced to (8).

Compared to MiPOD which requires to compute noise variance in an inefficient way, the variance ν_i in our proposed scheme can be estimated in a simple way as MG by using a 3×3 neighbourhood for estimation as follows:

$$\nu_i = \max\{0.01, E_N(\eta_i^2) - [E_N(\eta_i)]^2\} \quad (17)$$

where E_N is the mean function for neighbourhood and the lower boundary 0.01 is used for preventing zero denominator.

As indicated in [1], a larger value of FI will lower the steganographic performance. As a result, we propose to employ various of filters and use the maximum FI values for each location to compute the embedding costs. The final FI values are computed as

$$I_i(0) = \max\{I_i^{\mathbf{H}_k}(0)\}, \mathbf{H}_k \in \mathbb{H}. \quad (18)$$

where \mathbb{H} denotes the high-pass filter set. As MG, β_i can be solved by minimizing (1) with constraint (9), and the embedding costs ξ_i can be solved according to (10).

In [5], the spreading rule (SR) has been proposed and later implemented in [10, 13, 14] to improve undetectability. We also apply SR for costs by using an average filter \mathbf{L} to obtain the final embedding costs:

$$\rho = \xi \otimes \mathbf{L}. \quad (19)$$

The whole processing pipeline of the proposed scheme is illustrated in Fig. 1.

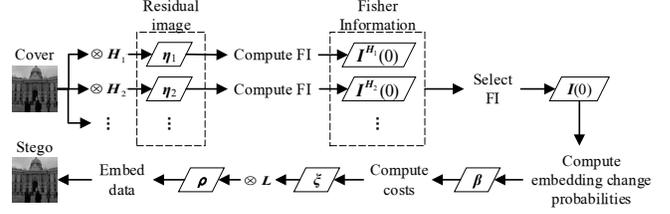


Fig. 1. The processing pipeline of the proposed MGR scheme.

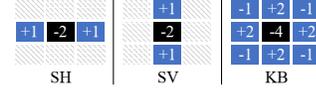


Fig. 2. Typical high-pass filters used in this paper.

Table 1. \bar{P}_E of MG and that of MGR with different high-pass filters under different payload α against SRM. MGR* denotes the scheme using SH, SV, and KB filters together.

α	0.05	0.1	0.2	0.3	0.4	0.5
MG	0.3715	0.2935	0.2131	0.1654	0.1339	0.1119
MGR(SH)	0.4083	0.3467	0.2686	0.2142	0.1733	0.1400
MGR(KB)	0.4327	0.3668	0.2745	0.2066	0.1617	0.1253
MGR(KV)	0.4155	0.3511	0.2485	0.1884	0.1443	0.1129
MGR*	0.4516	0.3951	0.3081	0.2383	0.1882	0.1518

Table 2. \bar{P}_E of MGR* with $h \times h$ average filter under different payloads α against SRM.

α	0.05	0.1	0.2	0.3	0.4	0.5
$h = 3$	0.4584	0.4108	0.332	0.2741	0.2193	0.1782
$h = 5$	0.4653	0.4296	0.358	0.2961	0.2473	0.2020
$h = 7$	0.4668	0.4289	0.3624	0.3015	0.2506	0.2103
$h = 9$	0.4644	0.4276	0.3587	0.2991	0.2488	0.2079
$h = 11$	0.4613	0.4258	0.3565	0.2974	0.2463	0.2065

4. EXPERIMENTS

4.1. Setup

We conducted experiments to verify the performance of the proposed method. All experiments were performed on BOSSBase ver 1.01 database [18] containing 10000 gray-scale images of 512×512 pixels. Five existing content-adaptive steganographic schemes were used for comparison. Three of them were designed heuristically: WOW [8], S-UNIWARD [9] (with the stable constant $\sigma = 1$), and HILL [10]. Two of them were model-based: MG [11] and MiPOD [14]. The ternary optimal embedding simulator was used for all methods. The payload rate was measured by bit per pixel (bpp).

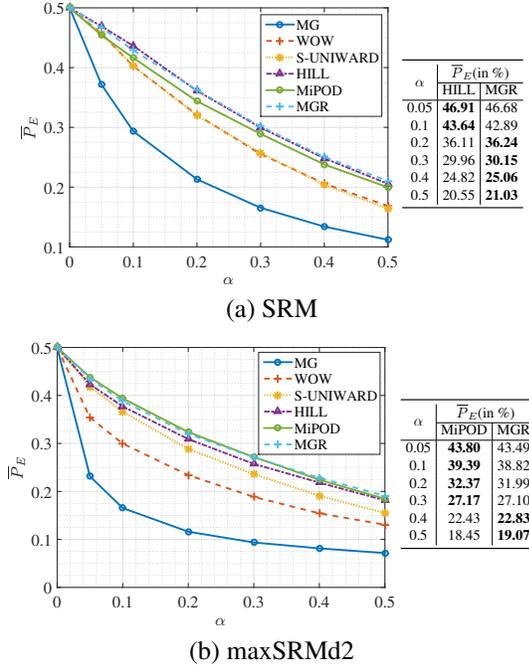


Fig. 3. Steganalytic performance (\bar{P}_E) of the proposed scheme when compared to existing methods.

We used two 34671-D steganalytic feature sets, including SRM [15] and its selection-channel version maxSRMd2 [16], to evaluate security performance. The Fisher linear discriminant based ensemble classifier [19] was used to perform classification. For each steganographic scheme, we randomly split the image set into 5000 cover/stego pairs for training and the rest for testing. The performance was evaluated by the testing error defined as

$$P_E = \min_{P_{FA}} \frac{1}{2} (P_{FA} + P_{MD}), \quad (20)$$

where P_{FA} and P_{MD} were the false-alarm and the missed-detection probabilities, respectively. We repeated the random split of the training and testing sets 10 times and reported the averaged value \bar{P}_E .

4.2. Impact of the parameters

The high-pass filter has an impact on the performance. We used several filters individually, including second-order horizontal derivative (SH) filter, KB filter, KV filter, etc, which were used in [15]. The experimental results are shown in Table 1, and the results of MG are also included for comparison. We can observe that SH and KB are the two best performing filters. As a result, we used three filters, including SH filter, second-order vertical derivative (SV) filter, and KB filter together, which are shown in Fig. 2. The resulting scheme is denoted MGR*. It can be found that the scheme with various filters can improve the performance.

Table 3. The averaged elapsed time (in second) used in computing FI for MiPOD and MGR.

Scheme	MiPOD	MGR
Elapsed time	0.4329	0.0542

The size of the $h \times h$ average filter L also has an impact on the steganographic performance and it is determined experimentally. Table 2 shows the performance of the proposed MGR* against SRM under different filter size. It can be observed that MGR* achieves the best performance when $h = 7$. In the rest experiments, we use the 7×7 average low-pass filter as default. In the rest of the paper, we use MGR to stand for MGR* incorporated L with $h = 7$.

4.3. Comparison to Existing Methods

We compare MGR with WOW, S-UNIWARD, HILL, and MiPOD against SRM and maxSRMd2 under different payloads. Fig. 3 shows the results. We can observe that MGR outperforms MiPOD and acts comparable to HILL against SRM, and outperforms HILL and acts comparable to MiPOD against maxSRMd2.

Comparing MGR with MiPOD, the main difference is the way to obtain FI. We randomly selected 1000 images to evaluate elapsed time of computing FI for both MiPOD and MGR under payload 0.4 bpp. The experiments were performed with a computer with Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz and 32G memory. Averaged results are shown in Table 3. It is worthy to note that the elapsed time of computing FI of MGR is significantly less than that of MiPOD.

5. CONCLUSION

In this paper, an extension of MG [11], called MGR, has been proposed based on modeling image filtered residuals with multivariate Gaussian distribution. Different from MG which models image elements, MGR explicitly considers the KL divergence in terms of image residuals, which are commonly used in steganalysis. It is not surprise to see that the mathematically derived FI is related to both Gaussian variance and high-pass filter coefficients. In order to further improve the performance, various filters can be employed in MGR by considering the maximum FI values. Experimental results have shown that the statistical model built from residual domain gets better performance in resisting residual-based steganalytic methods than its counterpart built from pixel domain, and achieves the best overall performance when compared with HILL and MiPOD.

We note that fusing the FI values from different high-pass filters is still performed in a heuristic manner. More insightful investigation such as making regulation on the filter coefficients should be done in the future work.

6. REFERENCES

- [1] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*, Cambridge University Press, 2010.
- [2] R. Böhme, *Advanced Statistical Steganalysis*, Springer Berlin Heidelberg, 2010.
- [3] B. Li, J. He, J. Huang, and Y. Q. Shi, “A survey on image steganography and steganalysis,” *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, Apr. 2011.
- [4] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevný, “Moving steganography and steganalysis from the laboratory into the real world,” in *Proc. IH&MMSec*, Montpellier, France, June 2013, pp. 45–58.
- [5] B. Li, S. Tan, M. Wang, and J. Huang, “Investigation on cost assignment in spatial image steganography,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 8, pp. 1264–1277, Aug. 2014.
- [6] T. Filler, J. Judas, and J. Fridrich, “Minimizing embedding impact in steganography using trellis-coded quantization,” in *Proc. SPIE, Media Forensics and Security*, San Jose, California, USA, Jan. 2010, vol. 7541, p. 754105.
- [7] T. Filler, J. Judas, and J. Fridrich, “Minimizing additive distortion in steganography using syndrome-trellis codes,” *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sept. 2011.
- [8] V. Holub and J. Fridrich, “Designing steganographic distortion using directional filters,” in *Proc. WIFS*, Costa Adeje, Tenerife, Spain, Dec. 2012, pp. 234–239.
- [9] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1, Jan. 2014.
- [10] B. Li, M. Wang, J. Huang, and X. Li, “A new cost function for spatial image steganography,” in *Proc. ICIP*, Paris, France, Oct. 2014, pp. 4206–4210.
- [11] J. Fridrich and J. Kodovský, “Multivariate gaussian model for designing additive distortion for steganography,” in *Proc. ICASSP*, Vancouver, BC, Canada, May 2013, pp. 2949–2953.
- [12] T. Pevný, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” in *Proc. IH*, Calgary, AB, Canada, June 2010, pp. 161–177.
- [13] V. Sedighi, J. Fridrich, and R. Cogranne, “Content-adaptive pentary steganography using the multivariate generalized gaussian cover model,” in *Proc. SPIE, Media Watermarking, Security, and Forensics 2015*, San Francisco, California, USA, Mar. 2015, pp. 9409 – 9409 – 13.
- [14] V. Sedighi, R. Cogranne, and J. Fridrich, “Content-adaptive steganography by minimizing statistical detectability,” *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, Feb. 2016.
- [15] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, June 2012.
- [16] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, “Selection-channel-aware rich model for steganalysis of digital images,” in *Proc. WIFS*, Atlanta, GA, USA, Dec. 2014, pp. 48–53, IEEE.
- [17] B. Li, Z. Li, S. Zhou, S. Tan, and X. Zhang, “New steganalytic features for spatial image steganography based on derivative filters and threshold LBP operator,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 5, pp. 1242–1257, May 2018.
- [18] P. Bas, T. Filler, and T. Pevný, “Break our steganographic system: The ins and outs of organizing BOSS,” in *Proc. IH*, Prague, Czech Republic, May 2011, pp. 59–70.
- [19] J. Kodovský, J. Fridrich, and V. Holub, “Ensemble classifiers for steganalysis of digital media,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444, Apr. 2012.