MULTIVARIATE GAUSSIAN MODEL FOR DESIGNING ADDITIVE DISTORTION FOR STEGANOGRAPHY

Jessica Fridrich and Jan Kodovský

Binghamton University
Department of ECE
Binghamton, NY

ABSTRACT

Currently, the most successful approach to steganography in empirical objects, such as digital media, is to cast the embedding problem as source coding with a fidelity constraint. The sender specifies the costs of changing each cover element and then embeds a given payload by minimizing the total embedding cost. Since efficient practical codes exist that embed near the rate–distortion bound, the remaining task left to the steganographer is the fidelity measure – the choice of the costs. In the past, the costs were obtained either in an ad hoc manner or determined from the effects of embedding in a chosen feature space. In this paper, we adopt a different strategy in which the cover is modeled as a sequence of independent but not necessarily identically distributed quantized Gaussians and the embedding change probabilities are derived to minimize the total KL divergence within the chosen model for a given embedding operation and payload. Despite the simplicity of the adopted model, the resulting stegosystem exhibits security that is comparable to current state-of-the-art methods methods across a wide range of payloads.

Index Terms— Steganography, multivariate Gaussian cover, additive distortion function, syndrome-trellis codes, steganalysis

1. INTRODUCTION AND PRIOR ART

Fundamentally, there exist three types of steganographic systems – steganography by cover synthesis, cover selection, and cover modification [1]. While the first two are important for studying theoretical aspects of steganography, only the third one can be used to embed payloads that are large enough to make the stegosystem practical.

Steganography by cover modification can be approached from several different directions. Model-based approaches start with adopting a cover model that the embedding algorithm is forced to preserve [2, 3, 4]. Although the resulting stegosystem is undetectable within the chosen model, such systems are (sometimes extremely) detectable within alternative representations of the cover source. A more pragmatic approach is to admit that one will never construct a perfectly secure system for empirical objects and design the steganography to minimize a distortion function that is related to statistical detectability. Here, right from the beginning the sender gives up perfect security, and, instead, minimizes the steganographic Fisher information to maximize the size of the secure payload that can be embedded at a fixed level of statistical detectability. This approach has been extraordinarily successful and lead to practical embedding schemes that current best steganalyzers cannot reliably detect even at rather large payloads [5, 6, 7].

The most common distortion function is additive w.r.t. cover elements. The designer starts by assigning costs of changing each cover element (pixel or quantized JPEG DCT coefficient) and then embeds a given payload with the smallest possible distortion. This problem can be formulated as source coding with fidelity constraint [8] for which efficient near-optimal codes exist – the syndrome–trellis codes (STCs) [9]. Freed from having to invent coding schemes for every embedding scheme, the stego designer only needs to specify the pixel costs.

The caveat of this design is, of course, the costs. Ideally, they should be defined to minimize the statistical detectability. However, the relationship between costs and statistical detectability is currently not clear. Intuitively, the costs should be high in well-modelable smooth regions and low in noisy/textured content, where modeling the content becomes difficult. The cost could be parametrized and then optimized w.r.t. a specific cover representation and source (feature space and image database) as in HUGO [5]. Since such adaptive schemes concentrate the embedding modifications into smaller regions, one might need to properly model the interactions between the embedding changes, which inevitably leads to non-additive distortion functions and necessitates more complex methods, such as the Gibbs construc-

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9950-12-1-0124. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of AFOSR or the U.S. Government.

tion [10]. Non-additive distortions could be made additive using the so-called additive approximation or majorized by a bounding distortion [10], allowing again embedding using STCs. In [11], the costs were optimized w.r.t. a specific feature space to minimize the margin between the classes of cover and stego features. The method constructed in this manner was later shown to be extremely detectable in an appropriately extended (but still low-dimensional) feature space [12]. While applying the same idea with a richer representation of covers [13, 14] might be ultimately successful, it is unclear how to scale this approach with increased diversity of the features.

In this paper, we approach the problem of designing the pixel costs in a different manner. In the next three sections, we start with a simple cover model and embedding operation and compute the embedding change probabilities to minimize the KL divergence between the cover and stego objects.[1] The model is a sequence of independent (but not necessarily identically distributed) quantized Gaussians. Similar models are not new and were already used in steganography, e.g., in [15]. The non-stationarity of this model can capture the varying content in images, while the assumption of independence permits an especially simple analytical treatment. We chose the Least Significant Bit Matching (LSBM) as the embedding operation because one can easily derive its steganographic Fisher information [16, 17]. The method of Lagrange multipliers is used to derive the optimal embedding change probabilities for a given payload and image. Curiously, the embedding profile (cost) now depends on the payload. In Section 5.2, we test the security of the embedding algorithm, which we call MG (Multivariate Gaussian), on a database of images with rich cover models and compare the detectability with HUGO [5], which, at the time of writing this paper, was the most secure steganographic algorithm for images represented in the spatial domain. Despite the simplicity of the chosen cover model, the MG algorithm offers better security than HUGO for payloads larger than 0.3 bpp. For smaller payloads, HUGO is slightly more secure. Even though the MG algorithm does not lead to a major improvement over existing state of the art, we believe that the methodology introduced in this paper is significant and further improvement can be expected with more complex cover models.

2. COVER MODEL

Given a uniform scalar quantizer $Q_\triangle$ with quantization step $\triangle$ and range $\mathcal{M} = \{j\triangle | j \in \mathbb{Z}\}$, the cover will be modeled as a sequence of n independent random variables, $X = (X_1, \ldots, X_n)$, which are quantized zero-mean Gaussians $Q_\triangle(N(0, v_i))$ with p.m.f.'s $p^{(i)} = (p_j^{(i)})$, $j \in \mathcal{M}$.[2] In this article, the integers $i \in \{1, \ldots, n\}$ and $j \in \mathbb{Z}$ will be *exclusively* used to index pixels and bins in $\mathcal{M}$, respectively. Thus, below, we refrain from adding the respective ranges of both indices to declutter the text.

The embedding modifies each pixel independently with probability β_i, changing the cover to a sequence of independent random variables (stego object), $Y = (Y_1, \ldots, Y_n)$, with distributions $q^{(i)}(\beta_i) = (q_j^{(i)})$, $j \in \mathcal{M}$. One can say that β_i is the ith change rate.

With increasing β_i, the KL divergence between the cover and stego objects increases. For small change rates, the KL divergence is well-approximated with its leading quadratic term:[3]

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

where $I_i(0)$ is the steganographic Fisher information (FI)

$$I_i(0) = \sum_j \frac{1}{p_j^{(i)}} \left(\frac{\mathrm{d}q_j^{(i)}(\beta_i)}{\mathrm{d}\beta_i}\bigg|_{\beta_i=0} \right)^2. \tag{2}$$

3. EMBEDDING: ADAPTIVE LSBM

In this article, we consider LSBM as the embedding operation. It is used almost solely in all stegosystems designed for digital images in both raster and transfer-domain formats. LSBM changes pixel i by ± 1 with probabilities $\beta_i^+ = \beta_i^- = \beta_i$. Under these assumptions, the stego pixel distribution and its partial derivative become (we drop the pixel index i for better readability):

$$q_j(\beta) = (1 - 2\beta)p_j + \beta(p_{j-1} + p_{j+1}), \tag{3}$$

$$\frac{\partial q_j}{\partial \beta}\bigg|_{\beta=0} = -2p_j + p_{j+1} + p_{j-1}. \tag{4}$$

The FI will be computed in the fine quantization limit. Using

$$F_\triangle(x) \triangleq \int_{x-\triangle/2}^{x+\triangle/2} f_v(t)\mathrm{d}t \tag{5}$$

for Gaussian density $f_v(x)$ with variance v and zero mean, the Mean Value Theorem (MVT) gives for the quantized Gaussian cover

$$p_j = F_\triangle(j\triangle) = \triangle f_v(j'\triangle) \tag{6}$$

for some $j' \in (j-1/2, j+1/2)$. The values $p_{j\pm1} = F_\triangle((j \pm 1)\triangle)$ can be obtained using Taylor expansion of $F_\triangle(x)$ at $x = j\triangle$:

$$p_{j\pm1} = \sum_{l=0}^{\infty} F_\triangle^{(l)}(j\triangle)\frac{(\pm\triangle)^l}{l!}, \tag{7}$$

[1] Note that we are not forcing the stego algorithm to preserve the model but merely to disturb it in the least possible way.

[2] We note that the results derived in this paper hold under the slightly more general cover model $X_i \sim Q_\triangle(N(\mu_i, v_i))$ with an integer μ_i.

[3] In fact, this approximation is valid also for "large" change rates in the fine quantization limit (when $v_i \gg \triangle$).

where $F_{\triangle}^{(l)}$ is the lth derivative of $F_{\triangle}$. After substituting (6) and (7) in (4), simplifying, and using the MVT, for each $l \geq 1$ and x, $F_{\triangle}^{(l)}(x) = f_v^{(l-1)}(x + \triangle/2) - f_v^{(l-1)}(x - \triangle/2) = \triangle f_v^{(l)}(\phi_l)$ for some $\phi_l \in (x - \triangle/2, x + \triangle/2)$:

$$\left.\frac{\partial q_j}{\partial \beta}\right|_{\beta=0} = \triangle^3 f_v''(j\triangle) + \mathcal{O}(\triangle^4). \quad (8)$$

Finally, the Fisher information

$$I(0) = \sum_j \frac{1}{p_j}\left(\left.\frac{\partial q_j}{\partial \beta}\right|_{\beta=0}\right)^2 \approx \sum_j \frac{\triangle^6 \big(f_v''(j\triangle)\big)^2}{\triangle f_v(j'\triangle)} \quad (9)$$

$$\approx \triangle^4 \int_{\mathbb{R}} \frac{\big(f_v''(x)\big)^2}{f_v(x)}\mathrm{d}x = \frac{\triangle^4}{v^2}. \quad (10)$$

Eq. (10) was obtained by approximating the "Riemann sum" in (9) with an integral and evaluating it for Gaussian density f_v.

4. MINIMIZING THE KL DIVERGENCE

In this section, we derive the change rates β_i (and thus the pixel costs) for the payload-limited sender (PLS) that minimizes the KL divergence. The total relative payload that can be embedded in the image is the sum of entropies of p.m.f.'s $\{\beta_i, \beta_i, 1 - 2\beta_i\}$,

$$\alpha n = \sum_{i=1}^{n} h(\beta_i), \quad (11)$$

where $h(x) = -2x \ln x - (1 - 2x)\ln(1 - 2x)$ is expressed in nats. The optimal choice of β_i that minimizes the total KL divergence (1) subject to the payload constraint (11) can be found using the method of Lagrange multipliers. Differentiating the objective function w.r.t. β_i gives:

$$\frac{\partial}{\partial \beta_i}\left(\sum_{k=1}^{n} \frac{1}{2}\beta_k^2 I_k(0) - \frac{1}{\lambda}\left[\sum_{k=1}^{n} h(\beta_k) - \alpha n\right]\right) = 0, \quad (12)$$

$$\beta_i I_i(0) - \frac{2}{\lambda}\ln\frac{1 - 2\beta_i}{\beta_i} = 0, \quad (13)$$

which needs to be solved numerically for each pixel i. Solving (13) for β_i is equivalent to finding x satisfying $\lambda I_i(0)/2 = x\ln(x - 2)$, where $x = \beta_i^{-1} \geq 3$ since $h(x)$ is maximized for $x = 1/3$. To solve this equation quickly for all pixels, the inverse function to $y = x\ln(x - 2)$ was tabulated for $y \leq 10^3$ and an asymptotic iterative solution was implemented for $y > 10^3$. From the requirement that the found β_i be minima, the second derivative of the objective function w.r.t. β_i must be positive, which means that $\lambda > 0$. For the PLS, the Lagrange multiplier λ is determined from the payload constraint (11).

Since the probabilities minimizing an additive distortion function with pixel costs ρ_i satisfy $\beta_i = 1/(1 + \exp(\lambda\rho_i))$ (see, e.g., [9]), the pixel costs corresponding to embedding probabilities β_i are

$$\rho_i = \ln(1/\beta_i - 1). \quad (14)$$

Because the costs are unique up to a multiplicative constant, we normalize them so that $\max_i \rho_i = 1$. By ordering ρ_i from the smallest to the largest, we obtain the so-called cost profile.

5. EXPERIMENTS

5.1. Cover model estimation and embedding

For a given relative payload α and grayscale image $\mathbf{x} = \{x_i\}$, $x_i \in \{0, \ldots, 255\}$, the sender first computes the costs ρ_i using (14). Even though pixel values are not realizations of independent zero-mean Gaussians, the pixels are locally strongly correlated. Assuming that X_i from a small (e.g., 3×3) neighborhood $\mathcal{N}_i$ have the same mean and variance, the variance v_i of X_i can be estimated as

$$v_i = \max\{1, E_{\mathcal{N}_i}[x_i^2] - (E_{\mathcal{N}_i}[x_i])^2\}, \quad (15)$$

where $E_{\mathcal{N}_i}$ is the sample mean over $\mathcal{N}_i$. The maximum with 1 was added for numerical stability.

The actual embedding was simulated at the rate–distortion bound both for the MG algorithm and HUGO, which we included for comparison as the current state-of-the-art algorithm for images in raster format as of November 2012. In practice, the ternary version of STCs [9] could be used to implement the actual embedding algorithm near its payload–distortion bound.[4]

5.2. Steganalysis

To see how the detectability increases with increased payload, steganalysis was carried out using supervised machine learning by building a binary classifier for the class of cover images and stego images embedded with a fixed relative payload. Images were represented using the state-of-the-art spatial rich model (SRM) [13] with $q = 1$ with total dimensionality 12,753. The machine learning was the ensemble [18] run with default settings with the Fisher linear discriminant as the base learner. The cover source was the BOSSbase 1.01 [19] containing 10,000 grayscale 512×512 images originally obtained by seven cameras in raw format (DNG, CR2), demosaicked, and resized/cropped using a script also available online. The detection performance was evaluated in a standard manner using the minimal total error under equal prior probabilities of both hypotheses, $\overline{P}_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, averaged over ten random splits of the database into two halves.[5]

[4] STCs are known to have a small coding loss that diminishes to zero with increasing constraint height.

[5] P_{FA} and P_{MD} are the false-alarm and missed-detection rates.

Fig. 1. Test image (left) and the embedding changes displayed in white when embedding relative payload 0.4 bpp (right).

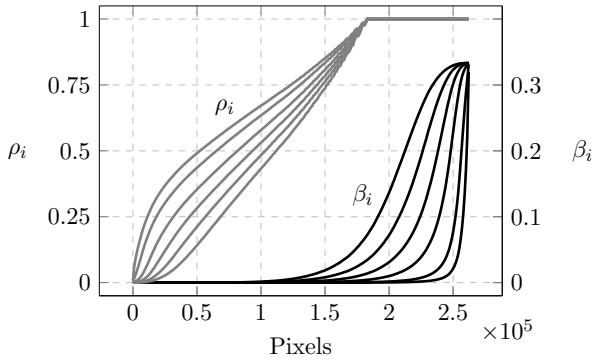

Fig. 2. Cost profiles ρ_i for image in Fig. 1 for six relative payloads 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 bpp (top curve corresponds to 0.05 bpp) and sorted embedding change probabilities β_i (top curve corresponds to 0.5 bpp).

In Fig. 1, we show one test image from BOSSbase and the embedding changes in white for payload 0.4 bpp (bits per pixel). Like HUGO, the MG algorithm is content-adaptive, concentrating the embedding changes in edges and textures. Working out the optimal embedding change probabilities for different payloads α for this image (Fig. 2), we discover that, in contrast with embedding schemes that fix the costs, the cost profile now *depends on the payload*. This is because we minimize the KL divergence in a multivariate Gaussian model rather than an embedding cost fixed for each pixel in the beginning.

Fig. 3 shows the average detection error $\overline{P}_{\mathrm{E}}$ as a function of the relative payload for the MG algorithm and for HUGO. HUGO was run with its parameters $\gamma = \sigma = 1$ and threshold $T = 255$. Both algorithms perform similarly for payloads up to 0.3 bpp. Then, the adaptivity of MG seems better than that of HUGO. The difference in detectability at $\alpha = 1$ is caused by the fact that MG uses ternary embedding and thus still preserves some adaptability while HUGO at this payload loses its adaptive character.

6. CONCLUSIONS

Constructing steganographic schemes using additive distortion functions is a modern trend in steganography for digital media. Since the coding part of the problem has been resolved, the most crucial element is the design of the individual pixel costs. In general, finding a relationship between the costs and statistical detectability is a very hard problem and one that will probably remain open for many years to come because of the complexity of digital media. In this paper, we adopt a rather simple model – a multivariate quantized Gaussian distribution and derive the pixel costs to minimize the KL divergence when embedding using least-significant bit matching. In contrast to schemes built by fixing the pixel costs, the distortion profile for this embedding algorithm depends on the payload. Despite the simplicity of the cover model, the MG algorithm exhibits security comparable to the current state-of-the-art algorithm HUGO. This provides hope that this approach to minimum-distortion steganography has a promise and might provide superior performance with more complex models.

With more complex models, the most problematic issue seems to be estimation of the local parameters, such as the covariance matrix for a joint Gaussian model or the transition probability matrix for a Markov model. Estimating these objects will inevitably run into the difficulty of having to trade off between estimator variance and bias.

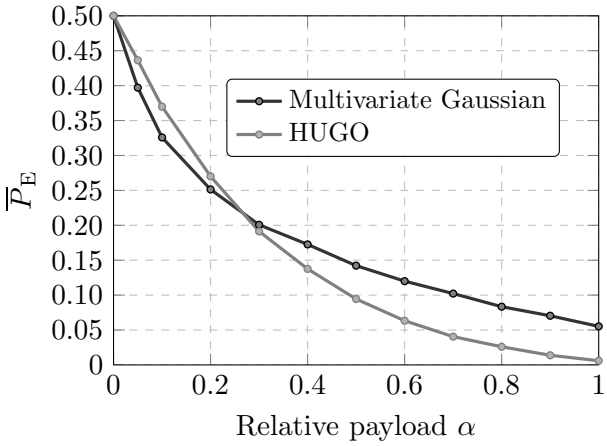

Fig. 3. Average detection error $\overline{P}_{\mathrm{E}}$ as a function of relative payload for MG and HUGO.

7. REFERENCES

[1] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*, Cambridge University Press, 2009.

[2] N. Provos, "Defending against statistical steganalysis,"

in *10th USENIX Security Symposium*, Washington, DC, August 13–17, 2001, pp. 323–335.

[3] P. Sallee, "Model-based methods for steganography and steganalysis," *International Journal of Image Graphics*, vol. 5, no. 1, pp. 167–190, 2005.

[4] E. Franz, "Embedding considering dependencies between pixels," in *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, Eds., San Jose, CA, January 27–31, 2008, vol. 6819, pp. D 1–12.

[5] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Information Hiding, 12th International Conference*, R. Böhme and R. Safavi-Naini, Eds., Calgary, Canada, June 28–30, 2010, vol. 6387 of Lecture Notes in Computer Science, pp. 161–177, Springer-Verlag, New York.

[6] C. Wang and J. Ni, "An efficient JPEG steganographic scheme based on the block–entropy of DCT coefficents," in *Proc. of IEEE ICASSP*, Kyoto, Japan, March 25–30, 2012.

[7] V. Sachnev, H. J. Kim, and R. Zhang, "Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding," in *Proceedings of the 11th ACM Multimedia & Security Workshop*, J. Dittmann, S. Craver, and J. Fridrich, Eds., Princeton, NJ, September 7–8, 2009, pp. 131–140.

[8] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, pp. 142–163, 1959.

[9] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, September 2011.

[10] T. Filler and J. Fridrich, "Gibbs construction in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 4, pp. 705–720, 2010.

[11] T. Filler and J. Fridrich, "Design of adaptive steganographic schemes for digital images," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics of Multimedia XIII*, A. Alattar, N. D. Memon, E. J. Delp, and J. Dittmann, Eds., San Francisco, CA, January 23–26, 2011, vol. 7880, pp. OF 1–14.

[12] J. Kodovský, J. Fridrich, and V. Holub, "On dangers of overtraining steganography to incomplete cover model," in *Proceedings of the 13th ACM Multimedia & Security Workshop*, J. Dittmann, S. Craver, and C. Heitzenrater, Eds., Niagara Falls, NY, September 29–30, 2011, pp. 69–76.

[13] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, June 2011.

[14] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics of Multimedia XIV*, A. Alattar, N. D. Memon, and E. J. Delp, Eds., San Francisco, CA, January 23–26, 2012, vol. 8303, pp. 0A–1–0A–13.

[15] R. Cogranne, C. Zitzmann, L. Fillantre, F. Retraint, I. Nikiforov, and P. Cornu, "A cover image model for reliable steganalysis," in *Information Hiding, 13th International Conference*, T. Filler, T. Pevný, A. Ker, and S. Craver, Eds., Prague, Czech Republic, May 18–20, 2011, vol. 6958 of Lecture Notes in Computer Science, pp. 178–192.

[16] T. Filler and J. Fridrich, "Fisher information determines capacity of ϵ-secure steganography," in *Information Hiding, 11th International Conference*, S. Katzenbeisser and A.-R. Sadeghi, Eds., Darmstadt, Germany, June 7–10, 2009, vol. 5806 of Lecture Notes in Computer Science, pp. 31–47, Springer-Verlag, New York.

[17] A. D. Ker, "Estimating steganographic fisher information in real images," in *Information Hiding, 11th International Conference*, S. Katzenbeisser and A.-R. Sadeghi, Eds., Darmstadt, Germany, June 7–10, 2009, vol. 5806 of Lecture Notes in Computer Science, pp. 73–88, Springer-Verlag, New York.

[18] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[19] T. Filler, T. Pevný, and P. Bas, "BOSS (Break Our Steganography System)," `http://www.agents.cz/boss`, July 2010.