JPEG STEGANOGRAPHY WITH SIDE INFORMATION FROM THE PROCESSING PIPELINE

Quentin Giboulot, Rémi Cogranne, Patrick Bas

JPEG STEGANOGRAPHY WITH SIDE INFORMATION FROM THE PROCESSING PIPELINE

Quentin Giboulot[+]*, Rémi Cogranne*[+] *and Patrick Bas*[†]

[+] ICD - M2S - ROSAS - FRE 2019 CNRS - Troyes University of Technology, Troyes, France
[†] CNRS, CRIStAL Lab., École Centrale de Lille, University of Lille, France

ABSTRACT

The current art in schemes using deflection criterion such as MiPOD for JPEG steganography is either under-performing or on par with distortion-based schemes. We link this lack of performance to a poor estimation of the variance of the model of the noise on the cover image. In this paper, we propose a method to better estimate the variances of DCT coefficients by taking into account the dependencies between pixels that come from the development pipeline. Using this estimate, we are able to extend statistically-informed steganographic schemesto the JPEG domain while significantly outperforming the current state-of-the-art JPEG steganography. An extension of Gaussian Embedding in the JPEG domain using quantization error as side-information is also formulated and shown to attain state-of-the-art performances.

Index Terms— Steganography, JPEG images, Statistical model, Side-information, Covariance estimation.

1. INTRODUCTION

The current trend in the design of steganographic schemes for JPEG images relies heavily on the use of so-called distortion functions which associate a cost of modification to each DCT coefficient of the image. Such schemes work by minimizing the sum of costs under the constraint of embedding a given payload (the Distortion Limited Sender). This framework is well exemplified by the current state of the art in JPEG steganography, namely J-UNIWARD [1], which assigns costs based on directional noise residuals estimated with a wavelet filter bank. The costs are heuristically linked to local content complexity and the rational is thus to assign high costs to smooth areas and low costs to areas which are "unpredictable" in every directions. The limitation of such an approach, however, is the lack of statistical interpretability since the distortion function is always defined heuristically.

This problem has been addressed in the spatial domain by the MiPOD steganographic scheme [2]. By modeling pixels of an image with independent Gaussian random variables, MiPOD directly minimizes the power of the most powerful detector under this model. This gives a direct interpretation of the cost of modifying each pixel. Despite its state-of-the-art performances in the spatial domain, the direct generalization of this framework to the JPEG domain, dubbed J-MiPOD, had little success competing with the current art [3].

Currently, the only competitive steganographic scheme which relies on statistically defined costs in the JPEG domain uses so called side-information. Side-information can refer to any knowledge to which the steganographer has access related to a given cover which can be used to improve the security of a steganographic scheme (e.g: another image of the same scene, the RAW image, knowledge of rounding errors, ...). In the case of JPEG images, a higher-resolution version of the given cover (a.k.a. the precover) consists of the non-rounded DCT coeffcients of the image. The state of the art for distortion-based scheme, SI-UNIWARD [1], significantly improves the security of J-UNIWARD by heuristically modulating its costs by $|0.5 - e_i|$ ($e_i \in]-0.5, 0.5]$ being the rounding error of the i-th DCT coefficient), embedding preferentially in coefficients which are close to a bin boundary. On the other hand, Model based SI-MiPOD [3], referred to as MB-MiPOD in this paper, formally derives a modulation factor of $|0.5 - e_i|^2$ by minimizing the KL-divergence between the cover and the stego image conditioned on the realization of the precover.

Two frameworks for statistically informed schemes currently exists: the aforementioned MiPOD and the recently proposed Gaussian Embedding [4]. They both suppose an independent Gaussian model of the noise residuals of pixels and as such, their performance relies on a proper esti-

This work has been funded in part by the French National Research Agency (ANR-18-ASTR-0009-02) ALASKA project https://alaska.utt.fr and by the French ANR DEFALS program (ANR-16-DEFA-0003) .

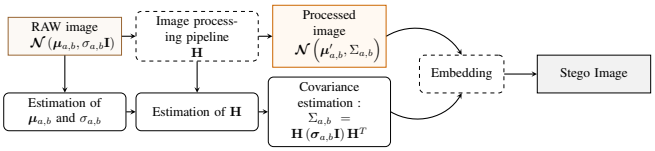

Fig. 1. Principle of steganography with side information coming from the processing pipeline: the estimation of the covariance matrix of the sensor-noise in the DCT domain enables to compute reliable variance estimates and to derive meaningful deflection coefficients, used to derive statistically founded costs for JPEG images.

mation of the variances of those residuals. However, these schemes estimate the variance in the spatial domain and, because of the independence assumption, the estimation of the variance in the JPEG domain from those estimates (as it was done for J-MiPOD) is extremely crude as it does not take into account the dependencies introduced very early in the processing pipeline during demosaicking. Consequently, better estimation of the variances in the JPEG domain ought to rely on the estimation of covariances of blocks of dependent pixels in the spatial domain.

In this paper, we show how to use the knowledge of the processing pipeline to obtain precise estimates of those covariances to make these two frameworks competitive in the JPEG domain. At the same time we propose a new side informed extension of the Gaussian embedding scheme into the JPEG domain. We show that our estimation method allows us to significantly outperform the current state of the art in JPEG steganography.

2. STATISTICAL MODEL OF THE SENSOR NOISE IN THE JPEG DOMAIN

In this section we describe the cover model that will be used in the rest of the paper as well as the method to estimate its parameters. We assume that the steganographer has both access to the original RAW image of a given cover, and to the knowledge of the processing pipeline that generates the cover. We also make the assumption that the operations of the processing pipeline can be approximated by a linear operator on blocks of pixels of the image. This last assumption might seem to reduce the range of potential image usable with our method, yet we show in Section 4 that it is robust to nonlinear processing pipelines.

Following [5, 6], we model the photo-site values as independent random variables following the heteroscedastic noise model:

$$X_{i,j}^{RAW} \sim \mathcal{N}\left(\mu_{i,j}, \sigma_{i,j}\right) \ ; \ \sigma_{i,j} = c_1\mu_{i,j} + c_2, \qquad (1)$$

where $\mu_{i,j}$ is the value that would be registered by the i,j-th photo-site if no sensor noise was present. In what follows, it will be easier to phrase the model in terms of $(8M + k) \times (8M + k)$ block of photo-sites following a multivariate Gaussian with diagonal covariance:

$$\mathbf{X}_{a,b}^{RAW} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}_{a,b}, \boldsymbol{\sigma}_{a,b}\mathbf{I}\right), \qquad (2)$$

where $\mathbf{X}_{a,b}^{RAW}$ corresponds to the (a,b)th 8×8 block of RAW image together with its $M-1$ neighboring 8×8 blocks and $16Mk + k^2$ coefficients at the margins.

We model the processing pipeline up to the DCT transform (i.e. demosaicking, white balancing, denoising, etc…) as a linear operator represented as a matrix $\mathbf{H}$ of dimension $(8N)^2 \times (8M + k)^2$. Since the processing pipeline introduces dependencies between pixels, we can model $8N \times 8N$ blocks of dependent DCT coefficients of the developed image as multivariate Gaussian random variables [7]:

$$\mathbf{X}_{\mathbf{a,b}}^{\mathbf{dev}} \sim \boldsymbol{\mathcal{N}}\left(\boldsymbol{\mu}'_{a,b}, \Sigma_{a,b}\right) \ ; \ \Sigma_{a,b} = \mathbf{H}\left(\boldsymbol{\sigma}_{a,b}\mathbf{I}\right)\mathbf{H}^T. \qquad (3)$$

Note that the constants M,N and k must be chosen depending on the range of dependencies introduced by the processing pipeline. For example, in the case of a purely linear processing pipeline consisting of bilinear demosaicking, RGB to greyscale conversion and DCT transform, one must choose $M = N = 3$ and $k = 2$ to take all the dependencies for one block into account (see [8] for details and explanations regarding the dependencies introduced by bilinear demosaicking followed by JPEG compression).

Since the embedding schemes presented in Section 3 suppose independent DCT coefficients, we eventually neglect those dependencies: only the diagonal of the covariance matrix is kept for each block and the DCT coefficients are then considered independent. We would like to emphasize to the reader that despite reintroducing the independence assumption, the estimation of the diagonal terms takes into account the dependencies between pixels, which is not the case for J-MiPOD.

In the rest of this section, we explain how to estimate $\mu_{i,j}$ and $\Sigma_{a,b}$. The method is summarized in Figure 1.

2.1. RAW model estimation and covariance estimation

To be able to estimate the covariances $\Sigma_{a,b}$, we need to estimate the variances associated to each photo-site which are themselves function of the true value of each photo-site $\mu_{i,j}$. To that end, we denoise the RAW image using the method in [9] based on the inverse Anscombe transform and the BM3D algorithm [10]. The heteroscedastic model parameters c_1 and c_2 are then estimated using the method detailed in [5, 6]. In practice the RAW model has to be estimated only once for each camera and each ISO among images in a given dataset. Once the parameters of the RAW model are estimated, the covariance matrix of each block $\Sigma_{a,b}$ can be estimated as:

$$\Sigma_{a,b} = \mathbf{H}\left(\boldsymbol{\sigma}_{a,b}\mathbf{I}\right)\mathbf{H}^T. \qquad (4)$$

Depending on the processing pipeline, $\mathbf{H}$ can be computed analytically such as in [8] for bilinear demosaicking. In our case, since the processing pipeline is assumed linear, we can blindly estimate $\mathbf{H}$ using a linear regression between the photo-site blocks and the developed blocks, that is, solving for $\mathbf{H}$:

$$\mathbf{X}_{a,b}^{dev} = \mathbf{H}\mathbf{X}_{a,b}^{RAW} + \mathbf{C}. \tag{5}$$

where $\mathbf{C}$ is the constant (a.k.a intercept) term.

2.2. Quantization

The covariance matrix has to take into account the quantization of the DCT coefficient according to a quantization table specific to each JPEG quality factorAs the covariance matrix is first estimated on non-rounded DCT coefficient (just before rounding, that is when each coefficient has already been divided by its corresponding entry in the quantization matrix), the quantization step is equivalent to a uniform quantizer with quantization step 1. When the diagonal variances are "high-enough", this is equivalent to adding $\frac{1}{12}$ on the diagonal terms of the covariance matrix [11] while "small" variances as well as their associated covariances should be set to zero. We define "small" variances heuristically as variances σ^2 associated to a given coefficient where the first bin boundary of the quantized value is at a distance of at least 3σ from the mean value of the coefficient (that is where the probability that its realization lies in the average bin is at least 99.7%). Such variances, as well as all their covariances are thus set to 0.

3. EMBEDDING

Once the parameters of the cover model are estimated, embedding using statistically defined costs becomes possible. We propose here two approaches. The first one, which we will refer to as $\Sigma-\mathrm{MiPOD}$, uses only the estimated cover model while the second, referred as Σ-SI-Gaussian, also uses the knowledge of the rounding errors in the DCT domain as side-information. The first approach is a straightforward expansion of MiPOD to the JPEG domain.The second approach is similar to the recently proposed Gaussian Embedding scheme [4], but here the embedding is performed in the JPEG domain and the development pipeline is taken into account. Due to space constraints, we only present the optimization problem to be solved for each embedding scheme and refer the reader to the original publications for the implementation details such as the optimization solving or the use of the Syndrome-Trellis Codes [12] for efficient implementation.

3.1. Embedding in the quantized DCT domain

Following Section 2, let the cover be modeled as a sequence of n independent random variables with the following pmf:

$$p_{\sigma_n}(k) = \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left(-\frac{k^2}{2\sigma_n^2}\right). \tag{6}$$

The embedding is carried out independently on each DCT coefficient, changing their value by at most ± 1 with probability β_n.

It can be shown that minimizing the power of the most powerful detector of this embedding scheme is equivalent to minimizing the so called deflection coefficient which is well approximated by (see [2] for details):

$$\varrho = \sum_{i=0}^{n} \beta_i^2 \sigma_i^{-4}. \tag{7}$$

This quantity must be minimized under the constraint of a given payload size M which leads to following optimization problem:

$$\begin{cases} \min\limits_{\beta_i} \varrho & = \sum\limits_{i=0}^{n} \beta_i^2 \sigma_i^{-4} \\ M & = \sum\limits_{i=0}^{n} \mathrm{H}_3(\beta_i) \end{cases} \tag{8}$$

with $\mathrm{H}_3(x) = -2x\log(x) - (1-2x)\log(1-2x)$.

In the JPEG domain, we will call this embeddings scheme $\Sigma-\mathrm{MiPOD}$ when using our method for estimating the variance and J-MiPOD for the classic version presented in [3].

3.2. Embedding with Side-Information from the pre-Cover

To use the side-information coming from the quantization error, we propose to recast the Gaussian embedding framework in the continuous domain. Let the pre-cover in the unquantized DCT domain be modeled as a sequence of n independent random variables $X = (X_1, \ldots, X_n)$ with:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^c). \tag{9}$$

The embedding is carried out independently on each DCT coefficient by adding to it a realization of a zero-mean Gaussian $\mathcal{N}(0, \epsilon_i)$ resulting in a "pre-stego" which is a sequence of n independent random variables $Z = (Z_1, \ldots, Z_n)$ with

$$Z_i \sim \mathcal{N}\left(\mu_i, \sigma_i^{s\,2}\right) \;\; ; \;\; \sigma_i^s = \sqrt{\left(\sigma_i^c\right)^2 + \epsilon_i^2}. \tag{10}$$

We want to minimize the KL-divergence between the pre-cover and the pre-stego under the constraint of embedding a given payload size M. Since the final stego image must eventually be quantized, the constraint has to be expressed in the

quantized domain. This leads to the following optimization problem:

$$\begin{cases} \min\limits_{\epsilon_i} D_{KL}(P||Z) & = \sum\limits_{i=0}^{n} \ln(\frac{\sigma_i^s}{\sigma_i^c}) + \frac{(\sigma_i^c)^2}{2(\sigma_i^s)^2} - 0.5 \\ M & = \sum\limits_{i=0}^{n} \sum\limits_{k\in\mathbb{Z}} \beta_i^k \log\left(\beta_i^k\right) \end{cases} \tag{11}$$

with $\beta_i^k = \phi(\frac{k-e_i}{\epsilon_i}+0.5) - \phi(\frac{k-e_i}{\epsilon_i}-0.5)$ being the probability of modifying the i-th coefficient by $+k$, ϕ being the cumulative distribution function the standard normal distribution and $e_i = x_i - [x_i]$ being the rounding error of i-th DCT coefficient. In practice the alphabet size of the embedding scheme must be finite, k is thus constrained to finite range and the β_i^k renormalized accordingly. For a fair comparison with other embedding schemes, we set $k \in \{-1, 0, 1\}$ for the rest of this paper. We will call this embedding scheme Σ-SI-Gaussian.

4. RESULTS

To correctly study the performance of our approach, it is necessary to separate the effect of the acquisition parameters, in particular the camera and ISO which determine the noise model of each image, and the effect of the processing pipeline. We thus chose to work on datasets with one camera sensor, and constant ISO sensitivity and constant processing pipeline. To that end we used two different RAW databases. The first, dubbed E1Base and used for Natural Steganography [13], is composed of 200 RAW images taken with a E1 Camera at ISO100 while the second, dubbed CanonBase is composed of 119 RAW images taken with a Canon EOS 500D at ISO1600. Those images were then developed using the *rawpy* library and either cropped to 256×256 or cropped to 512×512 then down-sampled to 256×256 without overlap using the *Pillow* python library. Bases of 5000 JPEG images were produced with a fixed sensor, and constant ISO and processing pipeline, which consists only of demosaicking, RGB to grayscale conversion, downscaling in the form of cropping or down-sampling and JPEG compression.

The $\mathbf{H}$ matrix is estimated once for each camera and each processing pipeline using simple least square regression. To that end, we use a synthetic constant RAW image to which sensor noise is added. This image is then processed using the relevant processing pipeline for each datasets. The RAW and developed image are then reshaped as arrays of 10×10 and 8×8 blocks respectively. We eventually compute $\mathbf{H}$ using Eq (5). This implies that the covariance matrix of each block was estimated without using neighboring blocks. Even though the estimation for one block should theoretically be carried out with all its neighboring blocks as shown in the Natural Steganography approach [8], extensive experiments with the E1Base showed no observable gain in security when using those neighboring blocks for the estimation.

Embedding is simulated using our approaches, Σ-SI-Gaussian and $\Sigma-\mathrm{MiPOD}$ as well as J-UNIWARD [1], SI-UNIWARD [1] and MB-MiPOD [3] using our method for estimating the variances. J-MiPOD and MB-MiPOD using J-MiPOD's estimated variances were also tested but the resulting P_E was always worse than J-UNIWARD and SI-UNIWARD respectively by more than 10% on average. Consequently we do not present those results for readability's sake. Steganalysis was carried out using the DCTR [14] feature set and the Low Complexity Linear Classifier [15] with 2500 images in the training and testing sets using 5-folds cross-validation for finding the optimal regularization parameter. The empirical security of the schemes are evaluated using the minimal total probability of error under equal priors: $P_E = \min_{P_{FA}} \left(P_{FA} + P_{MD}\right)/2$

The results are summarized in Figure 2 to 4, they confirm that using our method to estimate the variances consistently improves the security of the tested schemes independently of the processing pipelines. For example, P_E when using Σ-SI-Gaussian is, on average 1.5 times higher on the different E1Bases, up to 2 times higher on average on CanonBase QF95 when compared to SI-UNIWARD. This is true even for pipelines which use non-linear operations, such as the VNG or DCB demosaicking algorithms, demonstrating the robustness of our approach despite the linearity assumption. Despite using our method for estimating the variances, MB-MiPOD is still less secure than Σ-SI-Gaussian yet more secure than SI-UNIWARD. A similar gain in performance can be observed with $\Sigma-\mathrm{MiPOD}$ when compared to J-UNIWARD, though it should be noted that the comparison is not completely fair as J-UNIWARD does not use any side-information while $\Sigma-\mathrm{MiPOD}$ uses it in the form of the knowledge of the sensor noise parameters and of the development pipeline. Notwithstanding this caveat, it corroborates our claim that precise estimation of the variances in the JPEG domain would allow significant gain in security for steganographic schemes especially when compared to J-MiPOD's methodology.

5. CONCLUSION AND PERSPECTIVES

In this paper, we presented a new method to estimate the variances of the DCT coefficients using the knowledge of the processing pipeline. We also proposed an extension of the Gaussian Embedding scheme in the JPEG domain that also uses the quantization error as side-information. Using this approach, we show a significant increase of performance with current state-of-art schemes, side-informed or not. As future works, we will study how to remove the independence assumption in the cover model in order to fully use the covariance matrix of each block. Another line of work will pay attention to making the method more practical by allowing the estimation of the covariance matrix without the knowledge of the RAW image.

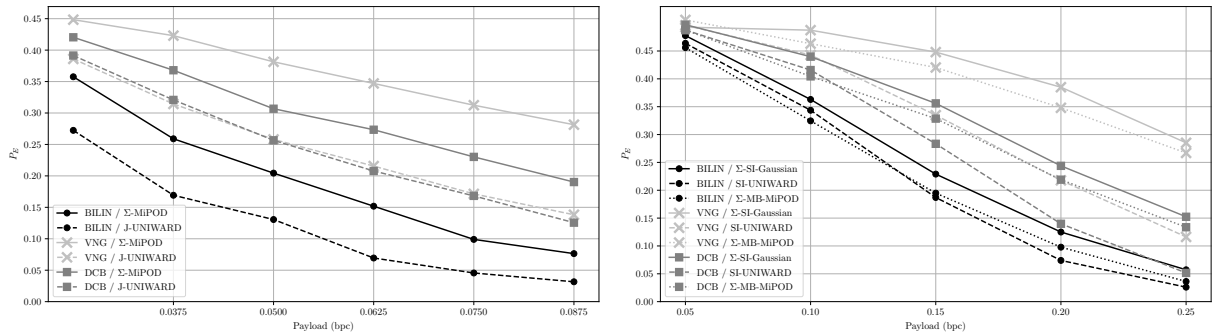

Fig. 2. P_E as a function of payload in bits per coefficients (bpc) for different demosaicking algorithms with cropped images on E1Base. Schemes using the knowledge of rounding errors are on the right while those which do not are on the left. All images are JPEG at QF100.

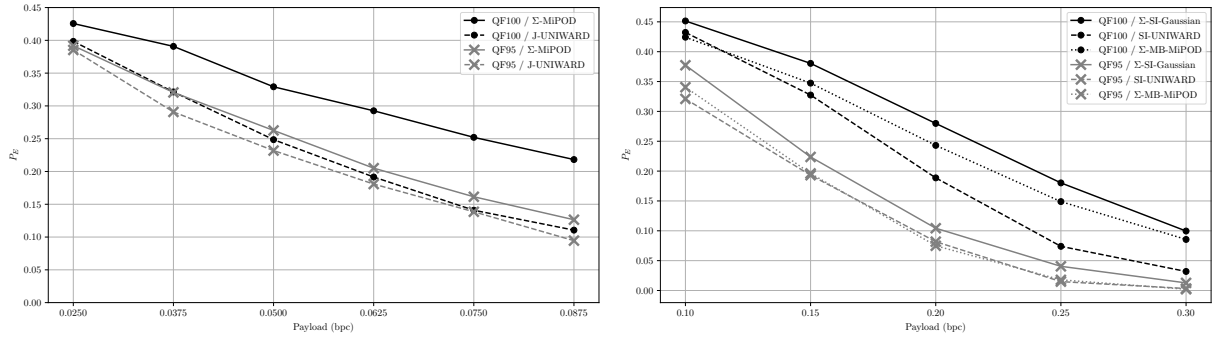

Fig. 3. P_E as a function of payload in bpc for different JPEG quality factors (QF) with cropped images on CanonBase. Schemes using the knowledge of rounding errors are on the right while those which do not are on the left.

6. REFERENCES

[1] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, jan 2014.

[2] Vahid Sedighi, Remi Cogranne, and Jessica Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, feb 2016.

[3] Tomáš Denemark and Jessica Fridrich, "Model based steganography with precover," *Electronic Imaging*, vol. 2017, no. 7, pp. 56–66, jan 2017.

[4] Mehdi Sharifzadeh, Mohammed Aloraini, and Dan Schonfeld, "Quantized gaussian embedding steganography," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. may 2019, IEEE.

[5] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, oct 2008.

[6] Thanh Hai Thai, Remi Cogranne, and Florent Retraint, "Camera model identification based on the heteroscedastic noise model," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250–263, jan 2014.

[7] Thanh Hai Thai, Remi Cogranne, and Florent Retraint, "Statistical model of quantized DCT coefficients: Application in the steganalysis of jsteg algorithm," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 1980–1993, may 2014.

[8] Théo Taburet, Patrick Bas, Jessica Fridrich, and Wadih Sawaya, "Computing dependencies between DCT coefficients for natural steganography in JPEG domain," in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security - IHMMSec2019*. 2019, ACM Press.

[9] Alessandro Foi, "Practical denoising of clipped or overexposed noisy images," in *2008 16th European Signal Processing Conference*. IEEE, 2008, pp. 1–5.

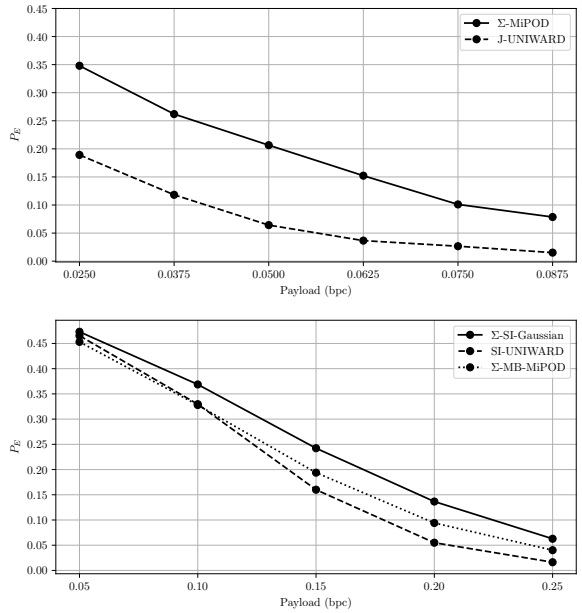

Fig. 4. P_E as a function of payload in bpc for downscaled images on CanonBase with a resizing factor of 2 using the 'LINEAR' kernel of *Pillow*. Schemes using the knowledge of rounding errors are down while those which do not are up.All images are JPEG at QF100.

[10] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, aug 2007.

[11] István Kollár Bernard Widrow, *Quantization Noise*, Cambridge University Press, 2015.

[12] 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, sep 2011.

[13] Patrick Bas, "Steganography via Cover-Source Switching," 2016, IEEE Workshop on Information Forensics and Security (WIFS).

[14] Vojtech Holub and Jessica Fridrich, "Low-complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, feb 2015.

[15] Remi Cogranne, Vahid Sedighi, Jessica Fridrich, and Tomas Pevny, "Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. nov 2015, IEEE.