Steganalysis of LSB Replacement Using Parity-Aware Features

Jessica Fridrich and Jan Kodovský

Department of ECE, Binghamton University, NY, USA
{fridrich,jan.kodovsky}@binghamton.edu

Abstract. Detection of LSB replacement in digital images has received quite a bit of attention in the past ten years. In particular, structural detectors together with variants of Weighted Stego-image (WS) analysis have materialized as the most accurate. In this paper, we show that further surprisingly significant improvement is possible with machine–learning based detectors utilizing co-occurrences of neighboring noise residuals as features. Such features can leverage dependencies among adjacent residual samples in contrast to the WS detector, which implicitly assumes that the residuals are mutually independent. Further improvement is achieved by adapting the features for detection of LSB replacement by making them aware of pixel parity. To this end, we introduce two key novel concepts – calibration by parity and parity-aware residuals. It is shown that, at least for a known cover source when a binary classifier can be built, its accuracy is markedly better in comparison with the best structural and WS detectors in both uncompressed images and in decompressed JPEGs. This improvement is especially significant for very small change rates. A simple feature selection algorithm is used to obtain interesting insight that reveals potentially novel directions in structural steganalysis.

1 Introduction

Least Significant Bit (LSB) replacement, also colloquially called LSB embedding, is arguably the oldest data hiding method. According to the CEO of WetStone Technologies, Inc., as of December 1, 2011 in their depository containing 836 data hiding products, 582 (70%) of them hide messages using LSB embedding. To the same day, the IEEE Xplore database registered 182 conference and 22 journal articles on LSB embedding, which further underlines the enormous popularity of this topic among researchers.

The first accurate detector of LSB replacement was the heuristic RS analysis [10] published in 2001, serendipitously discovered during research on reversible watermarking. The simplest case of RS analysis, the Sample Pairs (SP) analysis, was analyzed and reformulated by Dumitrescu et al. [5] into a framework amenable to further generalization and great improvement [6,4]. The least-squares version of SP by Lu *et al.* [24] later inspired further significant development mostly due to Ker, who derived the detectors from parity symmetries of natural images,

M. Kirchner and D. Ghosal (Eds.): IH 2012, LNCS 7692, pp. 31–45, 2013.

extended the framework to triples [14] and quadruples of pixels [15], and provided further valuable insight [17,16,18].

In 2004, a different kind of LSB detector was introduced [9] that was later dubbed Weighted Stego-image (WS) analysis and further improved in [19] by introducing moderated weights, a better pixel predictor, and a simpler yet more accurate bias correction. The WS detector differed from the structural detectors in that it did not utilize trace sets but instead incorporated the parity through a pixel predictor. The improved version of the WS detector was shown to outperform all other structural attacks in raw, never compressed images, while the triples analysis was identified as the most accurate for decompressed JPEGs. An unweighted version of WS equipped with a recompression predictor was shown to be very effective in decompressed JPEGs provided the quantization table can be estimated [2].

Recently, the WS detector was rederived [26] using invariant hypothesis testing by adopting a parametric model for the cover. An Asymptotically Universally Most Powerful (AUMP) test that seems to coincide with a generalized likelihood ratio was derived in [7]. This detector is a variant of the WS analysis with weights that give it Constant False Alarm Rate (CFAR) property, which allows threshold setting independent of the image source. Finally, we point out that with the exception of [3,7], all LSB replacement detectors mentioned above are quantitative in the sense that the detection statistic is an estimate of the change rate.[1]

Steganalysis of embedding operations other than LSB flipping went in a different direction due to the fact that parity symmetries are no longer useful even for rather trivial modifications of LSB embedding, such as LSB matching. For such embedding operations, the most accurate detectors today are built as classifiers using features obtained as sampled joint distributions (co-occurrence matrices) among neighboring elements of noise residuals [12,11,27,25,13]. These detectors perform equally well for both LSB replacement and LSB matching because features formed from noise residuals are generally blind to pixels' parity.

In contrast to modern steganalysis features (briefly outlined in Section 2), the WS method, which also works with noise residuals, makes an implicit assumption that adjacent residual samples are independent (Section 3). This suggests a potential space for improvement, which we confirm in Section 4 with a simple four-dimensional co-occurrence matrix obtained from the same noise residual that is typically used with WS analysis. With the help of feature selection, improvement over the state of the art (triples analysis) is achieved with as few as three co-occurrence bins for decompressed JPEGs. Besides better utilization of spatial dependencies through co-occurrences, we introduce calibration by parity and parity-aware residuals as two general methods to make features aware of pixel parity to further improve their sensitivity to LSB replacement. By scaling up the feature space complexity using rich models, the best results of this paper are reported in Section 5. The paper is summarized in Section 6.

[1] Since the relationship between the relative payload and change rate depends on the syndrome coding method employed (see, e.g., Chapter 8 in [8]), everywhere in this paper we strictly speak of change-rate estimators.

1.1 Notation

We use boldface symbols for vectors and capital-case boldface symbols for matrices or higher-dimensional arrays. The symbols $\mathbf{X} = (x_{ij}) \in \mathcal{X} = \mathcal{I}^{n_1 \times n_2}$ and $\mathbf{Y} = (y_{ij}) \in \mathcal{X}$, $\mathcal{I} = \{0, \dots, 255\}$, will always represent pixel values of 8-bit grayscale cover and stego images with $n = n_1 n_2$ pixels; $\mathbf{X}^{\mathrm{T}}$ denotes the transpose. We use $\mathbb{R}$ and $\mathbb{Z}$ for the set of real numbers and integers. The operation of rounding $x \in \mathbb{R}$ to an integer is $\mathrm{round}(x)$. Given $T > 0$, $\mathrm{trunc}_T(x) = x$ when $x \in [-T, T]$, and $\mathrm{trunc}_T(x) = T\mathrm{sign}(x)$ otherwise. We also define for $x \in \mathbb{Z}$, $\mathrm{LSB}(x) = \mathrm{mod}(x, 2)$, $\bar{x} = x + 1 - 2\mathrm{LSB}(x)$, which is x with its LSB "flipped." The symbol β stands for the change rate defined as the ratio between the number of embedding changes and the number of pixels. We reserve $\Pr(\mathrm{E})$ for the probability of event E.

1.2 Setup of All Experiments

All experiments in this paper are carried out on BOSSbase ver. 0.92 [1] and its JPEG compressed versions obtained using the Matlab `imwrite` command. The original database contains $9,074$ 512×512 images acquired by seven digital cameras in the RAW format (CR2 or DNG) and subsequently processed by resizing and cropping to the size of 512×512 pixels.

The classifiers we use are all instances of the ensemble proposed in [22,21] and available from `http://dde.binghamton.edu/download/ensemble`. It employs Fisher linear discriminants as base learners trained on random subspaces of the feature space. The out-of-bag estimate of the testing error on bootstrap samples of the training set is used to automatically determine the random subspace dimensionality and the number of base learners as described in [22]. The final classifier decision is obtained by fusing the decisions of its base learners. We train a separate classifier for each image source and payload.

The detection accuracy is evaluated in a standard fashion using the minimal total detection error under equal priors computed from the ROC from the testing set:

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{P_{\mathrm{FA}} + P_{\mathrm{MD}}(P_{\mathrm{FA}})}{2}, \tag{1}$$

where P_{FA} is the false alarm rate and P_{MD} is the missed detection rate. What is reported in all graphs and tables is the average value of this error, $\bar{P}_{\mathrm{E}}$, over ten random divisions of the database into equally-sized training and testing sets. The spread of the error over the database splits also includes the effects of randomness in the ensemble construction (e.g., formation of random subspaces and bootstrap samples). We measure this spread using Mean Absolute Deviation (MAD) defined as the mean of $|P_{\mathrm{E}}(i) - \bar{P}_{\mathrm{E}}|$, where $P_{\mathrm{E}}(i)$ is the testing error on the ith database split.

2 Steganalysis Features

Modern steganalysis features are built as co-occurrence matrices from noise residuals. Below, we summarize the approach taken in [11]. Denoting an estimate of

the cover image pixel x_{ij} from its neighborhood $\mathcal{N}(\mathbf{Y}, i, j)$ as $\mathrm{Pred}(\mathcal{N}(\mathbf{Y}, i, j))$, the noise residual, $\mathbf{Z} = (z_{ij})$,

$$z_{ij} = y_{ij} - \mathrm{Pred}(\mathcal{N}(\mathbf{Y}, i, j)), \tag{2}$$

is quantized with a quantization step $q > 0$ and truncated to a finite dynamic range $\mathcal{T} = \{-T, -T+1, \ldots, T\}$:

$$r_{ij} \triangleq \mathrm{trunc}_T\left(\mathrm{round}(z_{ij}/q)\right). \tag{3}$$

The statistical properties of $\mathbf{R} = (r_{ij})$ are captured as joint probability mass functions (pmfs) or co-occurrence matrices of m neighboring residual samples in the horizontal and vertical direction. The horizontal co-occurrence for residual $\mathbf{R}$ is

$$\mathbf{C}^{(\mathrm{h})}_{\mathbf{d}} = \Pr(r_{ij} = d_1 \wedge \ldots \wedge r_{i,j+m-1} = d_m), \quad \mathbf{d} = (d_1, \ldots, d_m) \in \mathcal{T}^m, \tag{4}$$

while the vertical matrix, $\mathbf{C}^{(\mathrm{v})}_{\mathbf{d}}$, is defined analogically. Both have $(2T+1)^m$ elements.

Most pixel predictors are realized as shift-invariant finite-impulse response linear filters captured by a kernel matrix. For example, the kernel

$$\mathbf{K} = \begin{pmatrix} -0.25 & 0.5 & -0.25 \\ 0.5 & 0 & 0.5 \\ -0.25 & 0.5 & -0.25 \end{pmatrix}, \tag{5}$$

proposed in [19] predicts the value of the central pixel from its local 3×3 neighborhood using the operation of convolution: $\mathbf{K} \star \mathbf{Y}$.

Symmetries are conveniently utilized to further reduce the dimensionality of the co-occurrences and to make them better populated. Given $\mathbf{d} \in \mathcal{T}^m$, we assume that for natural images $\mathbf{C}_{\mathbf{d}} \approx \mathbf{C}_{-\mathbf{d}}$ and $\mathbf{C}_{\mathbf{d}} \approx \mathbf{C}_{\overleftarrow{\mathbf{d}}}$, $\overleftarrow{\mathbf{d}} = (d_m, d_{m-1}, \ldots, d_1)$. Symmetrization by sign means merging the bins $\mathbf{C}_{\mathbf{d}} + \mathbf{C}_{-\mathbf{d}}$, while symmetrization by direction requires merging $\mathbf{C}_{\mathbf{d}} + \mathbf{C}_{\overleftarrow{\mathbf{d}}}$.

For example, for $T = 2$ and $m = 4$, which are the parameters solely used in this paper, the original co-occurrence matrix, $\mathbf{C}_{\mathbf{d}}$, with $(2 \times 2 + 1)^4 = 625$ elements is reduced to 325 elements using the directional symmetry or 338 elements using the sign symmetry. When both symmetrizations are applied, the dimension is reduced to 169.

3 Motivation

We now provide heuristic arguments for why detectors that utilize joint statistics of neighboring residual samples are likely to outperform variants of the WS analysis. It is because the WS detector can be derived from the assumption that the individual residual values are independent. Detailed technical arguments appear in [7] and require proper treatment of quantization effects. The author derives a CFAR variant of the WS detector starting with the independence assumption imposed on residual samples obtaining the detector in an asymptotic limit of infinite pixel bit-depth.

Deriving the detector while considering dependencies among residuals would require tackling the difficult problem of estimating the covariance between residuals as well as higher-order moments from a rather limited data. Instead, in this paper we represent groups of neighboring residual samples with co-occurrence matrices and use machine learning rather than the likelihood ratio test. While this approach is suboptimal, it is tractable and, as shown below, greatly improves the accuracy of all variants of WS.

Researchers have been aware for quite a long time that by leveraging the dependencies among neighboring residual samples,[2] one can obtain quite substantial improvement in detecting steganographic changes. Steganalyzers working with features formed as joint or transition probability distributions as features were shown to outperform [27,25,12,11,13] all previously proposed attacks on LSB matching and the content-adaptive HUGO. In summary, it makes perfect sense to expect that the accuracy of the WS detector can be improved as well by considering higher-order statistical constructs from the residual.

4 Making Features Parity Aware

Features computed from noise residuals, which are outputs of linear filters, such as (5), "do not see" pixel parity as this information is lost when, for example, taking a difference between two pixel values. This means that such features will detect LSB matching and LSB replacement with approximately the same accuracy.

We now describe several ways how to make the features parity aware. To this end, we introduce the following notation. For image $\mathbf{X} \in \mathcal{I}^n$, we denote by $\dot{\mathbf{X}}$, $\tilde{\mathbf{X}}$, $\bar{\mathbf{X}}$ the image $\mathbf{X}$ after setting all its LSBs to zero, randomizing all LSBs, and flipping all LSBs, respectively. Formally,

$$\dot{x}_{ij} = x_{ij} - \mathrm{LSB}(x_{ij}), \tag{6}$$

$$\tilde{x}_{ij} = \dot{x}_{ij} + \varphi, \quad \varphi \text{ r.v. uniform on } \{0, 1\}, \tag{7}$$

$$\bar{x}_{ij} = x_{ij} + 1 - 2\mathrm{LSB}(x_{ij}). \tag{8}$$

The residuals of $\mathbf{X}, \dot{\mathbf{X}}, \tilde{\mathbf{X}}$, and $\bar{\mathbf{X}}$ will be denoted correspondingly as $\mathbf{R}, \dot{\mathbf{R}}, \tilde{\mathbf{R}}$, and $\bar{\mathbf{R}}$. In general, a feature computed from a residual $\mathbf{R}$ will be denoted as $\mathbf{f}(\mathbf{R})$.

Borrowing the idea from the WS detector, we define the concept of a "parity-aware residual." Given a residual $\mathbf{R} = (r_{ij})$, its parity-aware version is

$$\mathbf{R}^{(\pi)} = (r_{ij}^{(\pi)}), \quad r_{ij}^{(\pi)} = (1 - 2\mathrm{LSB}(x_{ij}))\, r_{ij}. \tag{9}$$

To make a feature vector of image $\mathbf{X}$ parity aware, one can follow the idea of Cartesian calibration [20] and augment it with a reference feature computed from $\dot{\mathbf{R}}$, $\tilde{\mathbf{R}}$, or $\bar{\mathbf{R}}$. We call this "calibration by parity." Additionally, we can compute the feature from the parity-aware residual, $\mathbf{f}(\mathbf{R}^{(\pi)})$.

[2] The dependencies are due to in-camera processing, such as denoising, filtering, color interpolation, and also due to the traces of content in the residual.

Table 1. Average detection error $\bar{P}_{\mathrm{E}}$ for LSB replacement with change rate $\beta = 0.01$ in uncompressed and JPEG 80 BOSSbase. Six different feature sets and their symmetrizations are tested; the last five are parity aware. The last set, $\mathbf{f}^{(663)}$, is the 663-dimensional merger of $[\mathbf{f}(\mathbf{R}), \mathbf{f}(\dot{\mathbf{R}})]$ and $\mathbf{f}(\mathbf{R}^{(\pi)})$ symmetrized as explained in the text.

	Source	JPEG 80				Uncompressed			
	Symm.	None	Both	Dir	Sign	None	Both	Dir	Sign
1	$\mathbf{f}(\mathbf{R})$	0.0164	0.0162	0.0158	0.0159	0.3261	0.3282	0.3246	0.3305
2	$[\mathbf{f}(\mathbf{R}), \mathbf{f}(\dot{\mathbf{R}})]$	0.0114	0.0103	0.0103	0.0106	0.1958	0.1971	0.1959	0.2007
3	$[\mathbf{f}(\mathbf{R}), \mathbf{f}(\bar{\mathbf{R}})]$	0.0139	0.0130	0.0141	0.0135	0.2534	0.2524	0.2497	0.2531
4	$[\mathbf{f}(\mathbf{R}), \mathbf{f}(\bar{\mathbf{R}})]$	0.0128	0.0123	0.0129	0.0128	0.2239	0.2281	0.2242	0.2286
5	$\mathbf{f}(\mathbf{R}^{(\pi)})$	0.0165	0.0398	0.0163	0.0388	0.1253	0.3456	0.1249	0.3480
6	$\mathbf{f}^{(663)}$	0.0086				0.1154			

4.1 Testing

In the remainder of this section, we test the above features on BOSSbase and its JPEG compressed version to investigate the efficiency of calibration by parity and the parity-aware residual as well as the effect of symmetrization on detection performance for both types of features. Since these experiments are investigative in nature, they will be carried out only for one type of residual $\mathbf{R}$ obtained using the predictor $\mathbf{K}$ (5). The basic (parity-unaware) feature is

$$\mathbf{f}(\mathbf{R}) = \mathbf{C}^{(\mathrm{h})}_{\mathbf{d}} + \mathbf{C}^{(\mathrm{v})}_{\mathbf{d}}, \tag{10}$$

obtained as sum of the horizontal and vertical co-occurrences[3] with parameters $T = 2$ and $m = 4$, and with total dimensionality of 625 in its non-symmetrized version.

Table 1 shows $\bar{P}_{\mathrm{E}}$ on BOSSbase and its version compressed with JPEG quality 80. The results are for a fixed change rate $\beta = 0.01$, six different feature sets, and four types of symmetrization. As expected, the detection error is significantly lower for decompressed JPEGs than for uncompressed images. The symmetrization also has a very different impact on the features. In general, features computed from the parity-aware residual, $\mathbf{R}^{(\pi)}$, should be symmetrized only directionally but not by sign. The symmetrization has a much lesser impact on features calibrated by parity, for which both the directional and sign symmetries can be applied. The best calibration by parity is by zeroing out the LSB plane, i.e., $[\mathbf{f}(\mathbf{R}), \mathbf{f}(\dot{\mathbf{R}})]$. For JPEG images, this type of calibration gives the best results while features computed from the parity-aware residual are the best for uncompressed images. Finally, combining calibration by zeroing-out the LSBs with parity-aware residual is beneficial as can be seen from the last row ($\mathbf{f}^{(663)}$) showing the 663-dimensional merger of $[\mathbf{f}(\mathbf{R}), \mathbf{f}(\dot{\mathbf{R}})]$ symmetrized by both direction and sign with $\mathbf{f}(\mathbf{R}^{(\pi)})$ symmetrized directionally.

[3] The symmetry of the kernel $\mathbf{K}$ allows us to add both co-occurrences.

The fluctuations over the ten database splits are all statistically insignificant as the MAD of $P_{\mathrm{E}}(i)$ over the runs (not shown) was between 5×10^{-4} on JPEGs and 4×10^{-3} for uncompressed images.

4.2 Analysis by Cover Source

In this section, we apply feature selection to reveal several interesting facts about the detection of LSB replacement using parity-aware features from Table **1**.

The dimensionality of $\mathbf{f}(\mathbf{R})$ and $[\mathbf{f}(\mathbf{R}), \mathbf{f}(\dot{\mathbf{R}})]$ symmetrized using both symmetries is $d = 169$ and 338, respectively, while the directionally-symmetrized $\mathbf{f}(\mathbf{R}^{(\pi)})$ has dimensionality of $d = 325$. We use a simple forward feature selection (FFS) method in which the features are selected sequentially one by one based on how much they improve the detection w.r.t. the union of those already selected. We start with the feature with the lowest individual detection error estimated from the training set. Having selected $k \geq 1$ features, the $k + 1$st feature is selected as the one among the $d - k$ remaining features that leads to the biggest drop in the error estimate when the union of all $k + 1$ features is used. This strategy continuously utilizes feedback of the ensemble classifier as it greedily minimizes the detection error in every iteration, taking thus the mutual dependencies among individual features into account. This is an example of a wrapper [23], which is a feature selection method using the machine-learning tool as a black-box and is thus classifier-dependent.

Decompressed JPEGs. We start with the source of JPEG compressed images. Table 2 (left) shows the results of the FFS when applied to the 169-dimensional feature vector $\mathbf{f}(\mathbf{R})$. We used a larger change rate $\beta = 0.02$ to make the effects more pronounced. The most remarkable phenomenon is the large decrease in detection error when the second bin is supplied to the best individual bin. While the second bin by itself has a very poor performance almost equal to random guessing, it extremely well *complements* the first bin. The error drops further with added bins but does so rather gradually after the initial drop. Note that the first bin corresponds to a residual four-tuple with large differences among neighboring samples. Such a group of values seems to be much less frequent in decompressed JPEGs than in their stego versions (c.f. the last column in the table) because the compression smooths the covers and thus empties this bin while the embedding repopulates it. The second bin serves as a reference, which is approximately invariant to embedding, and the pair together facilitates a very accurate detection. In fact, *all four* next selected bins, $k = 2, 3, 4, 5$, have a rather poor individual performance, suggesting that they all serve as different references to the first bin.

Remarkably, after merging only the first *three* bins, the cumulative error of 0.0215 is already lower than for the triples analysis – the best prior art performer (see Table 5). When all 169 features are used, the error drops further to 0.005. We remind that this result is obtained for a feature vector that is *unaware* of the pixel parity! Applying the FFS to $\mathbf{f}(\mathbf{R})$ Cartesian-calibrated by parity, $\mathbf{f}(\dot{\mathbf{R}})$, returns the same first four bins as for $\mathbf{f}(\mathbf{R})$, which is why we are not showing the results. This also implies that the main power of the detection is drawn from

Table 2. Forward feature selection strategy with change rate $\beta = 0.02$ in JPEG 80: cumulative and individual $\bar{P}_{\mathrm{E}}$, selected bins, and average bin count in cover/stego images. Left: symmetrized $\mathbf{f}(\mathbf{R})$, dimension 169. Right: directionally symmetrized $\mathbf{f}(\mathbf{R}^{(\pi)})$, dimension 325. The last row is obtained when all features are used.

k	$\bar{P}_{\mathrm{E}}^{(\mathrm{cum})}$	$\bar{P}_{\mathrm{E}}^{(\mathrm{ind})}$	Bin	Bin count
1	0.2986	0.2986	(-1 2 -1 0)	1509/2291
2	0.0377	0.4798	(-1 -1 1 0)	4878/5061
3	0.0215	0.4582	(-2 0 0 0)	2939/2746
4	0.0190	0.4721	(-2 0 -1 1)	940/989
5	0.0149	0.4761	(-1 2 -2 0)	2155/2262
169	0.0050	-	-	-

$\bar{P}_{\mathrm{E}}^{(\mathrm{cum})}$	$\bar{P}_{\mathrm{E}}^{(\mathrm{ind})}$	Bin	Bin count
0.2226	0.2226	(-1 -1 -1 0)	2950/5730
0.0370	0.4660	(0 0 1 0)	10130/9470
0.0261	0.4712	(0 -1 -1 0)	3930/4190
0.0209	0.4433	(0 0 0 0)	116120/91530
0.0117	0.4970	(1 0 -2 2)	650/650
-	-	-	-

the singular property of the cover source (compression "empties out" certain bins) rather than the parity asymmetry of LSB replacement. This is additionally confirmed by the fact that LSB matching can be detected with the same feature vector $\mathbf{f}(\mathbf{R})$ equally reliably as LSB replacement.

Furthermore, the best individual bin $(-1, 2, -1, 0)$ seems to be universal across sources of images with suppressed noise, which immediately disperses any thoughts that the co-occurrence bins might somehow utilize JPEG compatibility for detection. We confirmed this by repeating the same experiment with the feature vector $\mathbf{f}(\mathbf{R})$ for BOSSbase images denoised using the 3×3 Wiener filter with noise variance $\sigma^2 = 2, 5, 10$ and for BOSSbase denoised using the 3×3 median filter.[4]

The 325-dimensional feature vector $\mathbf{f}(\mathbf{R}^{(\pi)})$ obtained from the parity-aware residual exhibits a similar initial phenomenon, see Table 2 (right). The best individually performing bin is now different than in images with suppressed noise, which only strengthens our interpretation above.

Uncompressed Images. The second experiment was carried out on the uncompressed BOSSbase. In Table 3 (left), we report the results for the best-performing bins obtained from the parity-aware residual. Although the cumulative error now falls off much slower than for decompressed JPEGs, we again observe a large initial drop – the best individual performer is supplied with a reference bin that is by itself a random guesser. Interestingly, the second selected bin is the negative of the first bin. In fact, the same is true for the first eight selected bin pairs! To obtain insight into why the bins pair up in this manner, realize that $E[r_{ij}^{(\pi)}] = 0$ for unchanged pixels, while $E[r_{ij}^{(\pi)}] = -1$ whenever the pixel ij was changed. Thus, while both bins, $\mathbf{d}, -\mathbf{d} \in \mathcal{T}^4$, occur equally likely in covers, in stego images the one with more negative values is more populated than its negative counterpart. The reason why the boundary bin $(-2, -1, 0, 0)$ was chosen as the best can be explained by its population. While there are other good individual performers with individual errors in the range $P_{\mathrm{E}} \approx 0.42 - 0.45$, they are less populated.

[4] We used Matlab commands `wiener2` and `medfilt2`.

Table 3. Forward feature selection strategy for $\mathbf{f}(\mathbf{R}^{(\pi)})$, dimension 325, for change rate $\beta = 0.02$: cumulative and individual $\bar{P}_{\mathrm{E}}$, selected bins, and average bin count in cover/stego images. Left: uncompressed images. Right: denoised images. The last row is obtained when all features are used.

k	$\bar{P}_{\mathrm{E}}^{(\mathrm{cum})}$	$\bar{P}_{\mathrm{E}}^{(\mathrm{ind})}$	Bin	Bin count	Wie 2	Wie 5	Wie 10	Med
1	0.4126	0.4126	(-2 -1 0 0)	1353/1536	0.3277	0.2988	0.2536	0.2729
2	0.2164	0.4954	(2 1 0 0)	1323/1320	0.1130	0.0709	0.0620	0.0474
3	0.1810	0.4866	(-2 -2 -1 -2)	1912/1976	0.0226	0.0491	0.0365	0.0111
4	0.1489	0.4910	(2 2 1 2)	1901/1868	0.0223	0.0354	0.0293	0.0092
5	0.1438	0.4915	(-2 0 -2 -2)	1503/1478	0.0222	0.0299	0.0236	0.0087
325	0.0384	-	-	-	0.0172	0.0133	0.0110	0.0021

About 30 features are enough to obtain a lower detection error than the best structural performer – the WS analysis with moderated weights with bias correction (see Table 5).

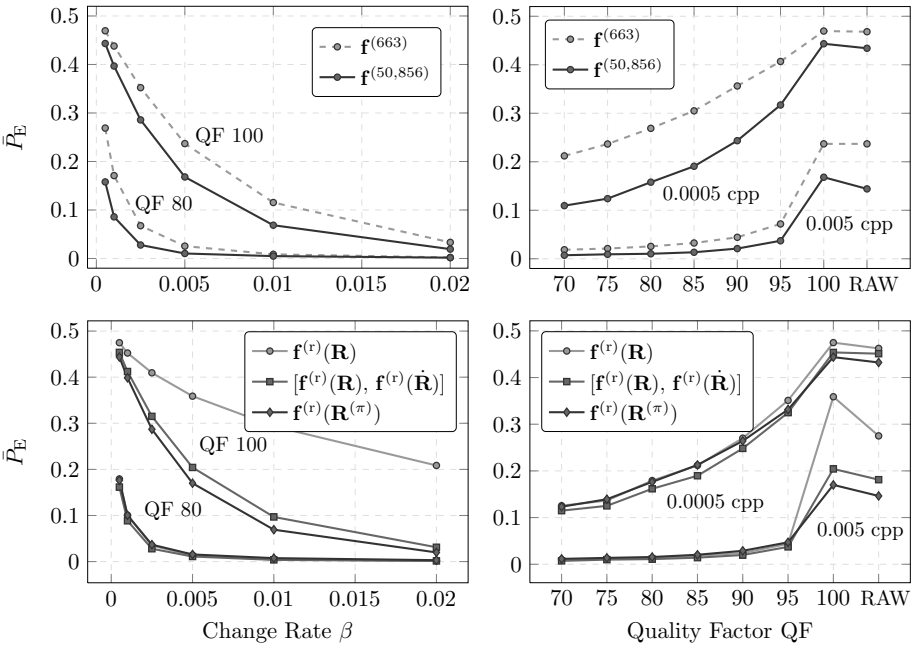

Fig. 1. Average detection error $\bar{P}_{\mathrm{E}}$ for different versions of the rich model (see text for details). Left: dependence on the change rate for two selected quality factors. Right: Dependence on the quality factor for two change rates.

Denoised Images. The last investigative experiment was carried out for four different versions of BOSSbase denoised using the 3×3 Wiener filter with

Table 4. Comparison of the average detection error $\bar{P}_{\rm E}$ for the best prior art detector, which is the triples analysis (Tr) and weighted stego-image with bias correction (WSb) marked by the symbol *, the feature $\mathbf{f}^{(663)}$ from Section 4, and the rich model $\mathbf{f}^{(50,856)}$

β	Det	70	75	80	85	90	95	100	UNCOMP.
0.0005	Tr/WSb	0.4022	0.4148	0.4190	0.4343	0.4464	0.4637	0.4767*	0.4776*
	$\mathbf{f}^{(663)}$	0.2121	0.2366	0.2689	0.3050	0.3563	0.4068	0.4695	0.4681
	$\mathbf{f}^{(50,856)}$	0.1095	0.1240	0.1579	0.1907	0.2435	0.3170	0.4433	0.4340
0.001	Tr/WSb	0.3168	0.3411	0.3521	0.3728	0.3961	0.4296	0.4547*	0.4536*
	$\mathbf{f}^{(663)}$	0.1270	0.1458	0.1709	0.2044	0.2558	0.3266	0.4380	0.4380
	$\mathbf{f}^{(50,856)}$	0.0610	0.0699	0.0858	0.1078	0.1439	0.2126	0.3968	0.3743
0.0025	Tr/WSb	0.1738	0.1973	0.2163	0.2372	0.2742	0.3350	0.3875*	0.3869*
	$\mathbf{f}^{(663)}$	0.0527	0.0575	0.0676	0.0816	0.1094	0.1704	0.3522	0.3522
	$\mathbf{f}^{(50,856)}$	0.0185	0.0245	0.0278	0.0365	0.0504	0.0869	0.2857	0.2512
0.005	Tr/WSb	0.0852	0.1014	0.1139	0.1283	0.1682	0.2346	0.2918*	0.2925*
	$\mathbf{f}^{(663)}$	0.0186	0.0211	0.0255	0.0325	0.0443	0.0718	0.2369	0.2369
	$\mathbf{f}^{(50,856)}$	0.0073	0.0092	0.0103	0.0134	0.0210	0.0371	0.1681	0.1441
0.01	Tr/WSb	0.0388	0.0464	0.0537	0.0628	0.0832	0.1341*	0.1697*	0.1662*
	$\mathbf{f}^{(663)}$	0.0045	0.0066	0.0086	0.0125	0.0186	0.0302	0.1154	0.1154
	$\mathbf{f}^{(50,856)}$	0.0027	0.0032	0.0049	0.0067	0.0113	0.0203	0.0686	0.0582
0.02	Tr/WSb	0.0199	0.0225	0.0268	0.0327	0.0430	0.0613	0.0675*	0.0664*
	$\mathbf{f}^{(663)}$	0.0009	0.0013	0.0021	0.0048	0.0079	0.0166	0.0332	0.0332
	$\mathbf{f}^{(50,856)}$	0.0010	0.0011	0.0017	0.0032	0.0066	0.0126	0.0193	0.0173

noise variance $\sigma^2 = 2, 5, 10$ and the 3×3 median filter. For the directionally-symmetrized $\mathbf{f}(\mathbf{R}^{(\pi)})$ we show in Table 3 (right) the cumulative detection error when selecting the five best bins using the FFS. The last row shows the detection error $\bar{P}_{\rm E}$ when using all 325 features $\mathbf{f}(\mathbf{R}^{(\pi)})$. The best performing bin was again $(-1, 2, -1, 0)$, as in case of decompressed JPEGs, with the exception of Wiener-filter images with $\sigma^2 = 2$ where the best bin was the same as the one found for uncompressed images. In all cases, we observed a sharp drop in detection error after the second bin is added to the best bin. Images processed by the median 3×3 filter appear to be particularly easy for detection of LSB replacement. For these four sources, the FFS did not seem to select the bins in pairs as observed for uncompressed images, which indicates that the detection utilizes the low level of noise of covers more than the singularity of LSB replacement.

5 Scaling Up the Image Model

In this section, we scale up our approach to the rich image model built in [11]. Due to the complexity of this model and the limited space in this paper, we cannot describe it here in detail and instead refer to the original publication. We use the predictors described in Section IV of [11] designed to better adapt to content around edges and in textures. The resulting set of 39 feature sets obtained with $T = 2$, $q = 1$, and $m = 4$ forms the rich model feature vector $\mathbf{f}^{(\rm r)}$.

Table 5. Detection error $\bar{P}_{\mathrm{E}}$ for five structural detectors, six change rates, β, and eight cover sources: uncompressed BOSSbase (UNC) and its JPEG compressed versions using quality factors 70,75,...,100. Shaded in gray are the best results for each change rate. The acronyms are explained in Appendix A.

β	Det	70	75	80	85	90	95	100	UNC.
0.0005	SP	0.4725	0.4727	0.4752	0.4754	0.4792	0.4800	0.4849	0.4855
	WSb	0.4265	0.4323	0.4388	0.4477	0.4571	0.4642	0.4767	0.4776
	WS	0.4246	0.4240	0.4347	0.4422	0.4538	0.4635	0.4783	0.4768
	Tr	0.4022	0.4148	0.4190	0.4343	0.4464	0.4637	0.4853	0.4839
	AUMP	0.4564	0.4559	0.4620	0.4656	0.4698	0.4746	0.4805	0.4813
0.001	SP	0.4458	0.4448	0.4501	0.4510	0.4587	0.4626	0.4709	0.4719
	WSb	0.3717	0.3768	0.3879	0.3978	0.4124	0.4316	0.4547	0.4536
	WS	0.3580	0.3654	0.3768	0.3911	0.4086	0.4310	0.4548	0.4542
	Tr	0.3168	0.3411	0.3521	0.3728	0.3961	0.4296	0.4702	0.4673
	AUMP	0.4135	0.4139	0.4236	0.4317	0.4386	0.4514	0.4611	0.4614
0.0025	SP	0.3681	0.3768	0.3812	0.3854	0.3954	0.4066	0.4275	0.4255
	WSb	0.2639	0.2690	0.2809	0.2922	0.3124	0.3436	0.3875	0.3869
	WS	0.2356	0.2460	0.2630	0.2804	0.3069	0.3437	0.3878	0.3898
	Tr	0.1738	0.1973	0.2163	0.2372	0.2742	0.3350	0.4243	0.4185
	AUMP	0.3037	0.3056	0.3205	0.3392	0.3547	0.3812	0.4056	0.4044
0.005	SP	0.2766	0.2842	0.2909	0.2981	0.3106	0.3271	0.3595	0.3600
	WSb	0.1831	0.1838	0.1907	0.1990	0.2121	0.2386	0.2918	0.2925
	WS	0.1415	0.1563	0.1690	0.1848	0.2109	0.2392	0.2975	0.2939
	Tr	0.0852	0.1014	0.1139	0.1283	0.1682	0.2346	0.3548	0.3432
	AUMP	0.1962	0.2015	0.2153	0.2316	0.2494	0.2867	0.3256	0.3276
0.01	SP	0.1756	0.1802	0.1879	0.1949	0.2035	0.2195	0.2594	0.2576
	WSb	0.1083	0.1120	0.1164	0.1181	0.1251	0.1341	0.1697	0.1662
	WS	0.0730	0.0848	0.0935	0.1048	0.1232	0.1397	0.1770	0.1722
	Tr	0.0388	0.0464	0.0537	0.0628	0.0832	0.1377	0.2494	0.2383
	AUMP	0.1064	0.1081	0.1195	0.1316	0.1513	0.1818	0.2146	0.2162
0.02	SP	0.0916	0.0931	0.0989	0.0979	0.1094	0.1168	0.1447	0.1410
	WSb	0.0550	0.0565	0.0587	0.0592	0.0599	0.0613	0.0675	0.0664
	WS	0.0319	0.0359	0.0408	0.0494	0.0585	0.0676	0.0769	0.0714
	Tr	0.0199	0.0225	0.0268	0.0327	0.0430	0.0696	0.1392	0.1277
	AUMP	0.0498	0.0516	0.0563	0.0629	0.0790	0.1029	0.1231	0.1181

We test the following four versions of the rich model (the dimensionalities are in brackets):

1. $\mathbf{f}^{(\mathrm{r})}(\mathbf{R})$ symmetrized by both sign and direction (12,753);
2. $\mathbf{f}^{(\mathrm{r})}(\mathbf{R}^{(\pi)})$ symmetrized only directionaly (25,350);
3. $[\mathbf{f}^{(\mathrm{r})}(\mathbf{R}), \mathbf{f}^{(\mathrm{r})}(\dot{\mathbf{R}})]$ symmetrized by both sign and direction (25,506);
4. Merger of 2) and 3): $\mathbf{f}^{(50,856)} = [\mathbf{f}^{(\mathrm{r})}(\mathbf{R}), \mathbf{f}^{(\mathrm{r})}(\dot{\mathbf{R}}), \mathbf{f}^{(\mathrm{r})}(\mathbf{R}^{(\pi)})]$ (50,856).

Note that we do not symmetrize $\mathbf{f}^{(\mathrm{r})}(\mathbf{R}^{(\pi)})$ by sign as this would compromise its parity awareness as seen in Table 1.

Table 4 contrasts the performance of $\mathbf{f}^{(50,856)}$ with $\mathbf{f}^{(663)}$ and the best prior-art detectors from Table 5. The top two charts in Figure (1) show that the $\mathbf{f}^{(50,856)}$ model brings improvement over the 663-dimensional model especially for small change rates and high quality factors / uncompressed images. The two bottom charts inform us about the importance of making the feature vector $\mathbf{f}^{(\mathrm{r})}$ parity aware. The gain is the biggest for high-quality JPEGs and uncompressed images and it also increases with the change rate.

6 Conclusion

In 2005, the author of [14] expressed the following opinion about the state of the art in detection of LSB replacement: "... Because it makes full use of structural information, in some sense this framework [structural steganalysis] should be the last word on the detection of LSB replacement, although many practical questions remain open." In this paper, we challenge the supremacy of structural detectors and show that feature-based detectors with parity-aware features can significantly outperform all structural detectors as well as variants of WS analysis in both decompressed JPEG images and in uncompressed images. After all, it is only natural that the WS analysis with its limiting assumption of independent residual samples can be markedly improved as it has been shown in the literature before that utilizing dependencies in noise residual is quite important for detection of steganography.

Although the largest gain is demonstrated for high-dimensional rich models, state of the art can be outperformed using as few as three co-occurrence bins in decompressed JPEGs and thirty bins for uncompressed images. Our analysis shows that features built as co-occurrences of neighboring noise residuals are especially effective for detection in images with low level of noise, such as decompressed JPEGs or low-pass filtered images. In fact, here the detection strength is almost entirely in the peculiarity of the cover source rather than the asymmetry of the embedding operation (LSB replacement) as comparable detection accuracy can be obtained for LSB matching.

We introduce and study two general methods for making features parity aware – by calibration by parity (adding features computed from the image with zeroed-out LSBs) and by computing the features from a parity-aware residual. The latter is especially effective for steganalysis in uncompressed images.

Our approach has some obvious limitations imposed by the necessity to build a classifier. In particular, it is only feasible when sufficiently many images from a given source are available. For an unknown source, the accuracy of detection will undoubtedly be negatively affected by the mismatch between the training and testing data. Thus, for practical applications, quantitative LSB detectors and especially the CFAR detector of [7] will still be very important and useful tools. If the cover source is known, however, classifiers, such as those proposed here, offer a definitive advantage in terms of detection accuracy. The rich models,

and in general any high-dimensional steganalysis, require extensive computing resources, which limits them to primarily off-line applications rather then real-time traffic monitoring. We note that the classifier training in high dimensions is quite feasible with tools, such as the ensemble classifier [22]. It is the time needed to compute the feature vector, that needs to be done for each analyzed image, that limits the practical use of such highly complex detectors.

Last but not least, our study seems to hint at new directions in structural steganalysis. We noticed a surprising universality across a wide spectrum of cover sources. Certain co-occurrence bins appear to be the overall best performers when accompanied with suitable reference features that by themselves are random guessers. In uncompressed images, bins of the parity-aware residual should be combined in mutually-negative pairs. A study with a simplified version of the residual, such as the second-order differences, may reveal well-defined flows between "trace sets" indexed by the residuals that might eventually lead to novel structural attacks. This work also reveals a possible way how to describe in a unified manner the WS analysis and structural detectors, which is a very exciting topic that we do not further elaborate on in this paper due to lack of space.

Acknowledgments. The work on this paper was partially supported by Air Force Office of Scientific Research under the research grant number FA9550-08-1-0084. 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. The authors would like to thank Vojtěch Holub, Miroslav Goljan, and Rainer Böhme for useful discussions, and Lionel Fillatre for help with implementing the AUMP detector.

A Prior Art

To establish a baseline and to identify the current most accurate LSB replacement detectors, we report here the results of five attacks that we consider state of the art: SP analysis [5], WS analysis with prediction kernel $\mathbf{K}$ (5) with moderated weights with (WSb) and without (WS) bias correction [19], triples analysis with $m, n \in \{-5, \ldots, 5\}$ (notation used as in [14]), and the AUMP detector [7] implemented with the recommended pixel block size $m = 16$, $q = 6$ (polynomial degree 5), and, per author's recommendation and in contrast to the paper, $\max\{1, \hat{\sigma}\}$ as an estimate of the standard deviation to assure numerical stability. The code for all detectors is available for download at: `http://dde.binghamton.edu/download/structural_lsb_detectors`.

Table 5 portrays triples analysis as the most accurate for decompressed JPEGs up to the quality factor of about 95 when it is outperformed by WSb, which is the best also for raw images. Our results for SP, WSb, WS, and triples seem compatible with previous art, at least as much as one can judge by results on different image sources. However, we observed a disturbingly large discrepancy between our results

and what was reported on the *same image database* in [7] for WS as well as the SP. The author reports the entire ROC curves for relative payload $R = 0.05$, which corresponds to change rate $\beta = 0.025$ since the author is not considering any matrix embedding at the sender. Reading out the P_{E} from the ROC as the most distant point to the main diagonal in Fig. 5 in [7], the WS method and the weighted SP achieve $P_{\mathrm{E}} \approx 0.2$ and $P_{\mathrm{E}} \approx 0.45$, which is significantly worse than our results, $\bar{P}_{\mathrm{E}} = 0.0664$ and $\bar{P}_{\mathrm{E}} = 0.1410$, respectively, obtained for the change rate $\beta = 0.02$ (which is additionally slightly smaller than $R/2 = 0.025$).

References

1. Bas, P., Filler, T., Pevný, T.: "Break Our Steganographic System": The Ins and Outs of Organizing BOSS. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 59–70. Springer, Heidelberg (2011)
2. Böhme, R.: Advanced Statistical Steganalysis. Springer, Heidelberg (2010)
3. Cogranne, R., Zitzmann, C., Fillatre, L., Retraint, F., Nikiforov, I., Cornu, P.: A Cover Image Model For Reliable Steganalysis. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 178–192. Springer, Heidelberg (2011)
4. Dumitrescu, S., Wu, X.: LSB steganalysis based on higher-order statistics. In: Proceedings of the 7th ACM Multimedia & Security Workshop, New York, August 1-2, pp. 25–32 (2005)
5. Dumitrescu, S., Wu, X., Memon, N.D.: On steganalysis of random LSB embedding in continuous-tone images. In: Proceedings IEEE, International Conference on Image Processing, ICIP 2002, Rochester, NY, September 22-25, pp. 324–339 (2002)
6. Dumitrescu, S., Wu, X., Wang, Z.: Detection of LSB steganography via sample pair analysis. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 355–372. Springer, Heidelberg (2003)
7. Fillatre, L.: Adaptive steganalysis of least significant bit replacement in grayscale images. IEEE Transactions on Signal Processing 60, 556–569 (2012)
8. Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press (2009)
9. Fridrich, J., Goljan, M.: On estimation of secret message length in LSB steganography in spatial domain. In: Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI, San Jose, CA, January 19–22, vol. 5306, pp. 23–34 (2004)
10. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in grayscale and color images. In: Proceedings of the ACM, Special Session on Multimedia Security and Watermarking, Ottawa, Canada, October 5, pp. 27–30 (2001)
11. Fridrich, J., Kodovský, J.: Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security (to appear, 2012)
12. Fridrich, J., Kodovský, J., Holub, V., Goljan, M.: Steganalysis of Content-Adaptive Steganography in Spatial Domain. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 102–117. Springer, Heidelberg (2011)
13. Gul, G., Kurugollu, F.: A New Methodology in Steganalysis: Breaking Highly Undetectable Steganograpy (HUGO). In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 71–84. Springer, Heidelberg (2011)
14. Ker, A.D.: A General Framework for Structural Steganalysis of LSB Replacement. In: Barni, M., Herrera-Joancomartí, J., Katzenbeisser, S., Pérez-González, F. (eds.) IH 2005. LNCS, vol. 3727, pp. 296–311. Springer, Heidelberg (2005)

15. Ker, A.D.: Fourth-order structural steganalysis and analysis of cover assumptions. In: Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII, San Jose, CA, January 16–19, vol. 6072, pp. 25–38 (2006)
16. Ker, A.D.: A Fusion of Maximum Likelihood and Structural Steganalysis. In: Furon, T., Cayre, F., Doërr, G., Bas, P. (eds.) IH 2007. LNCS, vol. 4567, pp. 204–219. Springer, Heidelberg (2008)
17. Ker, A.D.: Optimally weighted least-squares steganalysis. In: Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, January 29-February 1, vol. 6505, pp. 6-1–6-16 (2007)
18. Ker, A.D.: Steganalysis of embedding in two least significant bits. IEEE Transactions on Information Forensics and Security 2, 46–54 (2007)
19. Ker, A.D., Böhme, R.: Revisiting weighted stego-image steganalysis. In: Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, San Jose, CA, January 27–31, vol. 6819, pp. 5-1–5-17 (2008)
20. Kodovský, J., Fridrich, J.: Calibration revisited. In: Proceedings of the 11th ACM Multimedia & Security Workshop, Princeton, NJ, September 7–8, pp. 63–74 (2009)
21. Kodovský, J., Fridrich, J.: Steganalysis in high dimensions: Fusing classifiers built on random subspaces. In: Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics of Multimedia XIII, San Francisco, CA, January 23–26, vol. 7880, pp. OL 1–OL 13 (2011)
22. Kodovský, J., Fridrich, J., Holub, V.: Ensemble classifiers for steganalysis of digital media. IEEE Transactions on Information Forensics and Security 7(2), 432–444 (2012)
23. Lal, T.N., Chapelle, O., Weston, J., Elisseeff, A.: Embedded Methods. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (eds.) Feature Extraction: Foundations and Applications. STUDFUZZ, vol. 207, pp. 137–165. Springer, Heidelberg (2006)
24. Lu, P., Luo, X., Tang, Q., Shen, L.: An Improved Sample Pairs Method for Detection of LSB Embedding. In: Fridrich, J. (ed.) IH 2004. LNCS, vol. 3200, pp. 116–127. Springer, Heidelberg (2004)
25. Pevný, T., Bas, P., Fridrich, J.: Steganalysis by subtractive pixel adjacency matrix. IEEE Transactions on Information Forensics and Security 5(2), 215–224 (2010)
26. Zitzmann, C., Cogranne, R., Retraint, F., Nikiforov, I., Fillatre, L., Cornu, P.: Statistical Decision Methods in Hidden Information Detection. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 163–177. Springer, Heidelberg (2011)
27. Zo, D., Shi, Y.Q., Su, W., Xuan, G.: Steganalysis based on Markov model of thresholded prediction-error image. In: Proceedings IEEE, International Conference on Multimedia and Expo, Toronto, Canada, July 9-12, pp. 1365–1368 (2006)