Phase-Aware Projection Model for Steganalysis of JPEG Images

Vojtěch Holub and Jessica Fridrich

Department of ECE, SUNY Binghamton, NY, USA

State-of-the-art JPEG steganographic algorithms, such as J-UNIWARD, are currently better detected in the spatial domain rather than the JPEG domain. Rich models built from pixel residuals seem to better capture the impact of embedding than features constructed as co-occurrences of quantized JPEG coefficients. However, when steganalyzing JPEG steganographic algorithms in the spatial domain, the pixels' statistical properties vary because of the underlying 8×8 pixel grid imposed by the compression. In order to detect JPEG steganography more accurately, we split the statistics of noise residuals based on their phase w.r.t. the 8×8 grid. Because of the heterogeneity of pixels in a decompressed image, it also makes sense to keep the kernel size of pixel predictors small as larger kernels mix up qualitatively different statistics more, losing thus on the detection power. Based on these observations, we propose a novel feature set called PHase Aware pRojection Model (PHARM) in which residuals obtained using a small number of small-support kernels are represented using first-order statistics of their random projections as in the projection spatial rich model PSRM. The benefit of making the features "phase-aware" is shown experimentally on selected modern JPEG steganographic algorithms with the biggest improvement seen for J-UNIWARD. Additionally, the PHARM feature vector can be computed at a fraction of computational costs of existing projection rich models.

Keywords: Steganalysis, JPEG, projection, phase, PHARM, steganography

1. INTRODUCTION

Detectors of JPEG steganography can be divided into two major categories – those designed using hypothesis testing based on a cover model and detectors constructed as classifiers trained on examples of cover and stego images. Detectors from the first category can be guaranteed to be optimal within the chosen cover model and thus designed to maximize the detection power for a prescribed false-alarm probability. However, they are limited to steganographic schemes that introduce easily detectable artifacts into the statistics of DCT coefficients, such as Jsteg[18,22] or OutGuess.[19] However, modern steganographic schemes, example of which is J-UNIWARD[11] and Uniform Embedding Distortion (UED),[5] and many side-informed schemes, such as SI-UNIWARD,[11] Normalized Perturbed Quantization (NPQ),[12] and Entropy-Block Steganography (EBS),[20] do not introduce easily detectable artifacts into the statistics of quantized DCT coefficients. Such steganographic techniques are best detected using feature-based detectors built using machine learning.

Statistical features used for JPEG steganalysis can be divided into two types depending on whether the feature vector is computed from quantized DCT coefficients or from pixels of the decompressed JPEG image. In the former case, the features are typically formed as two-dimensional co-occurrences of neighboring intra and inter-block coefficients. Examples of this approach appear in Refs [2,16,17] and, most prominently, in the JPEG Rich Model (JRM),[14] which can be considered as the current state of the art. In the latter case, the features are extracted from the spatial-domain representation of the JPEG image. One may directly use the Spatial Rich Model (SRM),[3] the Projection Spatial Rich Model (PSRM),[8] or its version, PSRMQ3, specifically adapted for detection in the JPEG domain by increasing the quantization step.[8]

As pointed out in Ref. [9], features extracted directly from the quantized DCT coefficients (e.g., JRM) appear more effective against older JPEG steganographic algorithms that introduce characteristic artifacts into the distribution of JPEG coefficients, such as nsF5.[4,21] Detection of the modern JPEG steganographic schemes listed above is far more accurate with features that were originally designed for the spatial domain, such as the PSRMQ3, or with features combined across domains – the JSRM, which is the union of JRM and SRM.

E-mail: vojtech_holub@yahoo.com, fridrich@binghamton.edu; `http://dde.binghamton.edu`.

Recently, the authors of this paper introduced a new low-complexity feature set called DCTR,[10] which is computed in the spatial domain but was purposely designed to detect JPEG steganography. Unlike the above rich feature representations, DCTR uses only first-order statistics (histograms) of noise residuals obtained with DCT bases as the pixel predictor kernels. Most importantly, before computing the histograms, the residuals are surgically divided into subsets based on their position within the 8×8 JPEG grid and the bases' symmetries. It is this splitting of the residuals, the so-called phase-awareness, that gives DCTR its power because when a JPEG image is decompressed to the spatial domain, the pixel statistics depend on their location within the JPEG grid.

In this paper, we take this observation to the next level by showing that it is also advantageous to use pixel predictors based on kernels with a small support to avoid mixing random variables with fundamentally different pedigree. Instead of diversifying the features by including a large number of pixel predictors, we use only a small number of small-support linear pixel predictors and instead diversify the model in a similar manner as in the projection spatial rich model (PSRM). We call the new feature set the PHase-Aware pRojection Model (PHARM). It offers the best detection performance on modern steganographic schemes while still enjoying a low computational complexity.

For completeness and ease of reading, in the next section we review the basic building blocks of the SRM and the PSRMQ3 feature sets. Section 3 describes the proposed PHARM, including experiments based on which we selected its parameters. In Section 4, we provide the results of all experiments aimed at comparing the accuracy and computational complexity of PHARM with current state-of-the-art rich models. The paper is concluded in Section 5.

2. PRELIMINARIES: THE SRM AND PSRM

In this section, we review some concepts from prior art, including the basic principles for building the SRM and its projection version, the PSRM. This is done in order to make the paper self-contained and easier to read.

The symbols $\mathbf{X}, \mathbf{Y} \in \{0, \ldots, 255\}^{n_1 \times n_2}$ will be used exclusively for two-dimensional arrays of pixel values in an $n_1 \times n_2$ grayscale cover and stego image, respectively. For simplicity, we will assume that both n_1 and n_2 are multiples of 8. Elements of a matrix will be denoted with the corresponding lower case letter. The pair of indices k, l will always be used to index elements in an $n_1 \times n_2$ matrix.

2.1 Residuals

Both the SRM and the PSRM start with estimating the noise component of a given image using 45 different pixel predictors. For a complete description of the predictors, see Ref. [3]. There are two types of pixel predictors – linear and non-linear. Each linear predictor is a shift-invariant finite-impulse response linear filter described by a kernel matrix $\mathbf{K}$. By subtracting the predicted image from the original image, we obtain the so-called noise residual $\mathbf{Z} = (z_{kl})$, which is a matrix of the same dimension as $\mathbf{X}$:

$$\mathbf{Z} = \mathbf{K} * \mathbf{X} - \mathbf{X}. \tag{1}$$

In (1), the symbol $'*'$ denotes the convolution with $\mathbf{X}$ mirror-padded so that $\mathbf{K} * \mathbf{X}$ has the same dimension as $\mathbf{X}$. This corresponds to the 'conv2' Matlab command with the parameter 'same'.

An example of a simple linear residual is $z_{kl} = x_{k,l+1} - x_{k,l}$, which is the difference between a pair of horizontally neighboring pixels. In this case, the predictor is $\mathbf{K} = (\begin{array}{cc} 0 & 1 \end{array})$, which means that we estimate the pixel value as its horizontally adjacent pixel. The submodel 'spam14h' in the SRM uses the same predictor.

All non-linear predictors in the SRM are obtained by taking the minimum or maximum of two or more residuals obtained using linear predictors. For example, one can obtain two predictions for pixel x_{ij} – one using the horizontal and one using the vertical neighbor obtaining thus one horizontal and one vertical residual $\mathbf{Z}^{(\mathrm{h})} = (z_{kl}^{(\mathrm{h})})$, $\mathbf{Z}^{(\mathrm{v})} = (z_{kl}^{(\mathrm{v})})$:

$$z_{kl}^{(\mathrm{h})} = x_{k,l+1} - x_{kl}, \tag{2}$$

$$z_{kl}^{(\mathrm{v})} = x_{k+1,l} - x_{kl}. \tag{3}$$

Using these two residuals, one can compute two non-linear 'minmax' residuals as:

$$z_{kl}^{(\min)} = \min\{z_{kl}^{(h)}, z_{kl}^{(v)}\}, \tag{4}$$

$$z_{kl}^{(\max)} = \max\{z_{kl}^{(h)}, z_{kl}^{(v)}\}. \tag{5}$$

Another example of a non-linear residual called the 'minmax41' in Ref. [3] is computed as

$$z_{kl}^{(\min)} = \min\{x_{k,l-1} - x_{kl}, \, x_{k,l+1} - x_{kl}, \, x_{k-1,l} - x_{kl}, \, x_{k+1,l} - x_{kl}\}.$$

$$z_{kl}^{(\max)} = \max\{x_{k,l-1} - x_{kl}, \, x_{k,l+1} - x_{kl}, \, x_{k-1,l} - x_{kl}, \, x_{k+1,l} - x_{kl}\}. \tag{6}$$

2.2 Projections of residuals

For the full description of the PSRMQ3, we refer the reader to the original publication.[8] Below, we provide only a highly condensed overview of the PSRMQ3 feature set needed to explain the PHARM features.

The PSRMQ3 features are computed from the spatial domain representation of the JPEG image after rounding the pixel values to integers. We note at this point that in the PHARM feature set, and the recently proposed DCTR,[10] we do not round the pixel values to avoid any potential loss of information.

Unlike the SRM, which captures the statistical properties of residuals using four-dimensional co-occurrences, the PSRMQ3 uses the first-order statistics of projections of residuals onto multiple random directions. Given a noise residual $\mathbf{Z}$, a slightly simplified algorithm for computing the PSRMQ3 is:

1. Generate ν random matrices $\mathbf{\Pi}^{(i)} \in \mathbb{R}^{r \times s}$, $i \in \{1, \ldots, \nu\}$.

   - r, s are uniformly randomly selected from $\{1, \ldots, s\}$, where $s > 0$ is an integer parameter,
   - the elements of $\mathbf{\Pi}^{(i)}$ are independent realizations of a standard normal random variable $\mathcal{N}(0, 1)$,
   - the elements are normalized so that the Frobenius norm* $\left\|\mathbf{\Pi}^{(i)}\right\|_F = 1$.

2. For each $i \in \{1, \ldots, \nu\}$, compute the projections $\mathbf{P}^{(i)} \triangleq \mathbf{Z} * \mathbf{\Pi}^{(i)}$.

3. Divide $\mathbf{P}^{(i)} = (p_{kl}^{(i)})$ by the quantization step q and quantize with a quantizer Q_T with $T + 1$ centroids $\mathcal{Q} = \{1/2, 3/2, \ldots, T + 1/2\}$ for linear residuals and $\mathcal{Q} = \{-T - 1/2, -T + 1/2, \ldots, T + 1/2\}$ with $2T + 2$ centroids for non-linear residuals:

$$p_{kl}^{(i)} \leftarrow Q_T(p_{kl}^{(i)}/q). \tag{7}$$

4. Compute ν separate histograms of the quantized values:

   - $\mathbf{h}_j^{(i)} = \left| \{(k, l) \,\middle|\, |p_{kl}^{(i)}| = j + 1/2\} \right|$, $j \in \{0, 1, \ldots, T - 1\}$, $i \in \{1, \ldots, \nu\}$ for linear residuals,
   - $\mathbf{h}_j^{(i)} = \left| \{(k, l) \,\middle|\, p_{kl}^{(i)} = j + 1/2\} \right|$, $j \in \{-T, \ldots, T - 1\}$, $i \in \{1, \ldots, \nu\}$ for non-linear residuals.

Notice that linear residuals are represented using only T bins because the last bin with centroid at $T + 1/2$ is uniquely determined by the other bins (the sum $\sum_{j \in \mathcal{Q}} \mathbf{h}_j^{(i)} = n_1 n_2$). Non-linear residuals are represented using only $2T$ bins because, in computing $\mathbf{h}_j^{(i)}$, we skip both the first and the last value corresponding to centroids $-T - 1/2$ and $T + 1/2$.

Also, symmetries of natural images can used to make the histograms better populated. Depending on the residual and the projection matrix $\mathbf{\Pi}^{(i)}$, the PSRMQ3 utilizes up to eight symmetries (rotation by multiples of 90 degrees, mirroring, etc.) for each random random matrix $\mathbf{\Pi}^{(i)}$.

The standard parameter setup for the PSRMQ3 is as follows. The number of projections per residual is $\nu = 55$, the maximum projection matrix size $s = 8$, the quantization step $q = 3$ (hence the name PSRMQ3), and the histogram threshold $T = 3$. This setup gives the PSRM the dimensionality of $12,870$.

---

*The Frobenius norm of matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_F = \sqrt{\operatorname{trace}(\mathbf{A}^\mathsf{T} \mathbf{A})}$.

3. CONSTRUCTING PHARM (PHASE-AWARE PROJECTION MODEL)

In this section, we provide a complete description of the PHARM. We start with a small motivational experiment to show the promise of phase-aware features, then we describe the PHARM in its full detail, and close the section with experiments aimed at determining the parameters of PHARM. All experiments in this section are carried out on the BOSSbase 1.01 image database,[1] compressed with JPEG quality factor 75 using Matlab's 'dct2' command, which corresponds to the 'slow' implementation of the DCT. As a test bed for setting the parameters, we generated stego images with J-UNIWARD[11] at 0.4 bits per non-zero AC DCT coefficient (bpnzac). The detectors were implemented as binary classifiers using the FLD ensemble classifier with default settings.[15] The detection accuracy is reported using the average of 10 OOB (out-of-bag) error estimates of the minimal total detection error $\overline{P}_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} (P_{\mathrm{FA}} + P_{\mathrm{MD}})/2$ obtained by running the ensemble with 10 different random seeds. The symbols P_{FA} and P_{MD} stand for the probability of false alarm and missed detection. The OOB error is known to be an unbiased estimate of the testing error.

3.1 Description

As already pointed out in the introduction, it is important to realize that in a decompressed JPEG (we remind the reader that we do not round the decompressed pixel values to interegers) the statistical properties of pixels are not spatially invariant – they depend on the position of the pixel and its neighborhood within the JPEG 8×8 pixel grid. Consequently, the histograms computed from all values of the projections $\mathbf{P}^{(i)}$, $i = 1, \ldots, \nu$, suffer from a lowered detection performance because in $\mathbf{P}^{(i)}$ we merge signals with different statistical properties. A more accurate detection can be obtained by computing the histograms from a properly subsampled $\mathbf{P}^{(i)}$ that takes into account the blocks of JPEG compression.

For every $i \in \{1, \ldots, \nu\}$, we select $u, v \in \{0, \ldots, 7\}$ uniformly randomly and subsample the projection values $\mathbf{P}^{(i)}$ to $\mathbf{P}^{(i;u,v)} \triangleq \left(p_{u+8 \cdot k, v+8 \cdot l}^{(i)} \right)$, $1 \leq k \leq n_1/8 - 1$, $1 \leq l \leq n_2/8 - 1$. The subsampled signal $\mathbf{P}^{(i;u,v)}$ is then quantized and captured using its histogram in the same manner as in the PSRM as explained in Steps 3 and 4 in Section 2.2. The phase pair (u, v) is always randomly chosen for each i.

Moreover, there are four symmetries that can be employed. Figure 1 shows how the symmetries can be utilized for a given residual $\mathbf{Z}$, projection matrix $\mathbf{\Pi}$, and phase pair (u, v). Merging histograms based on these symmetries improves the detection while preserving the feature dimensionality.

3.2 Single submodel

To obtain initial insight, we first test this approach on two submodels from Section 2.1 with the same parameter setup as in the PSRMQ3. The linear model is a merger of 'spam14h and 'spam14v' residuals obtained using (2) and (3). Because we have two linear residuals and $T = 3$, with ν projection matrices $\mathbf{\Pi}$, the dimensionality of this model is $2 \cdot 3 \cdot \nu$. Table 1 shows the detection error $\overline{E}_{\mathrm{OOB}}$ on J-UNIWARD at 0.4 bpnzac when the ν projections are used exactly as in the PSRMQ3 (see Section 2.2) and when they are used as described in Section 3.1 in the proposed phase-aware modification and its version that employs symmetries. The table shows the results for two values of ν.

For the concatenated spam14 residuals, the phase-aware versions offer a more accurate detection over the approach used in the PSRMQ3, especially when more projections per residual are utilized. As expected, the more robust statistic of the symmetrized features is more valuable for a smaller number of projections because symmetrization makes features better populated.

A similar experiment executed with the non-linear residual 'minmax41' revealed an interesting fact – the phase-awareness had worsened the performance. This can be easily explained by the nature of the 'minmax' features. The non-linear 'minmax' residuals take the maximum or minimum from multiple linear residuals that affect different neighborhoods. In a decompressed JPEG image, they thus aggregate values with very different statistics. Therefore, it makes sense that applying the proposed approach to already 'polluted' residuals does not bring any improvement; it only makes the statistic less robust. This also explains why the 'minmax' features are significantly worse than the 'spam14' features, even though they are both based on differences between neighboring pixels.
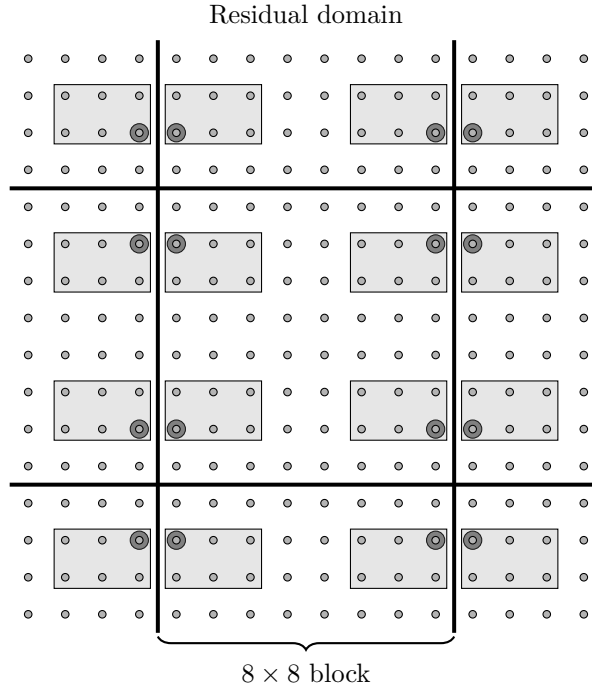
Residual domain

8×8 block

Figure 1. Four symmetries that can be utilized for a given 2×3 projection matrix $\mathbf{\Pi}$ and a given phase pair $(u, v) = (1, 0)$. The large dot represents the upper-left element of $\mathbf{\Pi}$ and shows how the matrix $\mathbf{\Pi}$ is flipped to preserve the symmetries.

3.3 Phase-Aware Projection Features (PHARM)

We now use the insight gained in the previous section to design a simple feature set to see what improvement can be obtained when splitting the projections according to their location w.r.t. the JPEG grid (phase). We only use linear ('spam' type) residuals with the following seven kernels $\mathbf{K}$:

$$
\begin{pmatrix} -1 & 1 \end{pmatrix} \quad \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -3 & 3 & -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ -3 \\ 3 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \quad \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.
$$

These kernels were obtained as the ones best complementing each other using a greedy forward feature-selection algorithm using the out-of-bag estimate of the detection error from the 25 prediction kernels listed below. The first, second, and third order kernels listed below use the same predictors as the ones used in the SRM oriented in the corresponding directions.[3]

- 1×2 first-order horizontal, vertical, major and and minor diagonal (4 predictors)

- 1×3 second-order horizontal, vertical, major and minor diagonal (4 predictors)

- 1×4 third-order horizontal, vertical, major and minor diagonal (4 predictors)

- 2×2 horizontal $[-1, 1; -1, 1]$, vertical $[-1, -1; 1, 1]$, and diagonal $[-1, 1; 1, -1]$ (notice that in this case diagonal and minor diagonal coincide) (3 predictors)

- 3×3 (KB) kernel[13] and its 4 EDGE (2×3) versions as in the SRM (5 predictors)

- 5×5 kernel used in SQUARE submodel in SRM[3] and its 4 EDGE (3×5) versions (5 predictors).

For a fair comparison with other steganalytic algorithms, we optimized this projection feature set with respect to all its parameters using J-UNIWARD at 0.4 bpnzac. Figure 2 shows how the detection error $\overline{E}_{\mathrm{OOB}}$ depends on the individual parameters s, T, and q when all other parameters are set to the following default values: $s = 8$, $T = 2$, $q = 5$, and $\nu = 900$.

Feature type	'spam14h' & 'spam14v'		'minmax41'	
	$\nu = 110$ dim 660	$\nu = 1000$ dim 6000	$\nu = 110$ dim 660	$\nu = 1000$ dim 6000
Standard	0.2587	0.2034	0.3054	0.2257
Phase-aware	0.2576	0.1536	0.3323	0.2421
Phase-aware symmetrized	0.2292	0.1582	0.3292	0.2409

Table 1. Detection error $\overline{E}_{\mathrm{OOB}}$ on J-UNIWARD at 0.4 bpnzac for the PSRMQ3 and two phase-aware versions with the concatenation of the residuals 'spam14h' and 'spam14v' and the 'minmax41' residual.

We selected the maximum neighborhood size $s = 8$ because it corresponds to the size of the JPEG blocks. The threshold is set $T = 2$ because it offers both the best detection rate and a small dimensionality. The optimal quantization step is significantly influenced by the JPEG quality factor (as in Ref. [10]). The best values were $q = 5$ for quality factor 75 and $q = 2$ for quality factor 95. To obtain a quantization step for a general quality factor $QF \leq 100$, we recommend using the formula $q = \frac{65}{4} - \frac{3}{20}QF$ obtained using linear interpolation.

Figure 3 shows the detection error as a function of the number of projections per residual ν. Although the detection rate decreases with increasing ν, it also increases the feature dimensionality and computational complexity. For the PHARM feature set, we decided to use $\nu = 900$ as a good compromise between detection accuracy and complexity, which gives PHARM the dimensionality of $7 \cdot T \cdot \nu = 7 \cdot 2 \cdot 900 = 12,600$. This allows us to directly compare the performance of PHARM with SRMQ1 $(12,753)$ and PSRMQ3 $(12,870)$ due to their similar dimensionality.

The time needed to extract the PHARM feature vector for one BOSSbase image is shown in Table 2. The phase-aware features are faster to compute than the PSRM features because they require only four dot products between the projection matrix and the corresponding segment of the noise residual per projection (see Figure 1)
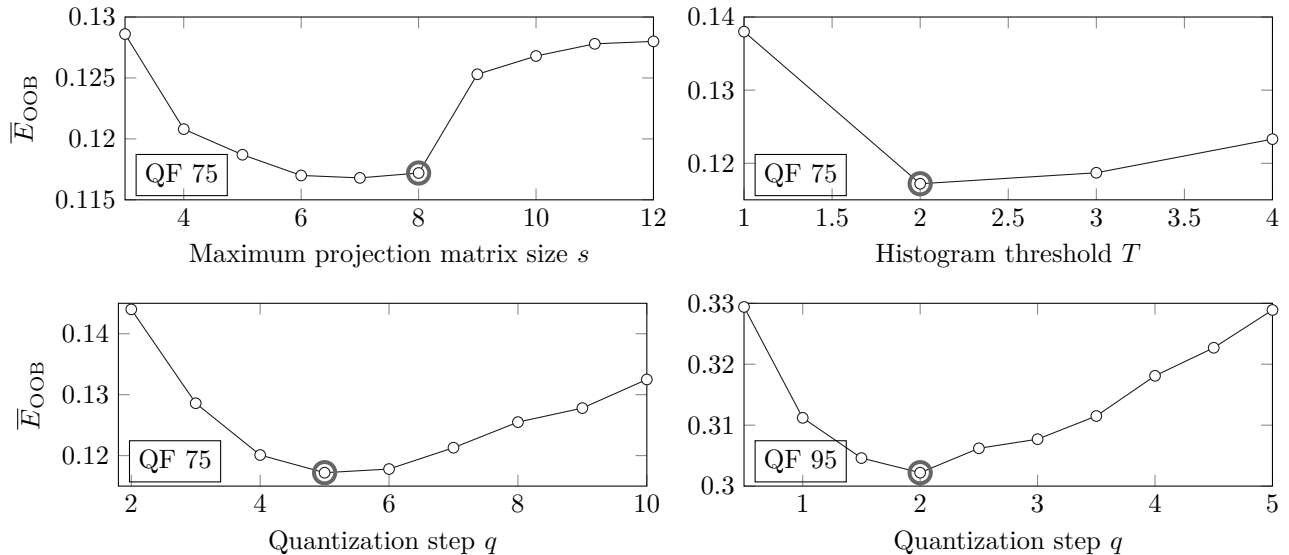

Figure 2. Detection error $\overline{E}_{\mathrm{OOB}}$ on J-UNIWARD 0.4 bpnzac as a function of the maximum projection matrix size s (top left), the histogram threshold T (top right), and the quantization step q for JPEG quality factor 75 (bottom left) and 95 (bottom right). The values of all parameters except the investigated one are fixed at $s = 8$, $T = 2$, $q = 5$, and $\nu = 900$.
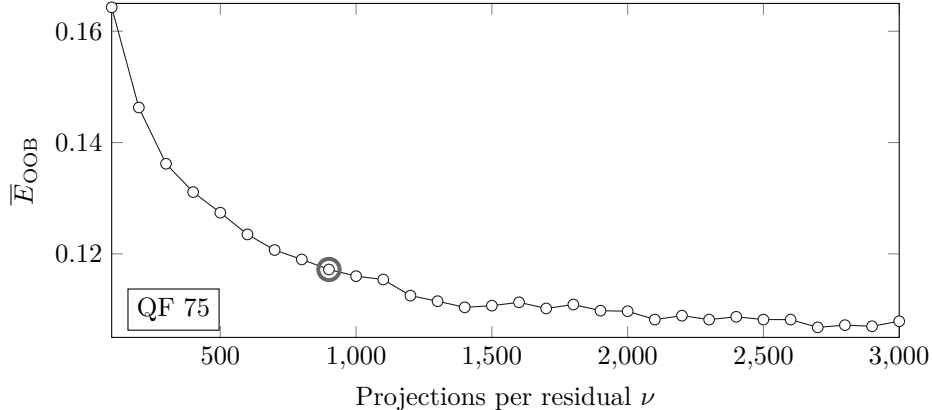
Figure 3. Detection error $\overline{E}_{\mathrm{OOB}}$ on J-UNIWARD at 0.4 bpnzac for JPEG quality factor 75 as a function of the number of projections per residual. The feature dimensionality is $14 \cdot \nu$. The selected value $\nu = 900$ is highlighted with a circle.

Feature set	PHARM	DCTR	JRM	SRMQ1	PSRMQ3
Dimensionality	12,600	8,000	22,510	12,753	12,870
Extraction time (s)	4.2	0.6	4.5	1.3	640

Table 2. Dimensionality and extraction time in seconds for one BOSSbase image for PHARM in comparison to other features. All feature extractors except JRM were implemented in C++ / MEX and run on an Intel i7 2 GHz laptop.

per 8×8 JPEG block instead of the 512 local convolutions (64 elements in JPEG block times 8 symmetries per projection) in the case of the PSRMQ3.

4. EXPERIMENTS AND COMPARISON WITH PRIOR ART

In this section, we evaluate the detection performance of PHARM w.r.t. other rich models when building detectors with the FLD ensemble for each tested payload. The tests are carried out on BOSSbase 1.01 images compressed with quality factors 75 and 95. The tested steganographic algorithms include mostly modern embedding schemes, such as J-UNIWARD, its side-informed version called SI-UNIWARD,[11] and the ternary implementation of UED.[6] We also included one example of an old stego scheme nsF5.[4] All payloads are relative and expressed in bits per non-zero AC coefficient (bpnzac). For the side-informed scheme SI-UNIWARD, the payload is expressed w.r.t. DCT coefficients of the precover compressed with the specified quality factor. We note that the implementation of both UNIWARD schemes used the stabilizing constant as described in Ref. [11], which improves the security w.r.t. the original publication.[7]

All experimental results appear in Figures 4–7, which show the detection error $\overline{E}_{\mathrm{OOB}}$ versus payload for PHARM, DCTR,[10] the JPEG rich model JRM,[14] the spatial rich model SRMQ1,[3] the union of JRM and SRMQ1 called JPSRM, and the PSRMQ3.[8] Since the PSRMQ3 is very expensive to compute, we only include the detection error for one small and one large payload.

The experiments can be summarized as follows. For J-UNIWARD and UED, PHARM achieves a markedly better performance than any other feature set irrespectively of the quality factor and payload. Since nsF5 significantly changes the distribution of DCT coefficients, it is better detected using features formed directly from the DCT coefficients (JRM and JSRM) instead of features computed from the spatial domain (PHARM, SRMQ1, PSRMQ3, and DCTR). On SI-UNIWARD, PHARM offered a small improvement for quality factor 75 but not for 95. When interpreting the results, one needs to take into account also the dimensionality of the feature sets and the time needed for their extraction.

It is also interesting that PHARM is better than SRMQ1 in all tested cases. This is despite the much larger diversity of residuals in SRMQ1. This confirms the observation made in this paper that considering the phase of the residuals is very important and that residuals obtained using small-support kernels work better than those with a larger support.
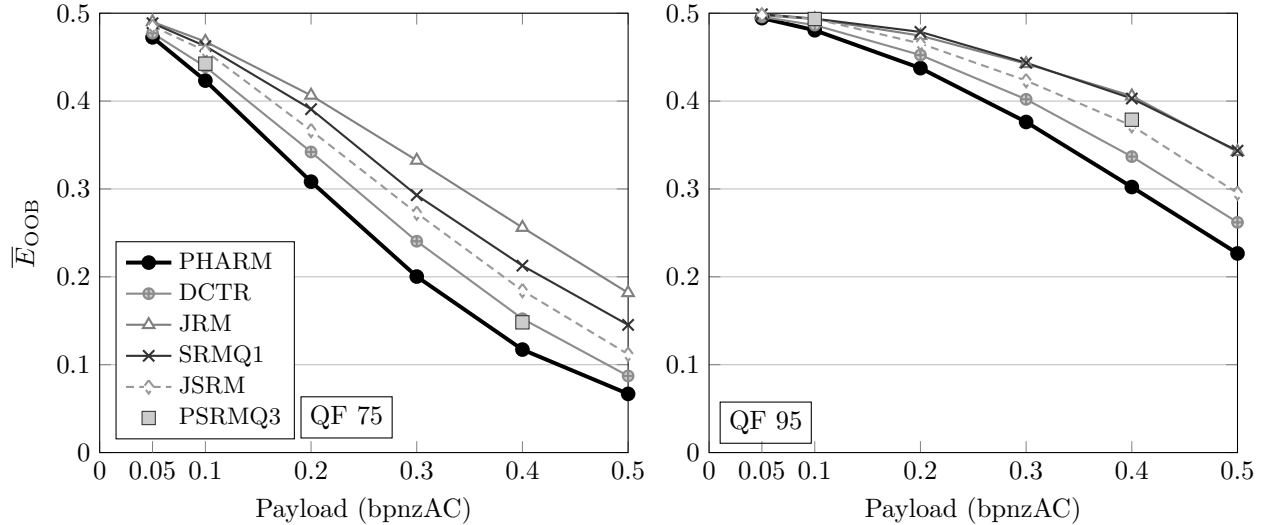
Figure 4. Detection error $\overline{E}_{\mathrm{OOB}}$ for J-UNIWARD for quality factors 75 and 95 when steganalyzed with the proposed PHARM and other rich feature sets.

5. CONCLUSION

Modern embedding algorithms for JPEG images, such as J-UNIWARD and UED, are currently better detected by decompressing the JPEG image into the spatial domain and steganalyzing using features formed by various descriptors of noise residuals obtained using pixel predictors. The main contribution of this paper is the key insight that in a decompressed JPEG image the pixels are not shift invariant and their statistical properties depend on their position or "phase" with respect to the 8×8 grid. Taking this into account, we propose to compute the statistics only from subsets of the residual determined by the underlying JPEG grid. Moreover, and in agreement with the above observation, smaller prediction kernels offer better performance than larger kernels. Thus, instead of diversifying the model over prediction kernels, as is commonly done in rich models, we diversify through projections in a similar manner as in the projection rich model, PSRM. In contrast to the "phase-unaware" PSRMQ3, the proposed feature set PHARM (PHase-Aware pRojection Model) is much more computationally efficient and also offers better detection accuracy of modern JPEG steganographic methods.

In the experimental part of this paper, on images embedded with J-UNIWARD we determine the parameters of PHARM, which include the number of projections per residual, the maximal size of the projection neighborhood, the quantization step, and the residual truncation threshold. Then, we subject the novel PHARM to tests on four JPEG steganographic algorithms on the standard image database BOSSbase 1.01 compressed with quality factors 75 and 95. The results clearly show the benefit of making the features phase-aware as the PHARM features outperform not only the spatial rich model (SRMQ1) but also the much more complex PSRMQ3. PHARM is also the best performer on all modern JPEG steganographic methods that we tested.

The code for the PHARM feature extractor is available from `http://dde.binghamton.edu/download/feature_extractors/`.
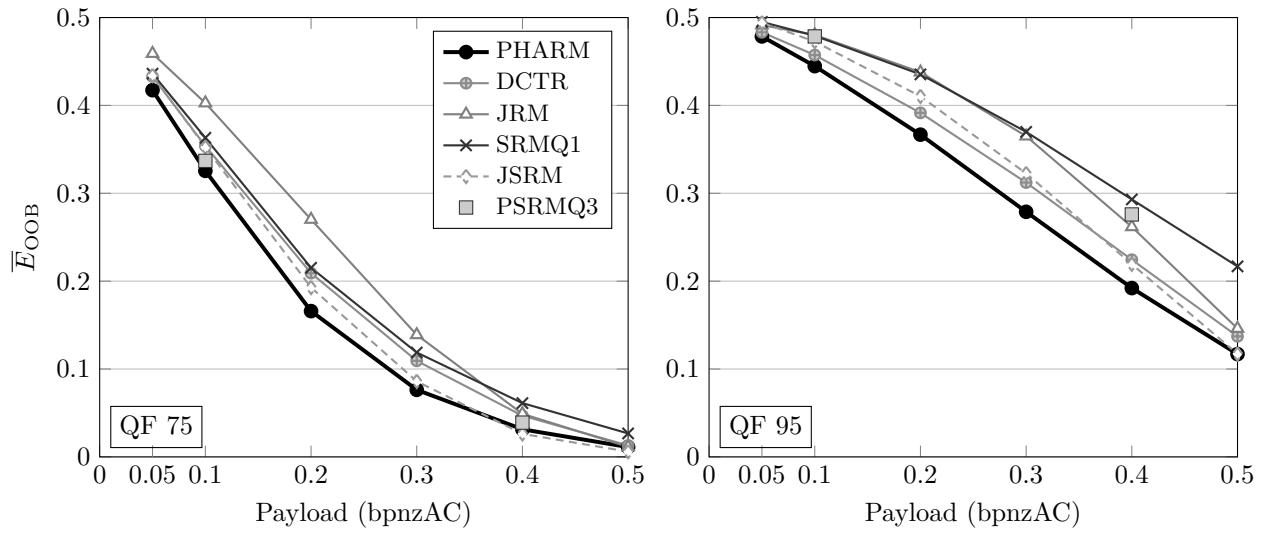
6. ACKNOWLEDGMENTS

Figure 5. Detection error $\overline{E}_{\mathrm{OOB}}$ for UED with ternary embedding for quality factors 75 and 95 when steganalyzed with the proposed PHARM and other rich feature sets.

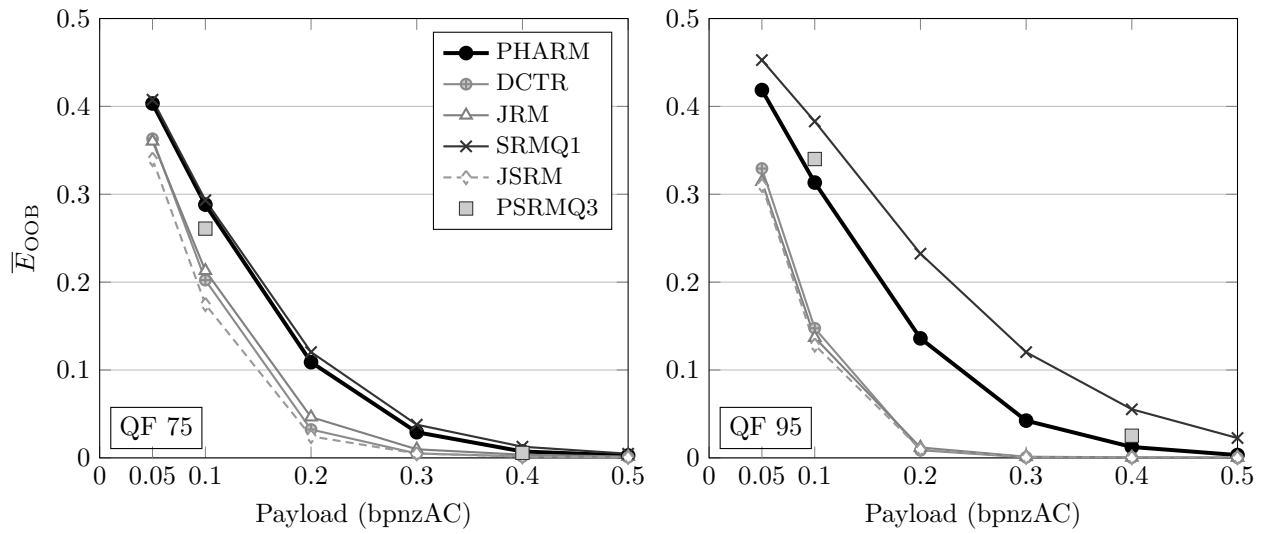Figure 6. Detection error $\overline{E}_{\mathrm{OOB}}$ for nsF5 for quality factors 75 and 95 when steganalyzed with the proposed PHARM and other rich feature sets.
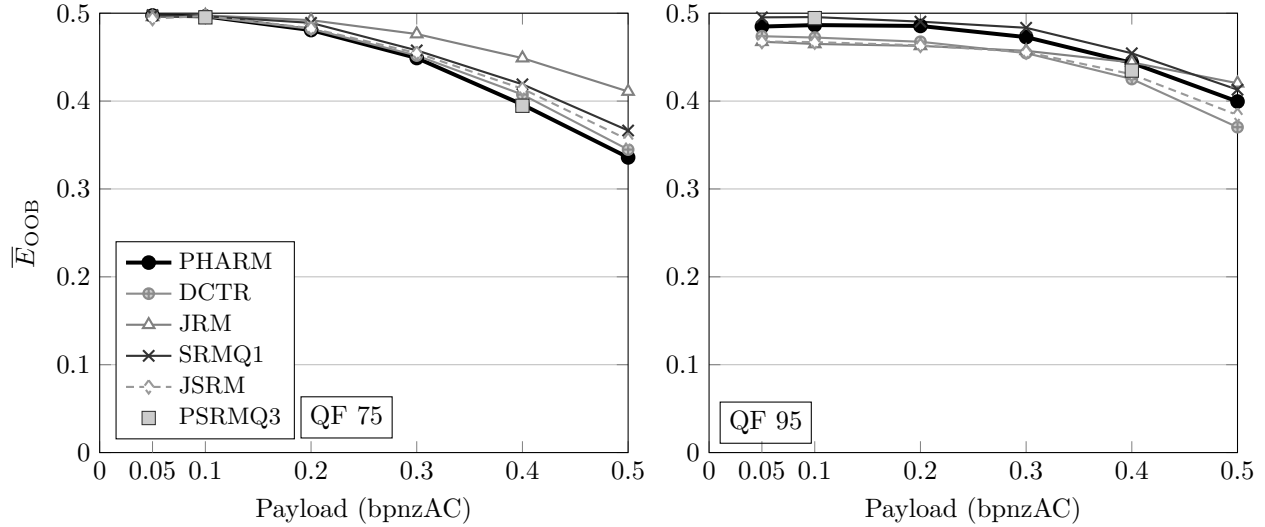
Figure 7. Detection error $\overline{E}_{\mathrm{OOB}}$ for side-informed SI-UNIWARD for quality factors 75 and 95 when steganalyzed with the proposed PHARM and other rich feature sets.

REFERENCES

1. P. Bas, T. Filler, and T. Pevný. Break our steganographic system – the ins and outs of organizing BOSS. In T. Filler, T. Pevný, A. Ker, and S. Craver, editors, *Information Hiding, 13th International Conference*, volume 6958 of Lecture Notes in Computer Science, pages 59–70, Prague, Czech Republic, May 18–20, 2011.

2. C. Chen and Y. Q. Shi. JPEG image steganalysis utilizing both intrablock and interblock correlations. In *Circuits and Systems, ISCAS 2008. IEEE International Symposium on*, pages 3029–3032, Seattle, WA, May, 18–21, 2008.

3. J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, June 2011.

4. J. Fridrich, T. Pevný, and J. Kodovský. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In J. Dittmann and J. Fridrich, editors, *Proceedings of the 9th ACM Multimedia & Security Workshop*, pages 3–14, Dallas, TX, September 20–21, 2007.

5. L. Guo, J. Ni, and Y.-Q. Shi. An efficient JPEG steganographic scheme using uniform embedding. In *Fourth IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5, 2012.

6. L. Guo, J. Ni, and Y.-Q. Shi. Uniform embedding for efficient JPEG steganography. *IEEE Transactions on Information Forensics and Security*, 9(5):814–825, May 2014.

7. V. Holub and J. Fridrich. Digital image steganography using universal distortion. In W. Puech, M. Chaumont, J. Dittmann, and P. Campisi, editors, *1st ACM IH&MMSec. Workshop*, Montpellier, France, June 17–19, 2013.

8. V. Holub and J. Fridrich. Random projections of residuals for digital image steganalysis. *IEEE Transactions on Information Forensics and Security*, 8(12):1996–2006, December 2013.

9. V. Holub and J. Fridrich. Challenging the doctrines of JPEG steganography. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014*, volume 9028, pages 02 1–8, San Francisco, CA, February 2–6, 2014.

10. V. Holub and J. Fridrich. Low complexity features for JPEG steganalysis using undecimated DCT. *IEEE Transactions on Information Forensics and Security*, 10(2):219–228, February 2015.

11. V. Holub, J. Fridrich, and T. Denemark. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security (Revised Selected Papers of 1st ACM IH&MMS Workshop)*, 2014:1–13.

12. F. Huang, J. Huang, and Y.-Q. Shi. New channel selection rule for JPEG steganography. *IEEE Transactions on Information Forensics and Security*, 7(4):1181–1191, August 2012.

13. A. D. Ker and R. Böhme. Revisiting weighted stego-image steganalysis. In E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, editors, *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages 5 1–17, San Jose, CA, January 27–31, 2008.

14. J. Kodovský and J. Fridrich. Steganalysis of JPEG images using rich models. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages 0A 1–13, San Francisco, CA, January 23–26, 2012.

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

16. Q. Liu. Steganalysis of DCT-embedding based adaptive steganography and YASS. In J. Dittmann, S. Craver, and C. Heitzenrater, editors, *Proceedings of the 13th ACM Multimedia & Security Workshop*, pages 77–86, Niagara Falls, NY, September 29–30, 2011.

17. Y. Q. Shi, C. Chen, and W. Chen. A Markov process based approach to effective attacking JPEG steganography. In J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, editors, *Information Hiding, 8th International Workshop*, volume 4437 of Lecture Notes in Computer Science, pages 249–264, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.

18. T. Thai, R. Cogranne, and F. Retraint. Statistical model of quantized DCT coefficients: Application in the steganalysis of Jsteg algorithm. *Image Processing, IEEE Transactions on*, 23(5):1–14, May 2014.

19. T. Hai Thai and R. Cogranne. Optimal detection of OutGuess using an accurate model of DCT coefficients. In *Sixth IEEE International Workshop on Information Forensics and Security*, Atlanta, GA, December 3–5 2014.

20. 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.

21. A. Westfeld. High capacity despite better steganalysis (F5 – a steganographic algorithm). In I. S. Moskowitz, editor, *Information Hiding, 4th International Workshop*, volume 2137 of Lecture Notes in Computer Science, pages 289–302, Pittsburgh, PA, April 25–27, 2001. Springer-Verlag, New York.

22. C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. Hidden information detection based on quantized Laplacian distribution. In *Proc. IEEE ICASSP*, Kyoto, Japan, March 25-30, 2012.