Random Projections of Residuals for Digital Image Steganalysis

Vojtech Holub and Jessica Fridrich, *Member, IEEE*

***Abstract*— The traditional way to represent digital images for feature based steganalysis is to compute a noise residual from the image using a pixel predictor and then form the feature as a sample joint probability distribution of neighboring quantized residual samples—the so-called co-occurrence matrix. In this paper, we propose an alternative statistical representation—instead of forming the co-occurrence matrix, we project neighboring residual samples onto a set of random vectors and take the first-order statistic (histogram) of the projections as the feature. When multiple residuals are used, this representation is called the projection spatial rich model (PSRM). On selected modern steganographic algorithms embedding in the spatial, JPEG, and side-informed JPEG domains, we demonstrate that the PSRM can achieve a more accurate detection as well as a substantially improved performance versus dimensionality trade-off than state-of-the-art feature sets.**

***Index Terms*— Image, steganalysis, random projection, residual.**

I. Introduction

STEGANALYSIS is the art of revealing the presence of secret messages embedded in objects. We focus on the case when the original (cover) object is a digital image and the steganographer hides the message by slightly modifying the numerical representation of the cover – either the pixel colors or the values of transform coefficients.

In general, a steganalysis detector can be built either using the tools of statistical signal detection or by applying a machine-learning approach. Both approaches have their strengths as well as limitations, which is the reason why they are both useful and will likely coexist in the foreseeable future. The former approach derives the detector from a statistical model of the cover source, allowing one to obtain error bounds on the detector performance. Normalized detection statistics are also less sensitive to differences between cover sources. On the other hand, to make this approach tractable, the adopted cover model must usually be sufficiently simple, which limits the detector optimality and the validity of the error bounds to the chosen cover model. Simple models, however, cannot capture all the complex relationships among individual image elements that exist in images of natural scenes acquired using imaging sensors. Moreover, this approach has so far been applied only to rather simple embedding operations, examples of which are the LSB (least significant bit) replacement and matching [5], [6], [8], [38], and may not be easily adapted to complex, content-adaptive embedding algorithms, such as HUGO [34], WOW [18], or the schemes based on UNIWARD [19]. This is because attacking these schemes would require working with models that allow for complex dependencies among neighboring pixels. However, given the highly non-stationary character of natural images, estimating such local model parameters will likely be infeasible.

The latter approach to steganalysis does not need the underlying cover distribution to build a detector. Instead, the task of distinguishing cover and stego objects is formulated as a classification problem. First, the image is represented using a feature vector, which can be viewed as a heuristic dimensionality reduction. Then, a database of cover and the corresponding stego images is used to build the detector using standard machine learning tools. The principal advantage of this approach is that one can easily construct detectors for arbitrary embedding algorithms. Also, for a known cover source, such detectors usually perform substantially better than detectors derived from simple cover models. The disadvantage is that the error bounds can only be established empirically, for which one needs sufficiently many examples from the cover source. While such detectors may be inaccurate when analyzing a single image of unknown origin, steganographic communication is by nature repetitive and it is not unreasonable to assume that the steganalyst has many examples from the cover source and observes the steganographic channel for a length of time.

In this paper, we assume that the analyst knows the steganographic algorithm and sufficiently many examples from the cover source are available. Since the embedding changes can be viewed as an additive low-amplitude noise that may be adaptive to the host image content, we follow a long-established paradigm [11], [16], [33], [39] and represent the image using a feature computed from the image noise component– the so-called noise residual.[1] To obtain a more accurate detection of content-adaptive steganography, various authors have proposed to utilize an entire family of noise residuals, obtaining thus what is now called rich image representations [11], [13], [16].

Traditionally, noise residuals were represented using either sample joint or conditional probability distributions of adjacent

Manuscript received May 14, 2013; revised August 13, 2013 and October 14, 2013; accepted October 15, 2013. Date of publication October 21, 2013; date of current version November 11, 2013. This work was supported by the Air Force Office of Scientific Research under Grant FA9950-12-1-0124. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick Bas.

The authors are with the Department of Electrical and Computer Engineering, Binghamton University, NY 13902 USA (e-mail: vholub1@binghamton.edu; fridrich@binghamton.edu).

Digital Object Identifier 10.1109/TIFS.2013.2286682

[1] The idea to compute features from noise residuals has already appeared in the early works on feature based steganalysis [1], [7], [15], [31].

1556-6013 © 2013 IEEE

quantized and truncated residual samples (co-occurrence matrices) [11], [16], [33], [39]. Higher-order co-occurrences detect steganographic changes better as they can capture dependencies across multiple pixels. Since the co-occurrence dimensionality increases exponentially with its order, the co-occurrence order one can use in practice is limited by the total number of pixels, and steganalysts had to quantize and truncate the residual (sometimes quite harshly) to obtain a reasonably low-dimensional and statistically significant descriptor for subsequent machine learning [11], [16], [33].

In this article, we propose an alternative statistical descriptor for noise residuals. Instead of forming co-occurrences of neighboring quantized residual samples, we use the unquantized values and project them on random directions, which are subsequently quantized and represented using histograms as steganalytic features. This brings several advantages over the representation based on co-occurrences. First, by using large projection neighborhoods one can potentially capture dependencies among a large number of pixels. Second, by selecting random neighborhood sizes, the statistical description can be further diversified, which improves the detection accuracy. Third, since more features will be statistically significant in comparison to high-dimensional co-occurrences where numerous boundary bins may be underpopulated, projections enjoy a much more favorable feature dimensionality vs. detection accuracy trade-off. Fourth, a greater design flexibility is obtained since the size and shape of the projection neighborhoods, the number of projection vectors, as well as the histogram bins can be incrementally adjusted to achieve a desired trade-off between detection accuracy and feature dimensionality. Finally, the novel feature representation appears to be universally effective for detection of modern steganographic schemes embedding in both the spatial and JPEG domains.

This work has evolved from an initial study by the same authors [20]. blackAmong the many differences and improvements between this prior art and the current manuscript, we name the following. The hand design of the projection neighborhoods and projection vectors was replaced with a fully randomized construction driven by a single parameter. We also investigate the effect of the quantizer design (bin width and the number of quantizer centroids) for detection in both the spatial and JPEG domains. Finally, the experiments were substantially enlarged and cover three different embedding domains for two cover sources and state-of-the-art steganographic methods in each domain.

In the next section, we introduce the common core of all experiments in this paper and a list of tested steganographic methods. Section III contains a brief description of the SRM (spatial rich model) [11] and the elements from which it is built. The same residuals are used to construct the PSRM (projection spatial rich model) proposed in Section IV. This section also contains several investigative experiments used to set the PSRM parameters. In Section V, we compare the detection performance of the proposed PSRM with the current state-of-the-art feature descriptors – the SRM and the JRM (JPEG rich model) proposed in [28]. The comparison is carried out on selected modern (and currently most secure) steganographic algorithms operating in the spatial, JPEG, and side-informed JPEG domains. The paper is concluded in Section VI.

High-dimensional arrays, matrices, and vectors will be typeset in boldface and their individual elements with the corresponding lower-case letters in italics. The calligraphic font is reserved for sets. For a random variable X, its expected value is denoted as $E[X]$. 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 $n = n_1 \times n_2$ pixels. For a set of L centroids, $\mathcal{Q} = \{q_1, \dots, q_L\}$, $q_1 \leq \dots \leq q_L$, a scalar quantizer is defined as $Q_{\mathcal{Q}}(x) \triangleq \arg\min_{q \in \mathcal{Q}} |x - q|$.

II. Preliminaries

A. Common Core of all Experiments

In this paper, we carry out experiments on two image sources. The first is the standardized database called BOSSbase 1.01 [2]. This source contains 10,000 images acquired by seven digital cameras in RAW format (CR2 or DNG) and subsequently processed by converting to 8-bit grayscale, resizing, and cropping to the size of 512×512 pixels. The script for this processing is also available from the BOSS competition web site.

The second image source was obtained using the Leica M9 camera equipped with an 18-megapixel full-frame sensor. A total of 3,000 images were acquired in the raw DNG format, demosaicked using UFRaw (with the same settings as the script used for creating BOSSbase), converted to 8-bit grayscale, and finally central-cropped to the size of 512×512. This second source is very different from BOSSbase 1.01 and was intentionally included as an example of imagery that has not been subjected to resizing, which has been shown to have a substantial effect on the detectability of embedding changes in the spatial domain [29]. By adjusting the image size of Leica images to that of the BOSSbase, we removed the effect of the square root law [25] on steganalysis, allowing interpretations of experiments on both sources in Section V.

For JPEG experiments, the databases were JPEG-compressed with standard quantization tables corresponding to quality factors 75 and 95. The JPEG format allows several different implementations of the DCT transform, DCT(.). The implementation may especially impact the security of side-informed JPEG steganography, in which the sender has the uncompressed (precover[2]) image and hides data while subjecting it to JPEG compression [10], [17], [19], [26], [35]. In this paper, we work with the DCT(.) implemented as 'dct2' in Matlab when feeding in pixels represented as 'double'. To obtain an actual JPEG image from a two-dimensional array of quantized coefficients $\mathbf{X}$ (cover) or $\mathbf{Y}$ (stego), we first create an (arbitrary) JPEG image of the same dimensions $n_1 \times n_2$ using Matlab's 'imwrite' with the same quality factor, read its JPEG structure using Sallee's Matlab JPEG Toolbox[3] and then merely replace the array of quantized coefficients in this structure with $\mathbf{X}$ and $\mathbf{Y}$ to obtain the cover and stego images, respectively.

[2]The concept of precover is due to Ker [22].

[3]http://dde.binghamton.edu/download/jpeg_toolbox.zip

The classifiers we use are all instances of the ensemble proposed in [30] and available from. They employ Fisher linear discriminants as base learners trained on random subspaces of the feature space. The ensemble is run in its default form in which the random subspace dimensionality and the number of base learners is determined automatically as described in the original publication [30]. We report the detection performance using the out-of-bag (OOB) estimate of the testing error. This error, which we denote E_{OOB}, is known to be an unbiased estimate of the testing error on unseen data [4]. It is computed by training on a subset of the database obtained by bootstrapping and testing on the remaining part that was unused for training. The unique images forming the training set span approximately two thirds of the database, while the testing error is estimated from the remaining unused third. We train a separate classifier for each combination of image source, embedding method, and payload. Even though the knowledge of the payload does not correspond to Kerckhoffs' principle, this testing is customary in research articles on steganography and steganalysis to inform the reader about how the security changes with payload.

B. Steganographic Algorithms

To evaluate the performance of the proposed projection rich model, we compare it against state-of-the-art rich feature sets on steganographic algorithms that represent the most secure algorithms for three embedding domains. All steganographic algorithms considered in this paper embed a given payload while minimizing a distortion function. We use embedding simulators that simulate embedding changes on the rate–distortion bound [9]. A practical data hiding algorithm would be embedding using a slightly increased distortion, e.g., using the syndrome-trellis codes (STCs) [9].

In the spatial domain, we use HUGO [34], the first content-adaptive algorithm that incorporated the STCs, WOW with its wavelet-based distortion [18], and S-UNIWARD [19], which can be thought of as a highly adaptive and simplified modification of WOW.

JPEG domain algorithms include the nsF5 [14], a modification of the original F5 algorithm [37], the Uniform Embedding Distortion (UED) algorithm [17], and J-UNIWARD [19].

We also include a comparison on steganographic algorithms embedding in the JPEG domain with "side-information" in the form of the uncompressed cover image. Such algorithms utilize the rounding errors of DCT coefficients to achieve a better security. We study two state-of-the-art side-informed algorithms – the Normalized Perturbed Quantization (NPQ) [21] and SI-UNIWARD [19]. The NPQ was chosen over older versions of the Perturbed Quantization algorithm [14] based on the superiority of NPQ over PQ reported in [21]. Both algorithms are modified so they avoid embedding in DCT modes $(0,0)$, $(0,4)$, $(4,0)$ and $(4,4)$ when the unquantized value is equal to $k+0.5$, $k \in \mathbb{Z}$. The reason for this modification can also be found in [19].

III. Spatial Rich Model

The statistical descriptor (feature vector) proposed in this article uses the same family of noise residuals as the SRM [11]. However, their statistical description in the proposed PSRM is different – instead of forming co-occurrences of quantized residuals, we project the unquantized residuals onto random directions and use the first-order statistics of the projections as features (see Section IV for details). To make this paper self-contained and to better contrast the differences between SRM and the proposed PSRM, we briefly describe the SRM residual family as well as the SRM feature vector while focusing on the conceptual part without going into details, which can be found in the original publication.

A. Noise Residuals

Each residual is tied to a pixel predictor, $\hat{x}_{ij}$, which is a mapping that assigns an estimate of the cover pixel x_{ij} as a function of pixel values from its immediate neighborhood, $\mathcal{N}(\mathbf{Y}, i, j)$, in the stego image $\mathbf{Y}$. The noise residual corresponding to this predictor is a matrix $\mathbf{Z} \in \mathbb{R}^{n_1 \times n_2}$ with elements

$$z_{ij} = \hat{x}_{ij}(\mathcal{N}(\mathbf{Y}, i, j)) - y_{ij}, \quad 1 \le i \le n_1, \quad 1 \le j \le n_2. \tag{1}$$

The SRM residuals are computed using 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 $\mathbf{K}$:

$$\mathbf{Z} = \mathbf{K} * \mathbf{Y} - \mathbf{Y}, \tag{2}$$

where the symbol $'*'$ denotes the convolution.

For example, the kernel

$$\mathbf{K}_3 = \tfrac{1}{4}\begin{pmatrix} -1 & 2 & -1 \\ 2 & 0 & 2 \\ -1 & 2 & -1 \end{pmatrix}, \tag{3}$$

which was originally proposed in [24] and theoretically justified in [3], estimates the value of the central pixel from its local 3×3 neighborhood. In contrast, the kernel

$$\mathbf{K}_3' = \frac{1}{4}\begin{pmatrix} -1 & 2 & -1 \\ 2 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4}$$

uses only a portion of the same 3×3 neighborhood and may return a better prediction in the presence of a horizontal edge going through the central pixel. The predictor with kernel $\mathbf{K}_3$ is non-directional because it does not prefer pixels from a certain direction (the kernel matrix is symmetrical). The predictor that utilizes $\mathbf{K}_3'$ is directional as its output depends only on six pixel values in the upper half of the 3×3 neighborhood.

There are numerous other linear predictors used in the SRM. Most are derived by assuming that the image content locally follows a polynomial model. For example, the pixel predictors

$$\hat{x}_{ij} = y_{i,j+1}, \tag{5}$$

$$\hat{x}_{ij} = (y_{i,j-1} + y_{i,j+1})/2, \tag{6}$$

$$\hat{x}_{ij} = (y_{i,j-1} + 3y_{i,j+1} - y_{i,j+2})/3, \tag{7}$$

are based on the assumption that image content is locally constant, linear, and quadratic, respectively. Note that the residuals computed using these three predictors are all directional as they only utilize horizontally adjacent neighbors

of y_{ij}. The vertical form of these residuals that uses only vertically adjacent pixels is obtained by simply swapping the subscripts in (5)–(7). In general, the kernel for the vertical predictor is a transpose of the one for the horizontal direction.

All non-linear predictors in the SRM are obtained by taking the minimum (maximum) of the output of two or more residuals obtained using linear predictors. For example, given a horizontal residual $Z^{(\mathrm{h})}$ and a vertical residual $Z^{(\mathrm{v})}$, the non-linear residuals (residuals computed using a non-linear predictor) are computed as:

$$z_{ij}^{(\min)} = \min\{z_{ij}^{(\mathrm{h})}, z_{ij}^{(\mathrm{v})}\}, \quad \forall i, j \tag{8}$$

$$z_{ij}^{(\max)} = \max\{z_{ij}^{(\mathrm{h})}, z_{ij}^{(\mathrm{v})}\}, \quad \forall i, j. \tag{9}$$

B. Quantization

The next step in forming the SRM constitutes quantizing $\mathbf{Z}$ to a set of centroids $\mathcal{Q} = \{-Tq, (-T+1)q, \ldots, Tq\}$, where $T > 0$ is an integer threshold and $q > 0$ is a quantization step:

$$r_{ij} \triangleq Q_{\mathcal{Q}}(z_{ij}), \quad \forall i, j. \tag{10}$$

C. Co-Occurrence Matrices and Submodels

The next step in forming the SRM feature vector involves computing a co-occurrence matrix of Dth order from D (horizontally and vertically) neighboring values of the quantized residual r_{ij} (10) from the entire image. As argued in the original publication [11], diagonally neighboring values are not included due to much weaker dependencies among residual samples in diagonal directions. To keep the co-occurrence bins well-populated and thus statistically significant, the authors of the SRM used small values for D and T: $D = 4$, $T = 2$, and $q \in \{1, 1.5, 2\}$. Finally, symmetries of natural images are leveraged to further marginalize the co-occurrence matrix to decrease the feature dimension and better populate the SRM feature vector (see Section II.C of [11]).

Note that non-linear residuals are represented using two co-occurrence matrices, one for $\mathbf{Z}^{(\min)}$ and one for $\mathbf{Z}^{(\max)}$, while linear residuals require a single co-occurrence matrix. The authors of the SRM combined the co-occurrences of two linear residuals into one "submodel" to give them after symmetrization approximately the same dimensionality as the union of co-occurrences from min / max non-linear residuals. Figure 3 in [11] illustrates the details of this procedure. This allowed a fair comparison of detection performance of individual submodels. The authors also used a simple forward feature selection on submodels to improve the dimensionality vs. detection accuracy trade-off. There are a total of 39 submodels in the SRM.

The predictors and residuals used in the proposed PSRM are the same as those used in the SRM – a complete list of predictors appears in Figure 2 of [11]. Everywhere in this article, we understand by SRM the full version of this model with all three quantization steps (its dimensionality is $34,671$). A scaled-down version of the SRM when only one quantization step, q, is used will be abbreviated as SRMQq. Its dimensionality is $12,753$.

IV. Projection Spatial Rich Model

In this section, we provide the reasoning behind the proposed projection spatial rich model and describe it in detail, including the experiments used to set the PSRM parameters.

A. Motivation

The residual is a realization of a two-dimensional random field whose statistical properties are closely tied to the image content (e.g., larger values occur near edges and in textures while smaller values are typical for smooth regions). Steganographic embedding changes modify the statistical properties of this random field. The steganalyst's task is to compute a test statistic from this random field that would detect the embedding changes as reliably as possible.

Traditionally, and as described in the previous section, the random field is first quantized and then characterized using a joint probability mass function (co-occurrence matrix) of D neighboring residual samples. The problem with this approach is the exponential growth of the co-occurrence size with its order D. With increasing D, a rapidly increasing number of co-occurrence bins become underpopulated, which worsens the detection–dimensionality trade-off and makes subsequent machine learning more expensive and the detection less accurate. This is because adding features that are essentially random noise may decrease the ability of the machine learning tool to learn the correct decision boundary. Also, with a small value of the truncation threshold T, some potentially useful information contained in the residual tails is lost, which limits the detection accuracy of highly adaptive schemes. Finally, since the co-occurrence dimensionality is $(2T+1)^D$, changing the parameters T and D gives the steganalyst rather limited options to control the feature dimensionality.

There are several possible avenues one can adopt to resolve the above issues. It is possible, for example, to overcome the problem with underpopulated bins by replacing the uniform scalar quantizer applied to each residual with a vector quantizer designed in the D-dimensional space of residuals and optimize w.r.t. the quantizer centroids. However, as the reference [32] shows, this approach lead to a rather negligible improvement in detection. A largely unexplored direction worth investigating involves representing adjacent residual samples with a high-dimensional joint distribution and then applying various dimensionality reduction techniques.

The avenue taken in this paper is to utilize dependencies among residual samples from a much larger neighborhood than what would be feasible to represent using a co-occurrence matrix. This way, we potentially use more information from the residual and thus improve the detection. Let us denote by $\mathcal{N}(\mathbf{Y}, i, j)$ an arbitrarily shaped neighborhood of pixel y_{ij} with $|\mathcal{N}|$ pixels. In the next section, we will consider rectangular $k \times l$ neighborhoods. Furthermore, we assume that the (unquantized) residual samples from $\mathcal{N}(\mathbf{Y}, i, j)$, $1 \leq i \leq n_1$, $1 \leq j \leq n_2$, are $|\mathcal{N}|$-dimensional vectors drawn from a probability distribution $\rho(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^{|\mathcal{N}|}$. Since for large $|\mathcal{N}|$, quantizing $\rho(\mathbf{x})$ and representing it using a co-occurrence matrix would not make a good test statistic due to heavily underpopulated bins, we instead project the

residual on random vectors $\mathbf{v} \in \mathbb{R}^{|\mathcal{N}|}$, $\mathbf{v} \neq 0$, and choose the first-order statistic of the projections as steganalysis features.

While it is certainly possible to use higher-order statistics for a fixed projection vector and neighborhood, in general, however, it is better to diversify the features by adding more projection neighborhoods and vectors rather than a more detailed description for one projection and neighborhood. See [12], [13], [16] for more details.

Intuitively, when selecting sufficiently many projection vectors $\mathbf{v}$, we improve our ability to distinguish between the distributions of cover and stego images. Furthermore, the random nature of vectors $\mathbf{v}$ is an important design element as it makes the steganalyzer key-dependent, making it harder for an adversary to design a steganographic scheme that evades detection by a specific steganalysis detector. The projection vectors could be optimized for a given cover source and stego method to obtain the best trade-off between feature dimensionality and detection accuracy. However, our goal is to present a universal feature vector capable of detecting potentially all stego schemes in arbitrary cover sources.

B. Residual Projection Features

In this section, we formally describe the process used to build the projection spatial rich model. We begin by introducing several key concepts. A specific instance of a projection neighborhood is obtained by first selecting two integers, $k, l \leq s$ randomly uniformly, where s is a fixed positive integer. The projection neighborhood is a matrix $\mathbf{\Pi} \in \mathbb{R}^{k \times l}$ whose elements, π_{ij}, are $k \cdot l$ independent realizations of a standard normal random variable $N(0, 1)$ normalized to a unit Frobenius norm $\|\mathbf{\Pi}\|_2 = 1$.[4] This way, the vector $\mathbf{v}$ obtained by arranging the elements of $\mathbf{\Pi}$, e.g., by rows, is selected randomly and uniformly from the surface of a unit sphere. This choice maximizes the spread of the projection directions.

To generate another instance of a projection neighborhood, we repeat the process with a different seed for the random selection of k, l as well as the elements of $\mathbf{\Pi}$. For a given instance of the projection neighborhood $\mathbf{\Pi}$ and residual $\mathbf{Z}$, the projection values $\mathbf{P}(\mathbf{\Pi}, \mathbf{Z})$ are obtained by convolving $\mathbf{Z}$ with the projection neighborhood $\mathbf{\Pi}$:

$$\mathbf{P}(\mathbf{\Pi}, \mathbf{Z}) = \mathbf{Z} * \mathbf{\Pi}. \tag{11}$$

Similarly to the features of the SRM, we utilize symmetries of natural images to endow the statistical descriptor with more robustness. In particular, we use the fact that statistical properties of natural images do not change with direction or mirroring. For non-directional residuals, such as the one obtained using the kernel (3), we can enlarge the set $\mathbf{P}$ (11) by adding to it projections with the matrix $\mathbf{\Pi}$ obtained by applying to it one or more following geometrical transformations: horizontal mirroring, vertical mirroring, rotation by 180 degrees, and transpose, respectively:

$$\overleftrightarrow{\mathbf{\Pi}} = \begin{pmatrix} \pi_{12} & \pi_{11} \\ \pi_{22} & \pi_{21} \end{pmatrix}, \tag{12}$$

$$\mathbf{\Pi}\updownarrow = \begin{pmatrix} \pi_{21} & \pi_{22} \\ \pi_{11} & \pi_{12} \end{pmatrix}, \tag{13}$$

$$\mathbf{\Pi}^{\circlearrowleft} = \begin{pmatrix} \pi_{22} & \pi_{21} \\ \pi_{12} & \pi_{11} \end{pmatrix}, \tag{14}$$

$$\mathbf{\Pi}^{T} = \begin{pmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{pmatrix}. \tag{15}$$

By combining these four transformations, one can obtain a total of eight different projection kernels.

The situation is a little more involved with directional residuals. The directional symmetry of natural images implies that we can merge the projections of a horizontal residual with projection kernels $\mathbf{\Pi}$, $\overleftrightarrow{\mathbf{\Pi}}$, $\mathbf{\Pi}\updownarrow$, and $\mathbf{\Pi}^{\circlearrowleft}$, and the projections obtained using their transposed versions applied to the vertical residual because its kernel is a transpose of the horizontal kernel.

Since a linear predictor (2) is a high-pass filter, the residual distribution for natural images will be zero mean and symmetrical about the y axis. Consequently, the distribution of the residual projections will also be symmetrical with a maximum at zero. Since we will be taking the first-order statistic (histogram) of the projections as the feature vector, the distribution symmetry allows us to work with absolute values of the projections and use either a finer histogram binning or a higher truncation threshold T. Denoting the bin width q, we will work with the following quantizer with $T + 1$ centroids:

$$\mathcal{Q}_{T,q} = \{q/2, 3q/2, \ldots, (2T + 1)q/2\}. \tag{16}$$

We would like to point out that by working with absolute values of the projections, our features will be unable to detect a steganographic scheme that preserves the distribution of the absolute values of projections yet one which violates the histogram symmetry. However, this is really only a minor issue as the projections are key-dependent and it would likely be infeasible to build an embedding scheme with this property for every projection vector and neighborhood. Moreover, an embedding scheme creating such an asymmetry would be fundamentally flawed as one could utilize this symmetry violation to construct a very accurate targeted quantitative attack. A good example is the Jsteg algorithm [36].

We now provide a formal description of the features. For a fixed set of quantizer centroids, $\mathcal{Q}_{T,q}$, the histogram of projections $\mathbf{P}$ is obtained using the following formula:

$$\mathbf{h}(l; \mathcal{Q}_{T,q}, \mathbf{P}) = \sum_{p \in \mathbf{P}} [Q_{\mathcal{Q}_{T,q}}(|p|) = l], \quad l \in \mathcal{Q}_{T,q}, \tag{17}$$

where [.] stands for the Iverson bracket defined as $[S] = 1$ when the statement S is true and 0 otherwise.

Considering the outputs of the residuals involved in computing a min (max) residual as independent random variables $Z_1, Z_2, \ldots, Z_r$, $E[\min\{Z_1, Z_2, \ldots, Z_r\}] < 0$ and $E[\max\{Z_1, Z_2, \ldots, Z_r\}] > 0$. Thus, the distribution of residuals obtained using the operations min (max) is not centered at zero and one can no longer work with absolute values

[4] The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{k \times l}$ is defined as $\|\mathbf{A}\|^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} a_{ij}^2$.

of residuals. Instead, we use the following expanded set of centroids:

$$\mathcal{Q}^{(\mathrm{x})}_{T,q} = \mathcal{Q}_{T,q} \cup \{-\mathcal{Q}_{T,q}\}, \tag{18}$$

which has double the cardinality of $\mathcal{Q}_{T,q}$. Because for any finite set $\mathcal{R} \subset \mathbb{R}$, $\min \mathcal{R} = -\max\{-\mathcal{R}\}$, the distribution of the projections $\mathbf{P}^{(\min)}$ of residuals $\mathbf{Z}^{(\min)}$ is a mirror image about the y axis of the distribution of $\mathbf{P}^{(\max)}$ of $\mathbf{Z}^{(\max)}$. One can use this symmetry to improve the robustness of the features and decrease their dimensionality by merging the projections $\mathbf{P}^{(\min)}$ and mirrored $\mathbf{P}^{(\max)}$ into one histogram:

$$\mathbf{h}(l; \mathcal{Q}^{(\mathrm{x})}_{T,q}, \mathbf{P}^{(\min)}, \mathbf{P}^{(\max)}) = \sum_{p \in \mathbf{P}^{(\min)}} [\mathcal{Q}_{\mathcal{Q}^{(\mathrm{x})}_{T,q}}(p) = l] + \sum_{p \in \mathbf{P}^{(\max)}} [\mathcal{Q}_{\mathcal{Q}^{(\mathrm{x})}_{T,q}}(-p) = -l], l \in \mathcal{Q}^{(\mathrm{x})}_{T,q}. \tag{19}$$

We note that the min a max residuals from the same submodel share the same projection neighborhood $\mathbf{\Pi}$.

To reduce the feature dimensionality, we do not include in the feature vector the last (marginal) bin $\mathbf{h}(l)$ corresponding to $l = (2T+1)q/2$ because its value can be computed from the remaining bins and is thus redundant for training the machine-learning-based classifier. Thus, for each linear residual $\mathbf{Z}$, the set of projections, $\mathbf{P}(\mathbf{Z}, \mathbf{\Pi})$, is represented in the PSRM using a T-dimensional vector $\mathbf{h}(l)$, $l \in \mathcal{Q}_{T,q} - \{(2T+1)q/2\}$. Similarly, and for the same reason, for a non-linear residual, we exclude the bins corresponding to $l = \pm(2T+1)q/2$, which gives us $2T$ features. Since in the SRM the features from two linear residuals are always paired up into one submodel (see Section II.C of [11]), we do the same in the proposed PSRM, which means that the projections of residuals from a given submodel are represented using exactly $2T$ features.

In summary, for a given submodel (a pair of residuals) and a projection neighborhood $\mathbf{\Pi}$ we obtain $2T$ values towards the PSRM. Since there are a total of 39 submodels in the SRM (and in the PSRM), the final dimensionality of the PSRM is

$$d(\nu) = 39 \cdot 2 \cdot T \cdot \nu, \tag{20}$$

where ν is the number of projection neighborhoods for each residual.

C. Parameter Setting

To construct the PSRM, we need to set the following parameters:

- ν ... the number of projection neighborhoods $\mathbf{\Pi}$ per residual;
- T ... the number of bins per projection neighborhood;
- s ... the maximum size of the projection neighborhood;
- q ... the bin width.

To capture a variety of complex dependencies among the neighboring residual samples, ν should be sufficiently large. Since larger ν increases the dimensionality of the feature space, $d(\nu)$, a reasonable balance must be stricken between feature dimensionality and detection accuracy.

Another parameter that influences the dimensionality is T– the number of bins per projection neighborhood. As

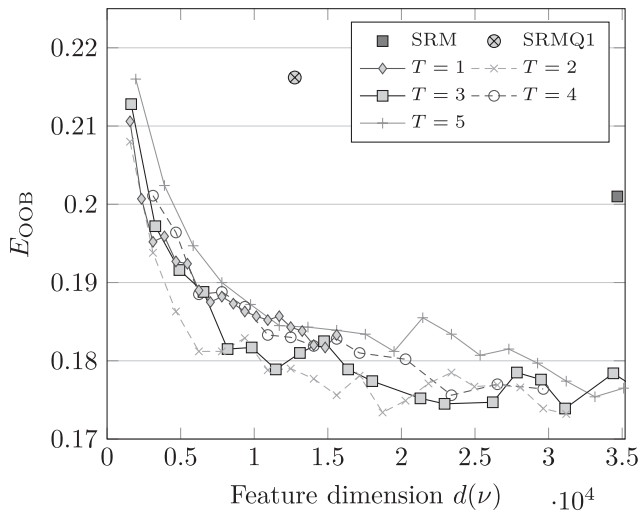

Fig. 1. Detection error E_{OOB} as a function of the PSRM feature-vector dimensionality $d(\nu)$ for $T \in \{1, \ldots, 5\}$ quantization bins per projection. Tested on S-UNIWARD on BOSSbase 1.01 at payload 0.4 bpp (bits per pixel).

mentioned in Section IV-A, the detection utilizes mainly the shape of the distribution, which is disturbed by the embedding process. Our experiments indicate that the number of bins necessary to describe the shape of the distribution of the projections can be rather small.

Figure 1 shows the detection–dimensionality tradeoff for different values of $d(\nu)$ and $T \in \{1, \ldots, 5\}$. The PSRM can clearly achieve the same detection reliability as SRM (SRMQ1) with much smaller dimensionality. One can trade a smaller value of T for larger ν to increase the performance for a fixed dimensionality. When choosing $\nu = 55$ and $T = 3$, the total dimensionality of the PSRM is $39 \cdot 2 \cdot T \cdot \nu = 12,870$, which makes its dimensionality almost the same of that of SRMQ1 (12, 753), allowing thus a direct comparison of both models. We opted for $T = 3$ as opposed to $T = 2$ because the performance for both choices is fairly similar and the choice $T = 3$ requires computing fewer projections for a fixed dimensionality, making the feature computation less computationally taxing.

The parameter s determines the maximal width and height of each projection neighborhood and thus limits the range of interpixel dependencies that can be utilized for detection. On the other hand, if the neighborhood is too large, the changes in the residual caused by embedding will have a small impact on the projection values, which will also become more dependent on the content. Moreover, the optimal value of s is likely to depend on the cover source. Experiments on BOSSbase 1.01 with S-UNIWARD at payload 0.4 bpp indicated a rather flat minimum around $s = 8$. We fixed s at this value and used it for all our experiments reported in this paper.

To capture the shape of the distribution, it is necessary to quantize the projection values. The impact of embedding manifests in the spatial domain differently depending on whether the actual embedding changes are executed in the spatial or the JPEG domain. Given the nature of JPEG

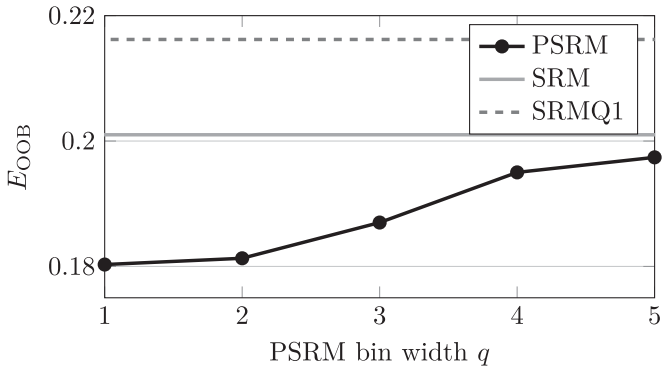

Fig. 2. Detection error as a function of the quantization bin width q when steganalyzing S-UNIWARD on BOSSbase at 0.4 bpp.

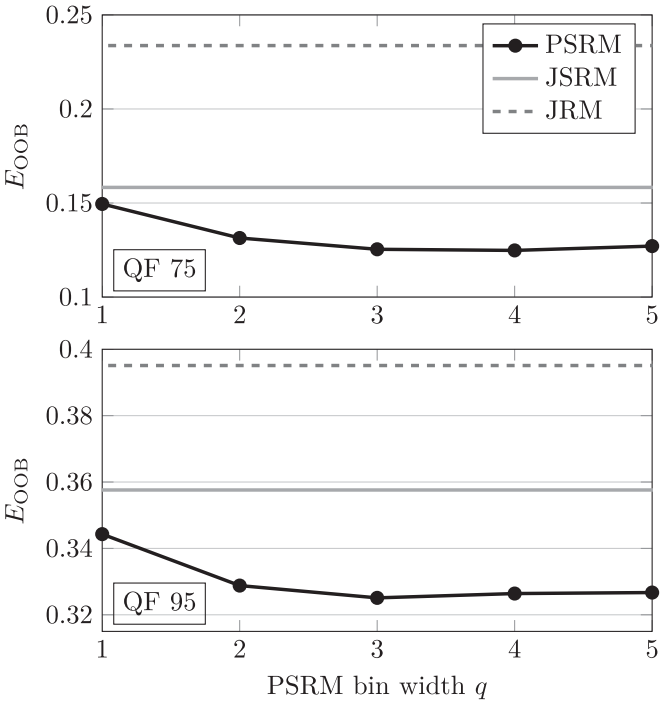

Fig. 3. Detection error as a function of the quantization bin width when steganalyzing q J-UNIWARD on BOSSbase compressed using quality factors 75 and 95.

compression, a change in a DCT coefficient has a more severe impact in the spatial domain depending on the quantization step of the particular DCT mode. Consequently, the best quantization bin width q will likely be different for detection of spatial- and JPEG-domain steganography. Figure 2 shows that the optimal value of q for spatial-domain embedding is $q = 1$, while the best value of q for steganalysis of JPEG-domain steganography is $q = 3$ (Figure 3). The PSRM versions used to detect embedding in the spatial and JPEG domains will be called PSRMQ1 and PSRMQ3, respectively.

V. Experiments

To evaluate the performance of the PSRM with dimension of 12,870, we ran experiments on multiple steganographic algorithms that embed messages in different domains. We contrast the results against several state-of-the-art domain-specific features sets. To show the universality of the proposed detection scheme, we added experiments on a markedly different cover source – the Leica database described in Section II-A.

In the spatial domain, we compare the PSRM with the SRM [11] (dimension 34,671) and the SRMQ1 (dimension 12,753). To the best knowledge of the authors, the SRM and SRMQ1 are the best spatial-domain feature sets available.

For JPEG-domain steganography, we compare with three rich models – the SRMQ1, the JPEG Rich Model (JRM) [28] with the dimension of 22,510, and JSRM, which is a merger of JRM and SRMQ1 with the total dimension of 35,263. Based on a thorough comparison reported in [28], the JSRM is currently the most powerful feature set for detection of JPEG domain steganography.

The empirical steganographic security in the JPEG domain is tested on two JPEG quality factors (QF) – 75 and 95. We selected these two quality factors as typical representatives of low quality and high quality compression factors.

We evaluate the performance of all feature sets on three payloads: 0.1, 0.2, and 0.4 bits per pixel (bpp) in the spatial domain and 0.1, 0.2, and 0.4 bits per non-zero AC coefficient (bpnzAC) in the JPEG domain. The main reason for using only three payloads is the high computational complexity involved with testing high-dimensional features on many algorithms covering three embedding domains. Moreover, as will become apparent from the experimental results revealed in the next section, showing the detection accuracy on a small, medium, and a large payload seems to provide sufficient information to compare the proposed PSRM with prior art.

In order to assess the statistical significance of the results, we measured the standard deviation of the E_{OOB} for all PSRM experiments measured on ten runs of the ensemble classifier with different seeds for its random generator that drives the selection of random subspaces as well as the bootstrapping for the training sets. The standard deviation was always below 0.3 %. We do not show it in the tables below to save on space and make the table data legible. The best performing features for every cover source, steganographic algorithm, and payload are highlighted in gray.

A. Spatial Domain

We first interpret the results on BOSSbase shown in Table I. Across all three embedding algorithms and payloads, the PSRM achieves a lower detection error than both SRMQ1 and SRM despite its almost three times larger dimensionality. Since the PSRM uses the same residuals as both SRM sets, it is safe to say that, for this image source, representing the residuals with projections is more efficient for steganalysis than forming co-occurrences. The actual improvement depends on the embedding algorithm. For HUGO, the PSRM lowers the detection error by about 2% w.r.t. the similar size SRMQ1. In light of the results of the BOSS competition reported at the 11th Information Hiding Conference [2], [12], [13], [16], this is a significant improvement. The difference between

TABLE I
DETECTION ERROR OF PSRM VERSUS SRMQ1 AND SRM FOR THREE CONTENT-ADAPTIVE STEGANOGRAPHIC ALGORITHMS EMBEDDING IN THE SPATIAL DOMAIN

Payload	0.1 bpp			0.2 bpp			0.4 bpp		
Features	PSRMQ1	SRMQ1	SRM	PSRMQ1	SRMQ1	SRM	PSRMQ1	SRMQ1	SRM
Dimension	12,870	12,753	34,671	12,870	12,753	34,671	12,870	12,753	34,671
BOSSbase									
HUGO	0.3564	0.3757	0.3651	0.2397	0.2701	0.2542	0.1172	0.1383	0.1278
WOW	0.3859	0.4119	0.3958	0.2950	0.3302	0.3117	0.1767	0.2170	0.1991
S-UNIWARD	0.3977	0.4182	0.4139	0.3025	0.3358	0.3159	0.1803	0.2162	0.2010
Leica									
HUGO	0.2170	0.2273	0.2110	0.0857	0.0802	0.0723	0.0213	0.0187	0.0177
WOW	0.2438	0.2418	0.2275	0.0997	0.0993	0.0903	0.0273	0.0245	0.0197
S-UNIWARD	0.2131	0.2188	0.2023	0.0800	0.0787	0.0722	0.0198	0.0192	0.0190

TABLE II
DETECTION ERROR OF PSRM VERSUS JRM AND JSRM FOR THREE JPEG-DOMAIN STEGANOGRAPHIC ALGORITHMS AND QUALITY FACTORS 75 AND 95

Payload	QF	0.1 bpnzAC					0.2 bpnzAC					0.4 bpnzAC				
Features		PSRMQ3	SRMQ1	JRM	JPSRM	JSRM	PSRMQ3	SRMQ1	JRM	JPSRM	JSRM	PSRMQ3	SRMQ1	JRM	JPSRM	JSRM
Dimension		12,870	12,753	22,510	35,380	35,263	12,870	12,753	22,510	35,380	35,263	12,870	12,753	22,510	35,380	35,263
BOSSbase																
nsF5	75	0.2609	0.2949	0.2115	0.1631	0.1742	0.0810	0.1162	0.0477	0.0188	0.0239	0.0057	0.0123	0.0036	0.0008	0.0013
UED ternary		0.3369	0.3621	0.3968	0.3393	0.3468	0.1856	0.2180	0.2680	0.1770	0.1934	0.0390	0.0612	0.0488	0.0202	0.0250
J-UNIWARD		0.4319	0.4578	0.4632	0.4350	0.4503	0.3244	0.3779	0.3990	0.3289	0.3564	0.1294	0.1933	0.2376	0.1228	0.1583
nsF5	95	0.3401	0.3831	0.1354	0.1220	0.1347	0.1749	0.2332	0.0114	0.0101	0.0089	0.0252	0.0540	0.0005	0.0005	0.0006
UED ternary		0.4785	0.4753	0.4750	0.4727	0.4786	0.4370	0.4331	0.4336	0.4133	0.4077	0.2759	0.2897	0.2604	0.2180	0.2205
J-UNIWARD		0.4943	0.4965	0.4923	0.4920	0.4940	0.4659	0.4752	0.4763	0.4622	0.4674	0.3256	0.3786	0.3951	0.3246	0.3576
Leica																
nsF5	75	0.2780	0.2965	0.2463	0.2040	0.2100	0.1060	0.1085	0.0783	0.0503	0.0458	0.0135	0.0114	0.0070	0.0047	0.0042
UED ternary		0.3028	0.3290	0.3643	0.2965	0.2987	0.1437	0.1570	0.2233	0.1295	0.1398	0.0270	0.0293	0.0525	0.0205	0.0200
J-UNIWARD		0.3627	0.3895	0.4233	0.3777	0.3803	0.2227	0.2538	0.3438	0.2225	0.2317	0.0610	0.0683	0.1398	0.0538	0.0593
nsF5	95	0.3833	0.4080	0.1425	0.1428	0.1370	0.2313	0.2580	0.0078	0.0090	0.0072	0.0473	0.0575	0.0002	0.0002	0.0002
UED ternary		0.4793	0.4792	0.4827	0.4767	0.4703	0.4283	0.4373	0.4410	0.4200	0.4115	0.2898	0.3020	0.2555	0.2300	0.2137
J-UNIWARD		0.4769	0.4802	0.4893	0.4797	0.4728	0.4363	0.4448	0.4517	0.4335	0.4315	0.3154	0.3380	0.3552	0.2940	0.2942

PSRMQ1 and SRMQ1 sets is even bigger (≈4%) for the highly adaptive WOW. This confirms our intuition that the projections do capture more complex interpixel dependencies and use them more efficiently for detection.

Table I clearly shows that steganalysis is easier in Leica images than in BOSSbase. This is mainly because of stronger interpixel dependencies in Leica images. Image downsampling without antialiasing used to create BOSSbase images weakens the dependencies and makes the detection more difficult [29]. Moreover, the BOSSbase database was acquired by seven different cameras, which makes it likely more difficult for the machine learning to find the separating hyperplane.

While we observed a significant detection improvement over the SRM for BOSSbase for the Leica database both PSRM and SRMQ1 offer a similar detection accuracy. The reader should realize that while the SRM achieves overall the lowest detection error, comparing SRM with PSRMQ1 is not really fair as the SRM has almost three times larger dimensionality. Since the parameters of both the PSRM and the SRM sets were optimized for maximal detection on BOSSbase, we attribute this observation to the fact that the much stronger pixel dependencies in Leica images make the co-occurrence bins much better populated, which improves the steganalysis.

B. JPEG Domain

Table II shows the results of all experiments in the JPEG domain on both BOSSbase and Leica databases for quality factors 75 and 95. In most cases, the PSRMQ3 achieved a lower detection error than SRMQ1, further fostering the claim already made in the previous section – that the projections are better suited for steganalysis than co-occurrences.

The JRM feature set, designed to utilize dependencies among DCT coefficients, shows a rather interesting behavior. Depending on the embedding algorithm and the embedding operation, the JRM's performance can be significantly better

TABLE III
DETECTION ERROR OF PSRM VERSUS JRM AND JSRM FOR TWO SIDE-INFORMED JPEG-DOMAIN STEGANOGRAPHIC ALGORITHMS AND QUALITY FACTORS 75 AND 95

Payload	QF	0.1 bpnzAC					0.2 bpnzAC					0.4 bpnzAC				
Features		PSRMQ3	SRMQ1	JRM	JPSRM	JSRM	PSRMQ3	SRMQ1	JRM	JPSRM	JSRM	PSRMQ3	SRMQ1	JRM	JPSRM	JSRM
Dimension		12,870	12,753	22,510	35,380	35,263	12,870	12,753	22,510	35,380	35,263	12,870	12,753	22,510	35,380	35,263
BOSSbase																
NPQ	75	0.4613	0.4677	0.4139	0.4076	0.4078	0.3609	0.3899	0.3171	0.2779	0.2871	0.0760	0.0990	0.0654	0.0345	0.0398
SI-UNIWARD		0.4952	0.4948	0.5004	0.4970	0.4965	0.4764	0.4872	0.4908	0.4770	0.4814	0.3744	0.4083	0.4470	0.3755	0.3989
NPQ	95	0.4950	0.4960	0.4295	0.4308	0.4313	0.4708	0.4708	0.3155	0.3136	0.3095	0.3358	0.3556	0.1471	0.1342	0.1349
SI-UNIWARD		0.4955	0.4950	0.4654	0.4672	0.4696	0.4830	0.4890	0.4651	0.4599	0.4602	0.3909	0.4337	0.4418	0.3790	0.4153
NPQ	75	0.4615	0.4637	0.4257	0.4127	0.4138	0.3457	0.3545	0.3257	0.2903	0.2968	0.0802	0.0862	0.0852	0.0483	0.0508
SI-UNIWARD		0.4933	0.4960	0.4963	0.4952	0.4953	0.4727	0.4777	0.4900	0.4848	0.4748	0.3712	0.3872	0.4473	0.3752	0.3802
NPQ	95	0.4868	0.4920	0.3435	0.3505	0.3518	0.4682	0.4785	0.2920	0.3030	0.2998	0.3727	0.3773	0.1660	0.1628	0.1477
SI-UNIWARD		0.4908	0.4957	0.4460	0.4415	0.4475	0.4872	0.4973	0.4480	0.4448	0.4563	0.4312	0.4475	0.4450	0.4083	0.4220

or worse than the performance of the spatial features (versions of PSRM and SRM). For example, the probability of detection error for the (by far) weakest nsF5 algorithm with payload 0.1 bpnzAC for quality factor 95 on BOSSbase using JRM is 13.54 % while it is 34.01 % for PSRMQ3 and 38.31 % for SRMQ1. This is caused by the nsF5's embedding operation designed to always decrease the absolute value of DCT coefficients. The JRM feature set is designed to exploit the effects of this "faulty" embedding operation. On the other hand, a qualitatively opposite behavior is observed for J-UNIWARD, which minimizes the relative distortion in the wavelet domain. Here, the spatial-domain features are generally much more effective than JRM since the embedding operation does not introduce artifacts in the distribution of quantized DCT coefficients detectable by the JRM.

As proposed in [27] and later confirmed in [28], the overall best detection of JPEG domain embedding algorithms is typically achieved by merging JPEG and spatial-domain features. It thus makes sense to introduce the merger of PSRMQ3 and JRM (JPSRM) whose dimensionality is similar to that of the JSRM (a merger of SRMQ1 and JRM). As expected, the JPSRM / JSRM provide the lowest detection error when compared to feature sets constrained to a specific embedding domain. On BOSSbase, the projection-based models provided the lowest detection error for almost all combinations of payload, embedding algorithm, and quality factor. On Leica, the performance of both JPSRM and JSRM was rather similar. Again, we attribute this to the fact that for the Leica source, the co-occurrences are generally better populated than for the BOSSbase. Finally, we would like to point out that for J-UNIWARD adding the JRM to PSRMQ3 generally brings only a rather negligible improvement, indicating that the main detection power resides in the spatial features (the PSRMQ3).

C. Side-Informed JPEG Domain

The performance comparison for side-informed JPEG-domain embedding methods shown in Table III strongly resembles the conclusions from the previous section. The merged feature spaces (JPSRM and JSRM) generally provide the lowest detection error when considering the statistical spread of the data (0.3%). It is worth pointing out that the JRM features are rather effective against the NPQ algorithm (see, e.g., the quality factor 95 and payload 0.4 bpnzAC). This indicates a presence of artifacts in the distribution of DCT coefficients that are well detected with the JRM, which further implies that the NPQ algorithm determines the embedding costs in the DCT domain in a rather suboptimal way. Also note that the detection errors for BOSSbase and Leica are much more similar in the JPEG domain when compared with the spatial domain. This is likely an effect of the lossy character of JPEG compression, which "erases" the high-frequency details (differences) between both sources.

VI. CONCLUSION

The key element in steganalysis of digital images using machine learning is their representation. Over the years, researchers converged towards a de facto standard representation that starts with computing a noise residual and then taking the sample joint distribution of residual samples as a feature for steganalysis. This co-occurrence based approach dominated the field for the past seven years. Co-occurrences, however, are rather non-homogeneous descriptors. With an increasing co-occurrence order, a large number of bins become underpopulated (statistically less significant), which leads to a feature dimensionality increase disproportional to the gain in detection performance. The co-occurrence order one can use in practice is thus limited, which prevents steganalysts from utilizing long-range dependencies among pixels that might further improve detection especially for content-adaptive steganographic schemes.

Aware of these limitations, in this article, we introduce an alternative statistical descriptor of residuals by projecting neighboring residual samples onto random directions and taking the first-order statistics of the projections as features. The resulting features are better populated and thus more statistically significant. Furthermore, the projection vectors as well as the size and shape of the projection neighborhoods further diversify the description, which boosts detection accuracy. The advantage of representing images using residual projections as opposed to co-occurrences is demonstrated on

several state-of-the-art embedding algorithms in the spatial, JPEG, and side-informed JPEG domains.

The new representation is called the projection spatial rich model (PSRM). We introduce two versions – one suitable for detection of spatial-domain steganography and one for the JPEG domain. Both versions differ merely in the quantization step used to quantize the projections. The PSRM is based on the exact same set of noise residuals as its predecessor – the spatial rich model. The fact that PSRM equipped with the same set of residuals as the SRM offers a better detection performance at the same dimensionality is indicative of the fact that the projections are indeed more efficient for steganalysis than co-occurrences.

The biggest advantage of PSRM over SRM becomes apparent for highly content adaptive algorithms, such as WOW or schemes employing the UNIWARD function. Besides a more accurate detection, the PSRM also enjoys a much better performance vs. dimensionality ratio. For spatial-domain algorithms, one can achieve the same detection accuracy as that of SRM with dimensionality 7–10 times smaller. This compactification, however, comes at a price, which is the computational complexity. This seems inevitable if one desires a descriptor that is more statistically relevant and diverse – the PSRM consists of a large number of projection histograms rather than a small(er) number of high-dimensional co-occurrences. The PSRM feature computation requires computing about 65,000 convolutions and histograms. A possible speed-up of the PSRM feature computation using graphical processing units (GPUs) was proposed in [23]. The PSRM feature extractor is available from.[5]

Finally, we make one more intriguing remark. The latest generation of currently most secure algorithms that embed messages in quantized DCT coefficients but minimize the embedding distortion computed in the spatial (wavelet) domain (J-UNIWARD and SI-UNIWARD) seem to be less detectable using features computed from quantized DCT coefficients and become, instead, more detectable using spatial-domain features (PSRM). This challenges the long heralded principle that the best detection is always achieved in the embedding domain. Unless the embedding rule is flawed (e.g, the embedding operation of LSB flipping or the F5 embedding operation), one should consider for detection representing the images in the domain in which the distortion is minimized.

ACKNOWLEDGMENT

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 Jan Kodovský for useful discussions.

[5]http://dde.binghamton.edu/download/feature_extractors/

REFERENCES

[1] I. Avcibas, N. D. Memon, and B. Sankur, "Steganalysis using image quality metrics," *Proc. SPIE*, vol. 4314, pp. 523–531, Jan. 2001.
[2] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system—The ins and outs of organizing BOSS," in *Proc. 13th Int. Conf.*, May 2011, pp. 59–70.
[3] R. Böhme, *Advanced Statistical Steganalysis*. New York, NY, USA: Springer-Verlag, 2010.
[4] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, Aug. 1996.
[5] R. Cogranne and F. Retraint, "An asymptotically uniformly most powerful test for LSB matching detection," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 3, pp. 464–476, Mar. 2013.
[6] R. Cogranne, C. Zitzmann, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu, "Statistical detection LSB matching using hypothesis testing theory," in *Proc. 14th Int. Conf.*, May 2012, pp. 46–62.
[7] H. Farid and L. Siwei, "Detecting hidden messages using higher-order statistics and support vector machines," in *Proc. 5th Int. Workshop*, Oct. 2002, pp. 340–354.
[8] L. Fillatre, "Adaptive steganalysis of least significant bit replacement in grayscale images," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 556–569, Feb. 2011.
[9] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 920–935, Sep. 2011.
[10] J. Fridrich, M. Goljan, and D. Soukal, "Perturbed quantization steganography using wet paper codes," in *Proc. 6th ACM*, Sep. 2004, pp. 4–15.
[11] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2011.
[12] J. Fridrich, J. Kodovský, M. Goljan, and V. Holub, "Breaking HUGO—The process discovery," in *Proc. 13th Int. Conf.*, May 2011, pp. 85–101.
[13] J. Fridrich, J. Kodovský, M. Goljan, and V. Holub, "Steganalysis of content-adaptive steganography in spatial domain," in *Proc. 13th Int. Conf.*, May 2011, pp. 102–117.
[14] J. Fridrich, T. Pevný, and J. Kodovský, "Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities," in *Proc. 9th ACM Multimedia Security Workshop*, Sep. 2007, pp. 3–14.
[15] M. Goljan, J. Fridrich, and T. Holotyak, "New blind steganalysis and its implications," *Proc. SPIE*, vol. 6072, pp. 1–13, Jan. 2006.
[16] G. Gül and F. Kurugollu, "A new methodology in steganalysis: Breaking highly undetectable steganography (HUGO)," in *Proc. 13th Int. Conf.*, May 2011, pp. 71–84.
[17] L. Guo, J. Ni, and Y.-Q. Shi, "An efficient JPEG steganographic scheme using uniform embedding," in *Proc. 4th IEEE Int. Workshop Inf. Forensics Security*, Dec. 2012, pp. 169–174.
[18] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proc. 4th IEEE Int. Workshop Inf. Forensics Security*, Dec. 2012, pp. 234–239.
[19] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *Proc. 1st ACM Workshop*, Jun. 2013, pp. 59–68.
[20] V. Holub and J. Fridrich, "Random projections of residuals as an alternative to co-occurrences in steganalysis," *Proc. SPIE*, vol. 8665, pp. 1–11, Feb. 2013.
[21] F. Huang, J. Huang, and Y.-Q. Shi, "New channel selection rule for JPEG steganography," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1181–1191, Aug. 2012.
[22] A. D. Ker, "A fusion of maximal likelihood and structural steganalysis," in *Proc. 9th Int. Workshop*, Jun. 2007, pp. 204–219.
[23] A. D. Ker, *Implementing the Projected Spatial Rich Features on a GPU*. San Francisco, CA, USA: TBD, Feb. 2014.
[24] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," *Proc. SPIE*, vol. 6819, pp. 1–17, Jan. 2008.
[25] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich, "The square root law of steganographic capacity," in *Proc. 10th ACM Multimedia Security Workshop*, Sep. 2008, pp. 107–116.
[26] Y. Kim, Z. Duric, and D. Richards, "Modified matrix encoding technique for minimal distortion steganography," in *Proc. 8th Int. Workshop Inf. Hiding*, Jul. 2006, pp. 314–327.
[27] J. Kodovský and J. Fridrich, "Calibration revisited," in *Proc.11th ACM Multimedia Security Workshop*, Sep. 2009, pp. 63–74.
[28] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," *Proc. SPIE*, vol. 8303, pp. 1–13, Jan. 2012.
[29] J. Kodovský and J. Fridrich, "Steganalysis in resized images," in *Proc. IEEE ICASSP*, May 2013, pp. 1–19.

[30] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 432–444, Apr. 2012.
[31] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 1, pp. 111–119, Mar. 2006.
[32] T. Pevný, "Co-occurrence steganalysis in high dimension," *Proc. SPIE*, vol. 8303, pp. 1–13, Jan. 2012.
[33] T. Pevný, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 215–224, Jun. 2010.
[34] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proc. 12th Int. Conf. Inf. Hiding*, Jun. 2010, pp. 161–177.
[35] V. Sachnev, H. J. Kim, and R. Zhang, "Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding," in *Proc. 11th ACM Multimedia Security Workshop*, Sep. 2009, pp. 131–140.
[36] D. Upham, (2013). *Steganographic Algorithm JSTeg* [Online]. Available: http://zooid.org/ paul/crypto/jsteg
[37] A. Westfeld, "High capacity despite better steganalysis (F5—A steganographic algorithm)," in *Proc. 4th Int. Workshop Inf. Hiding*, Apr. 2001, pp. 289–302.
[38] 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*, Mar. 2012, pp. 1793–1796.
[39] D. Zou, Y. Q. Shi, W. Su, and G. Xuan, "Steganalysis based on Markov model of thresholded prediction-error image," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 1365–1368.

Vojtech Holub is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY, USA. His main focus is on steganalysis and steganography. He received the M.S. degree in software engineering from Czech Technical University, Prague, in 2010.

Jessica Fridrich (M'05) holds the position of Professor of electrical and computer engineering at Binghamton University, Binghamton, NY, USA. She received the Ph.D. degree in systems science from Binghamton University in 1995 and the M.S. degree in applied mathematics from Czech Technical University, Prague, in 1987. Her main interests are in steganography, steganalysis, digital watermarking, and digital image forensic. Her research work has been generously supported by the U.S. Air Force and AFOSR. Since 1995, she has received 19 research grants totaling over $9 mil for projects on data embedding and steganalysis that lead to more than 140 papers and seven U.S. patents. She is a member of ACM.