Steganalysis of JPEG Images Using Rich Models

Jan Kodovský and Jessica Fridrich

Department of Electrical and Computer Engineering
Binghamton University, Binghamton, NY 13902-6000, USA

ABSTRACT

In this paper, we propose a rich model of DCT coefficients in a JPEG file for the purpose of detecting steganographic embedding changes. The model is built systematically as a union of smaller submodels formed as joint distributions of DCT coefficients from their frequency and spatial neighborhoods covering a wide range of statistical dependencies. Due to its high dimensionality, we combine the rich model with ensemble classifiers and construct detectors for six modern JPEG domain steganographic schemes: nsF5, model-based steganography, YASS, and schemes that use side information at the embedder in the form of the uncompressed image: MME, BCH, and BCHopt. The resulting performance is contrasted with previously proposed feature sets of both low and high dimensionality. We also investigate the performance of individual submodels when grouped by their type as well as the effect of Cartesian calibration. The proposed rich model delivers superior performance across all tested algorithms and payloads.

1. MOTIVATION

Modern image-steganography detectors consist of two basic parts: an image model and a machine learning tool that is trained to distinguish between cover and stego images represented in the chosen model. The detection accuracy is primarily determined by the image model, which should be sensitive to steganographic embedding changes and insensitive to the image content. It is also important that it captures as many dependencies among individual image elements (DCT coefficients) as possible to increase the chance that at least some of these dependencies will be disturbed by embedding. By measuring mutual information between coefficient pairs, it has been already pointed out [9] that the strongest dependencies among DCT coefficients are between close spatial-domain (inter-block) and frequency-domain (intra-block) neighbors. This fact was intuitively utilized by numerous researchers in the past, who proposed to represent JPEG images using joint or conditional probability distributions of neighboring coefficient pairs [1,3,14,15,17,22] possibly expanded with their calibrated versions [8, 11]. In [9, 10], the authors pointed out that by merging many such joint distributions (co-occurrence matrices), substantial improvement in detection accuracy can be obtained if combined with machine learning that can handle high model dimensionality and large training sets.

In this paper, we propose a complex (rich) model of JPEG images consisting of a large number of individual submodels. The novelty w.r.t. our previous contributions [9, 10] is at least three-fold: 1) we view the absolute values of DCT coefficients in a JPEG image as 64 weakly dependent parallel channels and separate the joint statistics by individual DCT modes; 2) to increase the model diversity, we form the same model from differences between absolute values of DCT coefficients; 3) we add integral joint statistics between coefficients from a wider range of values to cover the case when steganographic embedding largely avoids disturbing the first two models. Finally, the joint statistics are symmetrized to compactify the model and to increase its statistical robustness. This philosophy to constructing image models for steganalysis parallels our effort in the spatial domain [4]. We would like to point out that the proposed approach necessitates usage of scalable machine learning, such as the ensemble classifier that was originally described in [9] and then extended to a fully automatized routine in [10].

The JPEG Rich Model (JRM) is described in detail in Section 2. In Section 3, it is used to steganalyze six modern JPEG-domain steganographic schemes: nsF5 [5], MBS [19], YASS [21], MME [6], BCH, and BCHopt [18]. In combination with an ensemble classifier, the JRM outperforms not only low-dimensional models but also our previously proposed high-dimensional feature sets for JPEG steganalysis – the CC-C300 [9] and $\mathcal{CF}^*$ [10]. Afterwards, in Section 4, we subject the proposed JRM to analysis and conduct a series of investigative experiments revealing interesting insight and interpretations. The paper is concluded in Section 5.

E-mail: {jan.kodovsky, fridrich}@binghamton.edu; http://dde.binghamton.edu

Media Watermarking, Security, and Forensics 2012, edited by Nasir D. Memon, Adnan M. Alattar,
Edward J. Delp III, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 8303,
83030A · © 2012 SPIE-IS&T · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.907495

2. RICH MODEL IN JPEG DOMAIN

A JPEG image consists of 64 parallel channels formed by DCT modes which exhibit complex but short-distance dependencies of two types – frequency (intra-block) and spatial (inter-block). The former relates to the relationship among coefficients with similar frequency within the same 8×8 block while the latter refers to the relationship across different blocks. Although statistics of neighboring DCT coefficients were used as models in the past, the need to keep the model dimensionality low for the subsequent classifier training usually limited the model scope to co-occurrence matrices constructed from *all* coefficients in the DCT plane. Thus, despite their very different statistical nature, all DCT modes were treated equally.

Our proposed rich model consists of several qualitatively different parts. First, in the lines of our previously proposed $\mathcal{CF}^*$ features, we model individual DCT modes *separately,* collect many of these submodels a put them together. They will be naturally diverse since they capture dependencies among different DCT coefficients. The second part of the proposed JRM is formed as *integral* statistics from the whole DCT plane. The increased statistical power enables us to extend the range of co-occurrence features and therefore cover a different spectrum of dependencies than the mode-specific features from the first part. The features of both parts are further diversified by modeling not only DCT coefficients themselves, but also their *differences* calculated in different directions.

2.1 Notation and definitions

Quantized DCT coefficients of a JPEG image of dimensions $M \times N$ will be represented by a matrix $\mathbf{D} \in \mathbb{Z}^{M \times N}$. Let $\mathbf{D}^{(i,j)}_{xy}$ denote the (x, y)th DCT coefficient in the (i, j)th 8×8 block, $(x, y) \in \{0, \dots, 7\}^2$, $i = 1, \dots, \lceil M/8 \rceil$, $j = 1, \dots, \lceil N/8 \rceil$. Alternatively, we may access individual elements as $\mathbf{D}_{ij}$, $i = 1, \dots, M$, $j = 1, \dots, N$. We define the following matrices:

$$
\begin{aligned}
\mathbf{A}^{\times}_{i,j} &= |\mathbf{D}_{ij}|,\ i = 1, \dots, M,\ j = 1, \dots, N, && (1)\\
\mathbf{A}^{\rightarrow}_{i,j} &= |\mathbf{D}_{ij}| - |\mathbf{D}_{i,j+1}|,\ i = 1, \dots, M,\ j = 1, \dots, N-1, && (2)\\
\mathbf{A}^{\downarrow}_{i,j} &= |\mathbf{D}_{ij}| - |\mathbf{D}_{i+1,j}|,\ i = 1, \dots, M-1,\ j = 1, \dots, N, && (3)\\
\mathbf{A}^{\searrow}_{i,j} &= |\mathbf{D}_{ij}| - |\mathbf{D}_{i+1,j+1}|,\ i = 1, \dots, M-1,\ j = 1, \dots, N-1, && (4)\\
\mathbf{A}^{\rightrightarrows}_{i,j} &= |\mathbf{D}_{ij}| - |\mathbf{D}_{i,j+8}|,\ i = 1, \dots, M,\ j = 1, \dots, N-8, && (5)\\
\mathbf{A}^{\downdownarrows}_{i,j} &= |\mathbf{D}_{ij}| - |\mathbf{D}_{i+8,j}|,\ i = 1, \dots, M-8,\ j = 1, \dots, N. && (6)
\end{aligned}
$$

Matrix $\mathbf{A}^{\times}$ consists of the absolute values of DCT coefficients, matrices $\mathbf{A}^{\rightarrow}, \mathbf{A}^{\downarrow}, \mathbf{A}^{\searrow}$ are obtained as intra-block differences, and $\mathbf{A}^{\rightrightarrows}, \mathbf{A}^{\downdownarrows}$ represent inter-block differences. Individual submodels of the proposed JRM will be formed as 2D co-occurrence matrices calculated from the coefficients of matrices $\mathbf{A}^{\star}$, $\star \in \{\times, \rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$, positioned in DCT modes (x, y) and $(x + \Delta x, y + \Delta y)$. Formally, $\mathbf{C}^{\star}_T(x, y, \Delta x, \Delta y)$, $\star \in \{\times, \rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$, are $(2T + 1)^2$-dimensional matrices with elements

$$
c^{\star}_{kl}(x, y, \Delta x, \Delta y) = \frac{1}{Z} \sum_{i,j} \left| \left\{ \mathbf{T}^{(i,j)}_{xy} \middle| \mathbf{T} = \mathrm{trunc}_T(\mathbf{A}^{\star});\ \mathbf{T}^{(i,j)}_{xy} = k;\ \mathbf{T}^{(i,j)}_{x+\Delta x, y+\Delta y} = l \right\} \right|, \tag{7}
$$

where the normalization constant Z ensures that $\sum_{k,l} c_{kl} = 1$, and $\mathrm{trunc}_T(\cdot)$ is an element-wise truncation operator defined as

$$
\mathrm{trunc}_T(x) = \begin{cases} T \cdot \mathrm{sign}(x) & \text{if } |x| > T, \\ x & \text{otherwise.} \end{cases} \tag{8}
$$

In definition (7), we do not constrain Δx and Δy and allow $(x + \Delta x, y + \Delta y)$ to be out of the range $\{0, \dots, 7\}^2$ to more easily describe co-occurrences for inter-block coefficient pairs, e.g., $\mathbf{T}^{(i,j)}_{x+8,y} \equiv \mathbf{T}^{(i+1,j)}_{xy}$.

Assuming the statistics of natural images do not change after mirroring about the main diagonal, the symmetry of DCT basis functions w.r.t. the 8×8 block diagonal allows us to replace matrices $\mathbf{C}^{\star}_T$ with the more robust

$$
\begin{aligned}
\bar{\mathbf{C}}_T^{\times}(x,y,\Delta x,\Delta y) &\triangleq \frac{1}{2}\left(\mathbf{C}_T^{\times}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\times}(y,x,\Delta y,\Delta x)\right), && (9)\\
\bar{\mathbf{C}}_T^{\to}(x,y,\Delta x,\Delta y) &\triangleq \frac{1}{2}\Big(\mathbf{C}_T^{\to}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\downarrow}(y,x,\Delta y,\Delta x)\Big), && (10)\\
\bar{\mathbf{C}}_T^{\rightrightarrows}(x,y,\Delta x,\Delta y) &\triangleq \frac{1}{2}\Big(\mathbf{C}_T^{\rightrightarrows}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\downdownarrows}(y,x,\Delta y,\Delta x)\Big), && (11)\\
\bar{\mathbf{C}}_T^{\searrow}(x,y,\Delta x,\Delta y) &\triangleq \frac{1}{2}\Big(\mathbf{C}_T^{\searrow}(x,y,\Delta x,\Delta y)+\mathbf{C}_T^{\searrow}(y,x,\Delta y,\Delta x)\Big). && (12)
\end{aligned}
$$

Because the coefficients in $\mathbf{A}_{i,j}^{\times}$ are non-negative, most of the bins of $\bar{\mathbf{C}}_T^{\times}$ are zeros and its true dimensionality is only $(T+1)^2$. The difference-based co-occurrences $\bar{\mathbf{C}}_T^{\star}$, $\star \in \{\to, \searrow, \rightrightarrows\}$, are generally nonzero, however, we can additionally utilize their *sign symmetry* ($c_{kl}^{\star} \approx c_{-k,-l}^{\star}$) and define $\hat{\mathbf{C}}_T^{\star}$ with elements

$$\hat{c}_{kl}^{\star} = \frac{1}{2}\left(\bar{c}_{kl}^{\star} + \bar{c}_{-k,-l}^{\star}\right). \qquad (13)$$

The redundant portions of $\hat{\mathbf{C}}_T^{\star}$ can be removed to obtain the final form of the difference-based co-occurrences of dimensionality $\frac{1}{2}(2T+1)^2+\frac{1}{2}$, which we denote again $\hat{\mathbf{C}}_T^{\star}(x,y,\Delta x,\Delta y)$, $\star \in \{\to, \searrow, \rightrightarrows\}$. The rich model will be constructed only using the most compact forms: $\bar{\mathbf{C}}_T^{\times}$, $\hat{\mathbf{C}}_T^{\to}$, $\hat{\mathbf{C}}_T^{\searrow}$, and $\hat{\mathbf{C}}_T^{\rightrightarrows}$.

We note that the co-occurrences $\bar{\mathbf{C}}_T^{\times}$ evolved from the $\mathcal{F}^*$ feature set proposed in [10]. The difference is that $\mathcal{F}^*$ does not take absolute values before forming co-occurrences. Taking absolute values reduces dimensionality and makes the features more robust; it could be seen as another type of symmetrization. In Section 3, we compare the performance of the proposed rich model with the Cartesian calibrated $\mathcal{CF}^*$set [10].

2.2 DCT-mode specific components of JRM

Depending on the mutual position of the DCT modes (x,y) and $(x+\Delta x, y+\Delta y)$, the extracted co-occurrence matrices $\mathbf{C} \in \{\bar{\mathbf{C}}_T^{\times}, \hat{\mathbf{C}}_T^{\to}, \hat{\mathbf{C}}_T^{\searrow}, \hat{\mathbf{C}}_T^{\rightrightarrows}\}$ will be grouped into ten qualitatively different submodels:

1. $\mathcal{G}_{\mathrm{h}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,1) | 0 \le x;\ 0 \le y;\ x+y \le 5\}$,
2. $\mathcal{G}_{\mathrm{d}}(\mathbf{C}) = \{\mathbf{C}(x,y,1,1) | 0 \le x \le y;\ x+y \le 5\} \cup \{\mathbf{C}(x,y,1,-1) | 0 \le x < y;\ x+y \le 5\}$,
3. $\mathcal{G}_{\mathrm{oh}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,2) | 0 \le x;\ 0 \le y;\ x+y \le 4\}$,
4. $\mathcal{G}_{\mathrm{x}}(\mathbf{C}) = \{\mathbf{C}(x,y,y-x,x-y) | 0 \le x < y;\ x+y \le 5\}$,
5. $\mathcal{G}_{\mathrm{od}}(\mathbf{C}) = \{\mathbf{C}(x,y,2,2) | 0 \le x \le y;\ x+y \le 4\} \cup \{\mathbf{C}(x,y,2,-2) | 0 \le x < y;\ x+y \le 5\}$,
6. $\mathcal{G}_{\mathrm{km}}(\mathbf{C}) = \{\mathbf{C}(x,y,-1,2) | 1 \le x;\ 0 \le y;\ x+y \le 5\}$,
7. $\mathcal{G}_{\mathrm{ih}}(\mathbf{C}) = \{\mathbf{C}(x,y,0,8) | 0 \le x;\ 0 \le y;\ x+y \le 5\}$,
8. $\mathcal{G}_{\mathrm{id}}(\mathbf{C}) = \{\mathbf{C}(x,y,8,8) | 0 \le x \le y;\ x+y \le 5\}$,
9. $\mathcal{G}_{\mathrm{im}}(\mathbf{C}) = \{\mathbf{C}(x,y,-8,8) | 0 \le x \le y;\ x+y \le 5\}$,
10. $\mathcal{G}_{\mathrm{ix}}(\mathbf{C}) = \{\mathbf{C}(x,y,y-x,x-y+8) | 0 \le x;\ 0 \le y;\ x+y \le 5\}$.

The first six submodels capture intra-block relationships: $\mathcal{G}_{\mathrm{h}}$ – horizontally (and vertically, after symmetrization) neighboring pairs; $\mathcal{G}_{\mathrm{d}}$ – diagonally and minor-diagonally neighboring pairs; $\mathcal{G}_{\mathrm{oh}}$ – "skip one" horizontally neighboring pairs; $\mathcal{G}_{\mathrm{x}}$ – pairs symmetrically positioned w.r.t. the 8×8 block diagonal; $\mathcal{G}_{\mathrm{od}}$ – "skip one" diagonal and minor-diagonal pairs; $\mathcal{G}_{\mathrm{km}}$ – "knight-move" positioned pairs. The last four submodels capture inter-block relationships between coefficients from neighboring blocks: $\mathcal{G}_{\mathrm{ih}}$ – horizontal neighbors in the same DCT mode; $\mathcal{G}_{\mathrm{id}}$ – diagonal neighbors in the same mode; $\mathcal{G}_{\mathrm{im}}$ – minor-diagonal neighbors in the same mode, $\mathcal{G}_{\mathrm{ix}}$ – horizontal neighbors in modes symmetrically positioned w.r.t. 8×8 block diagonal. The two parts forming $\mathcal{G}_{\mathrm{d}}$ and $\mathcal{G}_{\mathrm{od}}$ were grouped together to give all submodels roughly the same dimensionality.

Since all ten groups of submodels are constructed for $\mathbf{C} \in \{\bar{\mathbf{C}}_3^{\times}, \hat{\mathbf{C}}_2^{\to}, \hat{\mathbf{C}}_2^{\searrow}, \hat{\mathbf{C}}_2^{\rightrightarrows}\}$, 40 DCT-mode specific submodels are obtained in total. For co-occurrences of absolute values of DCT coefficients, we fixed $T = 3$

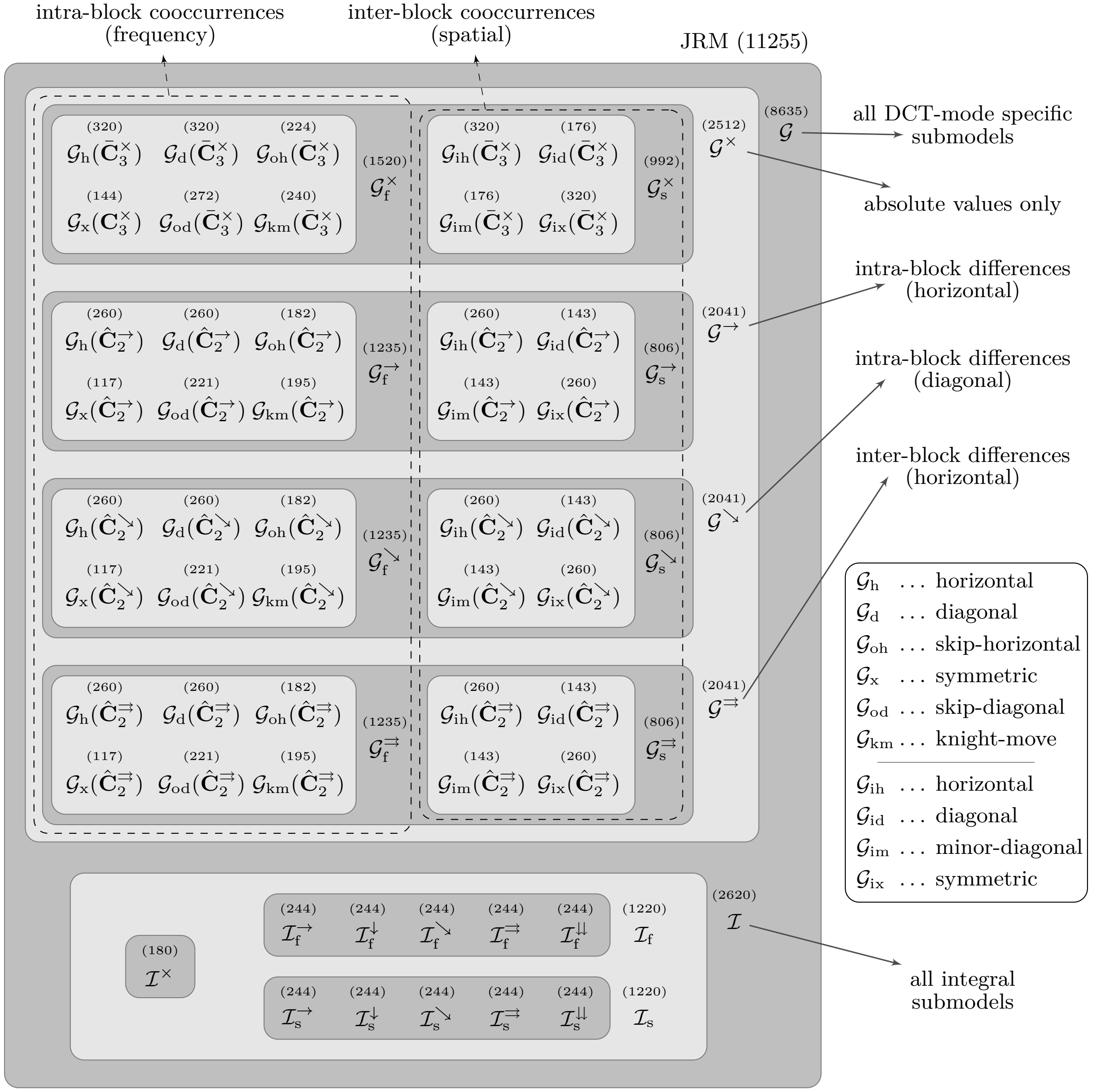
intra-block cooccurrences (frequency)
inter-block cooccurrences (spatial)
JRM (11255)
(320) $\mathcal{G}_{\mathrm{h}}(\bar{\mathbf{C}}_3^{\times})$ (320) $\mathcal{G}_{\mathrm{d}}(\bar{\mathbf{C}}_3^{\times})$ (224) $\mathcal{G}_{\mathrm{oh}}(\bar{\mathbf{C}}_3^{\times})$
(144) $\mathcal{G}_{\mathrm{x}}(\bar{\mathbf{C}}_3^{\times})$ (272) $\mathcal{G}_{\mathrm{od}}(\bar{\mathbf{C}}_3^{\times})$ (240) $\mathcal{G}_{\mathrm{km}}(\bar{\mathbf{C}}_3^{\times})$
(1520) $\mathcal{G}_{\mathrm{f}}^{\times}$
(320) $\mathcal{G}_{\mathrm{ih}}(\bar{\mathbf{C}}_3^{\times})$ (176) $\mathcal{G}_{\mathrm{id}}(\bar{\mathbf{C}}_3^{\times})$
(176) $\mathcal{G}_{\mathrm{im}}(\bar{\mathbf{C}}_3^{\times})$ (320) $\mathcal{G}_{\mathrm{ix}}(\bar{\mathbf{C}}_3^{\times})$
(992) $\mathcal{G}_{\mathrm{s}}^{\times}$
(2512) $\mathcal{G}^{\times}$
(8635) $\mathcal{G}$
all DCT-mode specific submodels
absolute values only
(260) $\mathcal{G}_{\mathrm{h}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (260) $\mathcal{G}_{\mathrm{d}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (182) $\mathcal{G}_{\mathrm{oh}}(\hat{\mathbf{C}}_2^{\rightarrow})$
(117) $\mathcal{G}_{\mathrm{x}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (221) $\mathcal{G}_{\mathrm{od}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (195) $\mathcal{G}_{\mathrm{km}}(\hat{\mathbf{C}}_2^{\rightarrow})$
(1235) $\mathcal{G}_{\mathrm{f}}^{\rightarrow}$
(260) $\mathcal{G}_{\mathrm{ih}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (143) $\mathcal{G}_{\mathrm{id}}(\hat{\mathbf{C}}_2^{\rightarrow})$
(143) $\mathcal{G}_{\mathrm{im}}(\hat{\mathbf{C}}_2^{\rightarrow})$ (260) $\mathcal{G}_{\mathrm{ix}}(\hat{\mathbf{C}}_2^{\rightarrow})$
(806) $\mathcal{G}_{\mathrm{s}}^{\rightarrow}$
(2041) $\mathcal{G}^{\rightarrow}$
intra-block differences (horizontal)
(260) $\mathcal{G}_{\mathrm{h}}(\hat{\mathbf{C}}_2^{\searrow})$ (260) $\mathcal{G}_{\mathrm{d}}(\hat{\mathbf{C}}_2^{\searrow})$ (182) $\mathcal{G}_{\mathrm{oh}}(\hat{\mathbf{C}}_2^{\searrow})$
(117) $\mathcal{G}_{\mathrm{x}}(\hat{\mathbf{C}}_2^{\searrow})$ (221) $\mathcal{G}_{\mathrm{od}}(\hat{\mathbf{C}}_2^{\searrow})$ (195) $\mathcal{G}_{\mathrm{km}}(\hat{\mathbf{C}}_2^{\searrow})$
(1235) $\mathcal{G}_{\mathrm{f}}^{\searrow}$
(260) $\mathcal{G}_{\mathrm{ih}}(\hat{\mathbf{C}}_2^{\searrow})$ (143) $\mathcal{G}_{\mathrm{id}}(\hat{\mathbf{C}}_2^{\searrow})$
(143) $\mathcal{G}_{\mathrm{im}}(\hat{\mathbf{C}}_2^{\searrow})$ (260) $\mathcal{G}_{\mathrm{ix}}(\hat{\mathbf{C}}_2^{\searrow})$
(806) $\mathcal{G}_{\mathrm{s}}^{\searrow}$
(2041) $\mathcal{G}^{\searrow}$
intra-block differences (diagonal)
(260) $\mathcal{G}_{\mathrm{h}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (260) $\mathcal{G}_{\mathrm{d}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (182) $\mathcal{G}_{\mathrm{oh}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(117) $\mathcal{G}_{\mathrm{x}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (221) $\mathcal{G}_{\mathrm{od}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (195) $\mathcal{G}_{\mathrm{km}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(1235) $\mathcal{G}_{\mathrm{f}}^{\rightrightarrows}$
(260) $\mathcal{G}_{\mathrm{ih}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (143) $\mathcal{G}_{\mathrm{id}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(143) $\mathcal{G}_{\mathrm{im}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$ (260) $\mathcal{G}_{\mathrm{ix}}(\hat{\mathbf{C}}_2^{\rightrightarrows})$
(806) $\mathcal{G}_{\mathrm{s}}^{\rightrightarrows}$
(2041) $\mathcal{G}^{\rightrightarrows}$
inter-block differences (horizontal)
$\mathcal{G}_{\mathrm{h}}$ ... horizontal
$\mathcal{G}_{\mathrm{d}}$ ... diagonal
$\mathcal{G}_{\mathrm{oh}}$ ... skip-horizontal
$\mathcal{G}_{\mathrm{x}}$ ... symmetric
$\mathcal{G}_{\mathrm{od}}$ ... skip-diagonal
$\mathcal{G}_{\mathrm{km}}$ ... knight-move
$\mathcal{G}_{\mathrm{ih}}$ ... horizontal
$\mathcal{G}_{\mathrm{id}}$ ... diagonal
$\mathcal{G}_{\mathrm{im}}$ ... minor-diagonal
$\mathcal{G}_{\mathrm{ix}}$ ... symmetric
(180) $\mathcal{I}^{\times}$
(244) $\mathcal{I}_{\mathrm{f}}^{\rightarrow}$ (244) $\mathcal{I}_{\mathrm{f}}^{\downarrow}$ (244) $\mathcal{I}_{\mathrm{f}}^{\searrow}$ (244) $\mathcal{I}_{\mathrm{f}}^{\rightrightarrows}$ (244) $\mathcal{I}_{\mathrm{f}}^{\downdownarrows}$ (1220) $\mathcal{I}_{\mathrm{f}}$
(244) $\mathcal{I}_{\mathrm{s}}^{\rightarrow}$ (244) $\mathcal{I}_{\mathrm{s}}^{\downarrow}$ (244) $\mathcal{I}_{\mathrm{s}}^{\searrow}$ (244) $\mathcal{I}_{\mathrm{s}}^{\rightrightarrows}$ (244) $\mathcal{I}_{\mathrm{s}}^{\downdownarrows}$ (1220) $\mathcal{I}_{\mathrm{s}}$
(2620) $\mathcal{I}$
all integral submodels

Figure 1. The proposed JPEG rich model and its decomposition into individual subgroups and submodels. The numbers denote the dimensionalities of the corresponding sets. Cartesian calibration doubles all shown values.

yielding the dimensionality of a single matrix $\bar{\mathbf{C}}_3^{\times}$ equal to 16. For difference-based co-occurrences, we fixed $T = 2$ to obtain a similar dimensionality of 13. Larger values of T would result in many underpopulated bins, especially for smaller images. A tabular listing of all introduced submodels, including their total dimensionalities, is shown in Figure 1.

2.3 Integral components of JRM

The mode-specific submodels introduced in Section 2.2 give the rich model a finer "granularity" but at the price of utilizing only a small portion of the DCT plane at a time. In order not to lose the integral statistical power of the whole DCT plane and to cover a larger range of DCT coefficients, we now finalize the rich model by supplementing additional co-occurrence matrices that are integrated over all DCT modes. We do so for both the co-occurrences of absolute values of DCT coefficients, $\bar{\mathbf{C}}_T^{\times}$, and their differences, $\hat{\mathbf{C}}_T^{\star}$, $\star \in \{\rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$. As the integral bins are better populated than DCT-mode specific bins, we increase T to 5. The integral submodels are defined as follows:

1. $\mathcal{I}^{\times} = \left\{ \sum_{x,y} \bar{\mathbf{C}}_5^{\times}(x, y, \Delta x, \Delta y) \middle| [\Delta x, \Delta y] \in \{(0,1), (1,1), (1,-1), (0,8), (8,8)\} \right\}$,
2. $\mathcal{I}_{\mathrm{f}}^{\star} = \left\{ \sum_{x,y} \hat{\mathbf{C}}_5^{\star}(x, y, \Delta x, \Delta y) \middle| [\Delta x, \Delta y] \in \{(0,1), (1,0), (1,1), (1,-1)\} \right\}$, $\star \in \{\rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$,
3. $\mathcal{I}_{\mathrm{s}}^{\star} = \left\{ \sum_{x,y} \hat{\mathbf{C}}_5^{\star}(x, y, \Delta x, \Delta y) \middle| [\Delta x, \Delta y] \in \{(0,8), (8,0), (8,8), (8,-8)\} \right\}$, $\star \in \{\rightarrow, \downarrow, \searrow, \rightrightarrows, \downdownarrows\}$,

For intra-block pairs, the summation in the above definitions is always over all DCT modes $(x, y) \in \{0, \ldots, 7\}^2$ such that both (x, y) and $(x + \Delta x, y + \Delta y)$ lie within the same 8×8 block. A similar constraint applies to the inter-block matrices whenever the indices would end up outside of the DCT array. DC modes are omitted in all definitions. The submodel $\mathcal{I}^{\times}$ covers both the spatial (inter-block) and frequency (intra-block) dependencies, and can be seen as an extension of feature sets $absNJ_1$ and $absNJ_2$ by Liu [14]. The difference-based submodels bear similarity to the Markov features proposed in [1], where the authors also utilized inter- and intra-block differences between absolute values of DCT coefficients. In order to obtain similar dimensionalities, the co-occurrences calculated from differences were divided into two distinct sets, capturing frequency ($\mathcal{I}_{\mathrm{f}}^{\star}$) and spatial ($\mathcal{I}_{\mathrm{s}}^{\star}$) dependencies separately.[*]

The union of DCT-mode specific submodels with the integral submodels form the JPEG domain rich model we propose. Its total dimensionality is 11,255. In order to improve the performance, we apply the Cartesian calibration [8] which doubles the dimensionality to 22,510. The structure of the entire JRM appears in Figure 1.

3. COMPARISON TO PRIOR ART

To demonstrate the power of the proposed JPEG rich model, we steganalyze six modern steganographic methods. We use the ensemble classifier [9, 10] for all experiments as it enables fast training in high-dimensional feature spaces and its performance on low-dimensional feature sets is comparable to the much more complex SVMs [10]. The classifier is constructed from base learners implemented as Fisher Linear Discriminants on random subspaces of the feature space. The number of subspaces and their dimensionality were determined automatically using the algorithms described in [10]. A publicly available implementation of the ensemble classifier can be downloaded from `http://dde.binghamton.edu/download/ensemble`.

3.1 Tested steganographic methods

The tested steganographic methods are: nsF5, MBS, YASS, MME, BCH, and BCHopt. The nsF5 algorithm [5] is an improved version of the popular F5 [24]. For experiments, we used a simulator of nsF5, available at `http://dde.binghamton.edu/download/nsf5simulator/`, that makes the embedding changes as if an optimal binary matrix coding scheme was used. We note that a near-optimal practical implementation can be achieved using syndrome-trellis codes [2].

MBS is a model-based steganography due to Sallee [19]. The implementation we used is available at `http://www.philsallee.com/mbsteg`. Both nsF5 and MBS start directly with the JPEG image to be modified and thus do not utilize any side information.

YASS, which hides data robustly in a transform domain, was introduced in [23] and later improved in [20, 21]. Even though it is easily detectable today [11, 13, 14], it played an important role to clarify the real purpose of the process of feature calibration [8]. We test five different settings of YASS, numbered 3, 8, 10, 11, and 12 in [11], as these were reported to be the most secure. YASS performs only full embedding and thus the reported payloads are averages over all images in the CAMERA database to be described next.

[*] Note that some of the submodels capture quite complex statistical dependencies among DCT coefficients. For example, $\mathcal{D}_{\mathrm{f}}^{\rightrightarrows}$ combines *inter*-block differences with *intra*-block co-occurrences.

MME [6] utilizes side information at the sender in terms of the uncompressed image and employs matrix embedding to minimize an appropriately defined distortion function. We tested its Java implementation that uses a Java JPEG encoder for image compression. Therefore, in order to steganalyze solely the impact of MME embedding, we need to create cover images using the same compressor to avoid artificially increasing the detection reliability by also detecting traces of a different JPEG compressor [7].

BCH and BCHopt [18] are side-informed algorithms that employ BCH codes to minimize the embedding distortion in the DCT domain defined using the knowledge of non-rounded DCT coefficients. BCHopt is an improved version of BCH that contains a heuristic optimization and also hides message bits into zeros. According to the experiments in [18], BCHopt is currently the most secure practical JPEG steganographic scheme. To the best of our knowledge, it has not been steganalyzed elsewhere.

3.2 Performance evaluation

The image source on which all experiments were carried out is the CAMERA database containing 6,500 JPEG images originally acquired in their RAW format taken by 22 digital cameras, resized so that the smaller size is 512 pixels with aspect ratio preserved, and converted to grayscale. The cover images for MME were created using a Java JPEG encoder. For the rest, we used Matlab's function `imwrite`. The JPEG quality factor was fixed to 75 in both cases.

For every steganographic method, we created stego images using a range of different payload sizes expressed in terms of bits per nonzero AC DCT coefficient (bpac), and trained a separate classifier to detect each of them. Before classification, all cover-stego pairs were divided into two halves for training and testing, respectively. We define the minimal total error P_{E} under equal priors achieved on the testing set as

$$P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{P_{\mathrm{FA}} + P_{\mathrm{MD}}(P_{\mathrm{FA}})}{2}, \tag{14}$$

where P_{FA} is the false alarm rate and P_{MD} is the missed detection rate. The performance is evaluated using the median value of P_{E} over ten random 50/50 splits of the database and will be denoted as $\bar{P}_{\mathrm{E}}$.

We compare the steganalysis performance of the following feature spaces (models); the numbers in brackets denote their dimensionality:

- CHEN (486) = Markov features utilizing both intra- and inter-block dependencies [1],
- CC-CHEN (972) = CHEN features improved by Cartesian calibration [8],
- LIU (216) = the union of *diff-absNJ-ratio* and *ref-diff-absNJ* features published in [14],
- CC-PEV (548) = Cartesian-calibrated PEV feature set [17],
- CDF (1,234) = CC-PEV features expanded by SPAM features [16] extracted from spatial domain,
- CC-C300 (48,600) = the high-dimensional feature space proposed in [9],
- $\mathcal{CF}^*$ (7,850) = compact rich model for DCT domain proposed in [10],
- JRM (11,255) = the rich model proposed in this paper, without calibration,
- CC-JRM (22,510) = Cartesian-calibrated JRM,
- J+SRM (35,263) = the union of CC-JRM and the Spatial-domain Rich Model (SRM) proposed in [4] (all 39 submodels of SRM were taken with a fixed quantization $q = 1c$, see [4] for more details).

Resulting errors $\bar{P}_{\mathrm{E}}$ are reported in Table 1. The proposed CC-JRM delivers the best performance among all feature sets that are extracted directly from the DCT domain, across all tested steganographic methods and all payloads. Adding the spatial-domain rich model [4] further improves the performance and delivers the overall best results.

Algorithm	Payload (bpac)	CHEN (486)	CC-CHEN (972)	LIU (216)	CC-PEV (548)	CDF (1,234)	CC-C300 (48,600)	$\mathcal{CF}^*$ (7,850)	JRM (11,255)	CC-JRM (22,510)	J+SRM (35,263)
nsF5	0.050	0.4153	0.3816	0.3377	0.3690	0.3594	0.3722	0.3377	0.3407	0.3298	0.3146
	0.100	0.3097	0.2470	0.1732	0.2239	0.2020	0.2207	0.1737	0.1782	0.1616	0.1375
	0.150	0.2094	0.1393	0.0706	0.1171	0.0906	0.1127	0.0720	0.0793	0.0663	0.0468
	0.200	0.1345	0.0708	0.0273	0.0549	0.0360	0.0486	0.0273	0.0338	0.0255	0.0150
MBS	0.010	0.4070	0.3962	0.3826	0.3876	0.3786	0.4038	0.3710	0.3478	0.3414	0.3260
	0.020	0.3178	0.2962	0.2780	0.2827	0.2684	0.3120	0.2560	0.2156	0.2122	0.1832
	0.030	0.2395	0.2100	0.1925	0.1965	0.1795	0.2241	0.1684	0.1266	0.1195	0.0983
	0.040	0.1770	0.1437	0.1288	0.1298	0.1135	0.1594	0.1087	0.0751	0.0670	0.0494
	0.050	0.1243	0.0946	0.0812	0.0833	0.0704	0.1176	0.0684	0.0427	0.0373	0.0282
YASS (12)	0.077	0.2009	0.1825	0.2324	0.2279	0.1268	0.0930	0.0532	0.0324	0.0303	0.0173
YASS (11)	0.114	0.1989	0.1585	0.2118	0.1573	0.0718	0.0701	0.0437	0.0349	0.0227	0.0111
YASS (8)	0.138	0.2520	0.1911	0.1886	0.1827	0.0742	0.0500	0.0271	0.0287	0.0178	0.0104
YASS (10)	0.159	0.2334	0.1476	0.1793	0.1341	0.0507	0.0370	0.0164	0.0210	0.0103	0.0054
YASS (3)	0.187	0.1277	0.0876	0.1301	0.0723	0.0224	0.0350	0.0146	0.0165	0.0081	0.0045
MME	0.050	0.4678	0.4546	0.4479	0.4492	0.4340	0.4427	0.4443	0.4424	0.4307	0.4194
	0.100	0.3001	0.2611	0.2574	0.2613	0.2501	0.3026	0.2466	0.2286	0.2091	0.1891
	0.150	0.2165	0.1735	0.1677	0.1721	0.1586	0.2299	0.1608	0.1404	0.1221	0.1027
	0.200	0.0217	0.0104	0.0127	0.0127	0.0124	0.0726	0.0153	0.0112	0.0080	0.0059
BCH	0.100	0.4599	0.4496	0.4448	0.4426	0.4390	0.4497	0.4290	0.4305	0.4229	0.4060
	0.200	0.3594	0.3124	0.3087	0.2974	0.2752	0.2958	0.2629	0.2707	0.2369	0.1946
	0.300	0.1383	0.0889	0.0862	0.0779	0.0697	0.0912	0.0663	0.0715	0.0536	0.0390
BCHopt	0.100	0.4726	0.4683	0.4558	0.4618	0.4595	0.4684	0.4550	0.4515	0.4480	0.4306
	0.200	0.4032	0.3712	0.3583	0.3548	0.3368	0.3517	0.3265	0.3253	0.3030	0.2582
	0.300	0.2400	0.1711	0.1719	0.1605	0.1356	0.1681	0.1289	0.1389	0.1102	0.0830

Table 1. Median testing error $\bar{P}_{\mathrm{E}}$ for six JPEG steganographic methods using different models. For easier navigation, the gray-level of the background in each row corresponds to the performance of individual feature sets: darker ⇒ better performance (lower error rate).

3.3 Discussion

Table 1 provides an important insight into the feature-building process and points to the following general guidelines for design of feature spaces for steganalysis:

- *High dimension is not sufficient for good performance.* This is clearly demonstrated by the rather poor performance of the $48,600$-dimensional CC-C300 feature set, often outperformed by the significantly lower-dimensional sets LIU, CC-PEV, and CC-CHEN. The failure of CC-C300 could be attributed to its lack of diversity (all co-occurrences are of the same type) and missing symmetrization, which make the model less robust and unnecessarily high-dimensional.
- *Calibration helps.* The positive effect of calibration has been demonstrated many times in the past, and here we confirm it by comparing the columns CHEN → CC-CHEN and JRM → CC-JRM. Notice that even for the high-dimensional JRM, the improvement may be substantial: $0.2707 \rightarrow 0.2369$ for BCH at 0.2 bpac and $0.1404 \rightarrow 0.1221$ for MME at 0.15 bpac. Moreover, researching alternative ways of calibration may bring additional improvements to feature-based steganalysis. This is indicated by a relatively good performance of the LIU feature set (compared to other low-dimensional sets), which utilizes two novel calibration principles: strenghtening the reference statistics by averaging over 63 different image croppings and calibrating by the *ratio* between original and reference features [14].
- *Steganalysis benefits from cross-domain models.* By combining CC-PEV with the spatial-domain SPAM features (CDF), sizeable improvement over CC-PEV is apparent across all steganographic methods. The

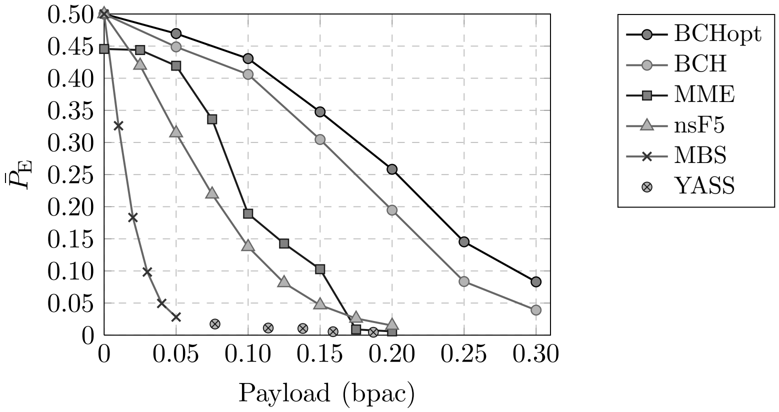

Figure 2. Testing error $\bar{P}_{\mathrm{E}}$ using the J+SRM feature space (dimension 35,263).

benefit of the multiple-domain approach is also clear from the last column – the union of the CC-JRM and the spatial domain rich model proposed in [4] further markedly improves the performance of CC-JRM and yields the lowest achieved error rates in all cases.

- *Future steganalysis will likely be driven by diverse and compact rich models.* The systematically constructed JPEG rich models $\mathcal{CF}^*$ and JRM/CC-JRM consistently outperform all low-dimensional sets. The superior performance of CC-JRM over $\mathcal{CF}^*$ is due to additional symmetrization, further diversification by co-occurrences of *differences*, and by its new integral components.

3.4 Comparison of steganographic methods

In Figure 2, we compare the performance of all tested stego schemes using the J+SRM feature set. We confirm that BCHopt is the most secure steganographic method, and its heuristic optimization brings improvement over BCH.

MBS and YASS are by far the least secure algorithms. The failure of YASS, already reported in [11, 14], suggests that embedding robustly in a different domain may not be the best approach for passive warden steganography as the robustness unavoidably yields significant and thus easily detectable distortion.

The nsF5 algorithm, which does not utilize any side information, is clearly outperformed by all schemes that utilize the knowledge of the uncompressed cover: MME, BCH, and BCHopt. The effect of this type of side information at the sender on steganographic security is, however, not well understood today. In particular, it is not clear how to utilize it in the best possible manner.

Let us conclude this section by commenting on two security artifacts of MME. First, we can see significant jumps in $\bar{P}_{\mathrm{E}}$ around payloads 0.09 and 0.16 bpac, which are due to suboptimal Hamming codes as already reported in [11]. This could be easily remedied by using more sophisticated coding schemes [2]. Second, note that the error at zero payload is $\bar{P}_{\mathrm{E}} \approx 0.45$ rather than random guessing. This is caused by the embedded message header whose size is independent of the message length and which is present in every stego image. We found that in case of MME, this message header is always embedded in the top left corner of the image, in many cases the area of sky, and thus creates statistically detectable traces. Even though this implementation flaw could be easily fixed, it illustrates that even the smallest implementation detail needs to be handled with caution when designing a practical steganographic scheme.

4. INVESTIGATIVE EXPERIMENTS

The purpose of this section is to study the contribution of the individual components of the CC-JRM to the overall performance. We also address the problem of finding a small subset of CC-JRM responsible for most of the detection accuracy for a fixed stego source. Our experiments are restricted only to selected steganographic methods and payloads. The notation follows Figure 1.

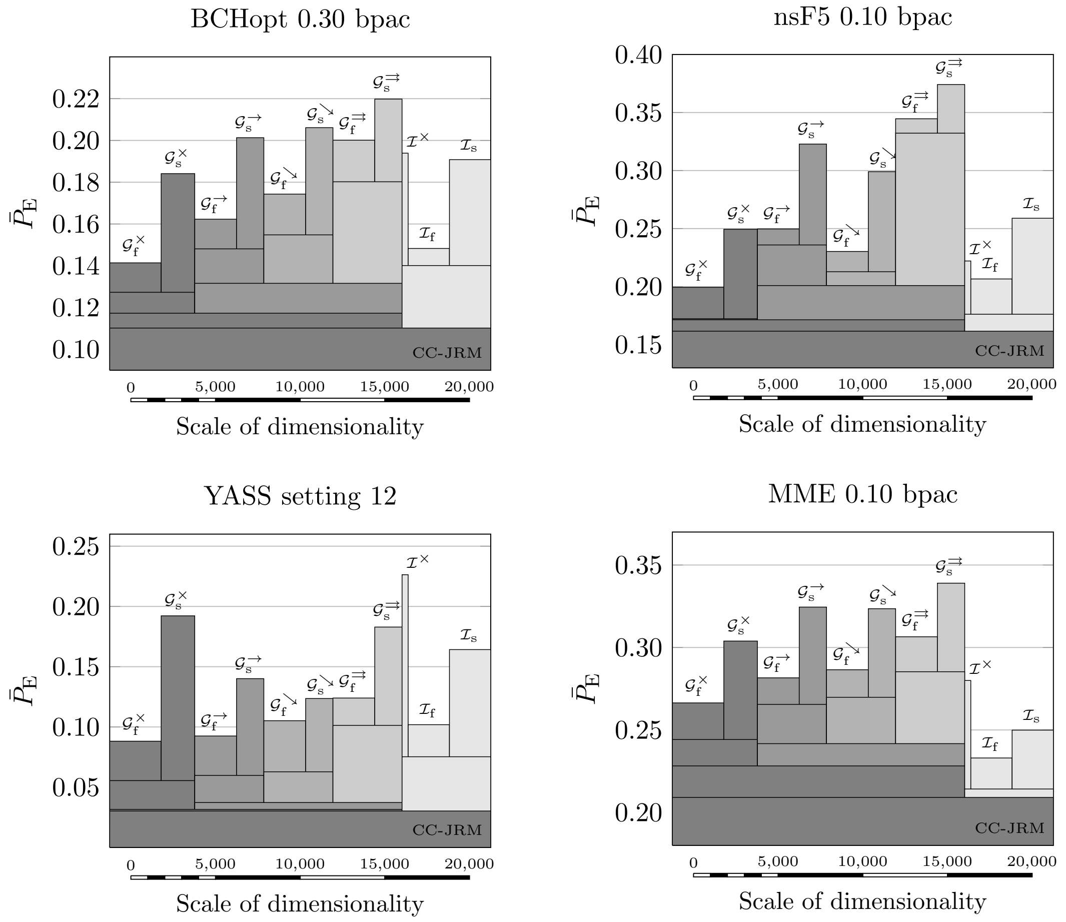

Figure 3. Systematic merging of the CC-JRM submodels and the progress of the testing error $\bar{P}_\mathrm{E}$. See Section 4.1 for explanation of the graphs and their interpretation.

4.1 Systematic merging of submodels

In the first experiment, we consider the following disjoint and qualitatively different subsets of the CC-JRM: $\mathcal{G}_\mathrm{f}^\times, \mathcal{G}_\mathrm{s}^\times, \mathcal{G}_\mathrm{f}^\rightarrow, \mathcal{G}_\mathrm{s}^\rightarrow, \mathcal{G}_\mathrm{f}^\searrow, \mathcal{G}_\mathrm{s}^\searrow, \mathcal{G}_\mathrm{f}^\rightrightarrows, \mathcal{G}_\mathrm{s}^\rightrightarrows, \mathcal{I}^\times, \mathcal{I}_\mathrm{f}, \mathcal{I}_\mathrm{s}$, and use them for steganalysis separately. Afterwards, we gradually and systematically merge them together, following the logic of Figure 1, until all of them are merged into the CC-JRM. All considered feature sets are Cartesian calibrated, yielding double the dimensionalities shown in Figure 1. The experiment was performed on the following steganographic schemes: BCHopt 0.30 bpac, nsF5 0.10 bpac, YASS setting 12, and MME 0.10 bpac. The training procedure was identical to the one used in the experiments of Section 3: training on a randomly selected half of the CAMERA database, testing on the other half. The obtained performance is reported in Figure 3 in terms of $\bar{P}_\mathrm{E}$.

In Figure 3, every submodel is represented by a bar whose height is the $\bar{P}_\mathrm{E}$. Conveniently, the width of each bar is proportional to the dimensionality of the corresponding submodel, allowing thus a continuous perception of the feature space sizes. For example, the rather thin bar of $\mathcal{I}^\times$ can be immediately perceived as more than five times smaller than the neighboring $\mathcal{I}_\mathrm{f}$. Intuitively, the union of several submodels is represented by an overlapping bar whose width is equal to the sum of its components. The overlapping bars do not interfere with the performance of their submodels because merging always decreases the error. For example, see the performance of submodels $\mathcal{G}_\mathrm{f}^\rightrightarrows$ and $\mathcal{G}_\mathrm{s}^\rightrightarrows$ in the top left graph (BCHopt). Their individual errors $\bar{P}_\mathrm{E}$ are 0.20 and 0.22, respectively, and their union (denoted $\mathcal{G}^\rightrightarrows$ in Figure 3) yields error 0.18, thence the corresponding height of the lower, wider bar. After adding additional submodels $\mathcal{G}_\mathrm{f}^\rightarrow, \mathcal{G}_\mathrm{s}^\rightarrow, \mathcal{G}_\mathrm{f}^\searrow, \mathcal{G}_\mathrm{s}^\searrow$, the error can be seen to drop further to roughly 0.13. The final performance of the CC-JRM is always represented by the lowest bar spanning the whole width of the graph. Finally, the readability is further improved by using different shades of gray for different types of submodels.

Figure 3 reveals interesting information about the types of features that are effective for attacking various steganographic algorithms. The four selected steganographic methods represent very different embedding paradigms, which is why the individual submodels contribute differently to the detection. The contribution of the integral features $\mathcal{I} = \{\mathcal{I}^{\times}, \mathcal{I}_{\mathrm{f}}, \mathcal{I}_{\mathrm{s}}\}$, for example, seems to be rather negligible for YASS because steganalysis *without* $\mathcal{I}$ delivers basically the same performance. For the other three algorithms, however, integral features noticeably improve the performance. This is most apparent for MME where the integral features $\mathcal{I}$ perform better than the rest of the features together despite their significantly lower dimensionality. As another example, compare the individual performance of the DCT-mode specific features extracted directly from absolute values of DCT coefficients, $\mathcal{G}^{\times} = \{\mathcal{G}_{\mathrm{f}}^{\times}, \mathcal{G}_{\mathrm{s}}^{\times}\}$, with the DCT-mode specific features extracted from the differences, $\mathcal{G}_{\mathrm{diff}} = \{\mathcal{G}_{\mathrm{f}}^{\rightarrow}, \mathcal{G}_{\mathrm{s}}^{\rightarrow}, \mathcal{G}_{\mathrm{f}}^{\searrow}, \mathcal{G}_{\mathrm{s}}^{\searrow}, \mathcal{G}_{\mathrm{f}}^{\rightrightarrows}, \mathcal{G}_{\mathrm{s}}^{\rightrightarrows}\}$. While for nsF5, $\mathcal{G}_{\mathrm{diff}}$ does not improve the performance of $\mathcal{G}^{\times}$ much, both seem to be important for the other three algorithms and especially for YASS.

We conclude that there is no subset of CC-JRM that is universally responsible for majority of detection accuracy across different steganographic schemes. The power of CC-JRM is in the *union* of its systematically built components, carefully designed to capture different types of statistical dependencies.

4.2 Forward feature selection

Despite its high dimension (22,510), ensemble classifiers make the training in the CC-JRM feature space computationally feasible. In fact, the bottleneck of steganalysis now becomes the feature extraction rather than the actual training of the classifier. To give the reader a better idea, we measured the running time needed for steganalysis of nsF5 at 0.10 bpac using the CC-JRM. The extraction of features from $6,500$ images took roughly 18 hours, while, on the same machine,[†] the classifier training took on average 5 minutes. From the practical point of view, the *testing* time may be an important factor – after the classifier is trained, the time needed to make decisions should be minimized. Although projecting the CC-JRM feature vector of the image under inspection into eigen-directions of individual FLDs of the ensemble classifier consists of a series of fast matrix multiplications, the extraction of the 22,510 complex features is quite costly. Therefore, one may want to consider investing more time into the training procedure, and perform a supervised feature selection in order to reduce the number of features needed to be extracted during testing, while keeping satisfactory performance. Note that we are interested specifically in feature selection rather than general dimensionality-reduction as the goal is to minimize the number of features needed.

Unfortunately, as shown in the previous investigative experiment in Figure 3, there is no compact subset of CC-JRM that would be universally effective against different types of embedding modifications. However, if we *fix* the steganographic channel, the problem becomes feasible. To demonstrate the feasibility of this direction, we performed a rather simple forward feature selection procedure applied to submodels (it was called the ITERATIVE-BEST strategy in [4]). It starts with all $N = 2 \times 51 = 102$ submodels of the CC-JRM.[‡] Once $k \geq 0$ submodels are selected, add the one that leads to the biggest drop in the OOB error estimate when all $k + 1$ submodels are used as a feature space. We use the out-of-bag (OOB) error estimation calculated from the training set [10] because no information about the testing images can be utilized. The procedure finishes after a pre-defined number of iterations or once the OOB values reach satisfactory values.

The ITERATIVE-BEST strategy greedily minimizes the OOB error at every iteration and takes the mutual dependencies among individual submodels into account. The ensemble classifier is used as a black box providing classification feedback. Such methods are known as wrappers [12].

We performed the ITERATIVE-BEST feature selection strategy on BCHopt 0.30 bpac, nsF5 0.10 bpac, YASS setting 12, and MME 0.10. The results are shown in Figure 4. The individual graphs show the progress of OOB error estimates for $k \leq 10$ as well as the list of the selected submodels. We follow the notation of the submodels

[†] Dell PowerEdge R710 with 12 cores and 48GB RAM when executed as a single process.

[‡] We treat the submodels and their reference submodels coming from Cartesian calibration separately.

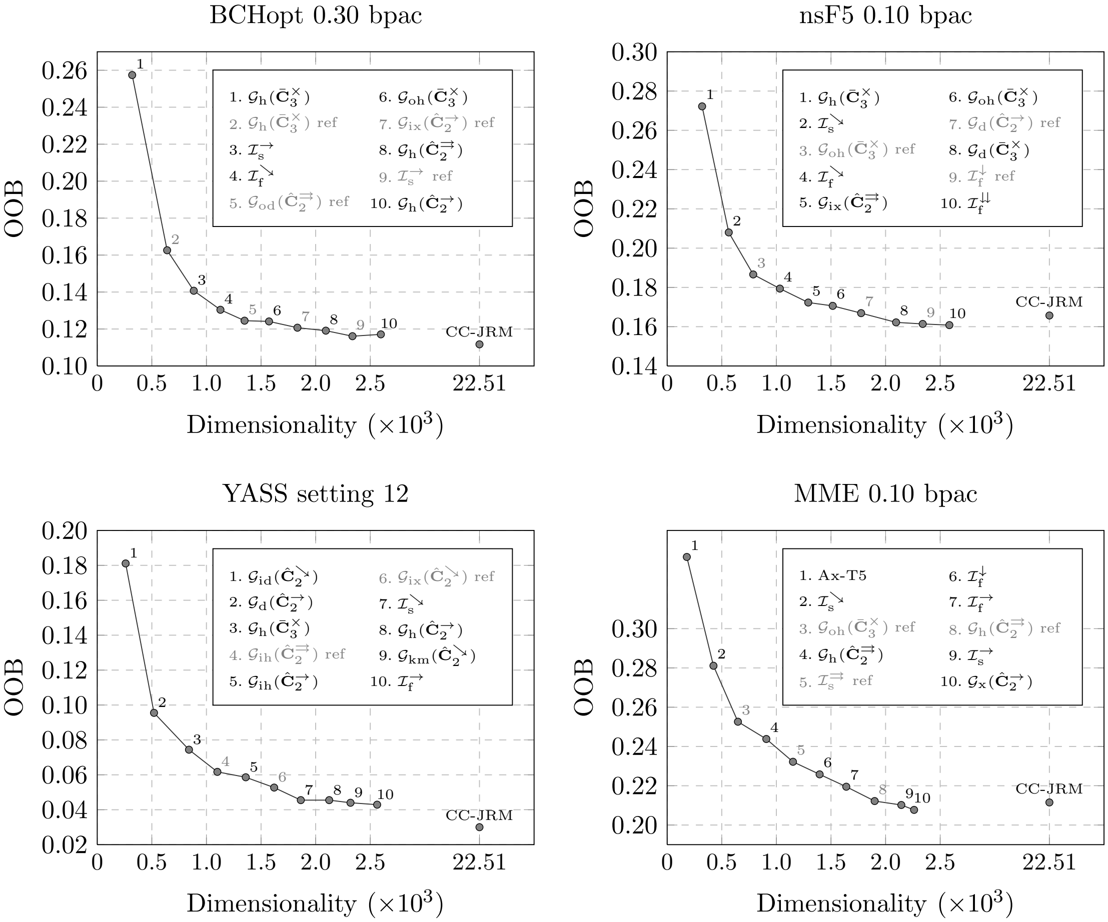

Figure 4. The result of the ITERATIVE-BEST feature selection strategy applied to four qualitatively different steganographic methods. The reported values are out-of-bag (OOB) error estimates calculated on the training set (half of the CAMERA database). For reference, we include the OOB error of the full CC-JRM feature space.

introduced in Figure 1 and distinguish the reference-version of the submodels by gray shade and adding the suffix "ref." For comparison, we also include the OOB-performance when the entire CC-JRM is used.

Figure 4 clearly demonstrates that it is indeed possible to obtain performance similar to CC-JRM with as few as one tenth of its submodels, reducing thus the testing time by one order of magnitude. The selected submodels are also generally different and algorithm-specific, which confirms our claim that there is no universally effective subset of CC-JRM.

The results provide us with another very interesting insight. Quite surprisingly, the appearance of a *reference* submodel often does not imply that the original version of the same submodel has been previously selected. In other words, a reference submodel may be useful as a complementary feature set to *other* types of features as well. Note, for example, the fourth selected submodel for YASS or the third for nsF5 and MME. This phenomenon indicates the intrinsic complexity of relationships among all extracted features and their reference values, and supports the hypothesis that appeared in [8]: individual features of complex feature spaces serve *each other* as references. For high-dimensional spaces, the concept of Cartesian calibration can thus be viewed simply as model enrichment that makes the feature space more diverse.

5. SUMMARY

Arguably, the most important element of today's feature-based steganalyzers is the feature-space design. In this paper, we follow the recent trend of constructing rich feature spaces consisting of many simple submodels, each of them capturing different types of dependencies among coefficients. We constructed a rich model of JPEG images, abbreviated as CC-JRM, and demonstrated its capability to detect a wide range of qualitatively different embedding schemes. When combined with a scalable machine learning, CC-JRM outperforms all previously published models of JPEG images.

Merging the JPEG domain rich model with the spatial domain rich model (SRM) recently proposed in [4] results in a 35,263-dimensional features space that further improves steganalysis across all six tested steganographic schemes and all tested payloads. This confirms the thesis that steganalysis benefits from multiple-domain approaches.

The experiments from Section 4 indicate that the proposed CC-JRM does not contain any universally effective subset that could replace CC-JRM while keeping its performance across different stegoschemes. However, if we are to construct a targeted steganalyzer for detection of a selected steganographic method, it is possible to significantly reduce the dimensionality by supervised feature selection.

The last experiment of Section 4 showed that reference features are often useful even without their original feature values, which sheds more light on the real benefit of Cartesian calibration in high dimensions.

Matlab implementation of all feature sets used in this paper is available at `http://dde.binghamton.edu/download/feature_extractors`.

6. ACKNOWLEDGMENTS

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9550-09-1-0147. 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.

REFERENCES

1. 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, May 2008.
2. T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 6:920–935, 2011.
3. J. Fridrich. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In J. Fridrich, editor, *Information Hiding, 6th International Workshop*, volume 3200 of Lecture Notes in Computer Science, pages 67–81, Toronto, Canada, May 23–25, 2004. Springer-Verlag, New York.
4. J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*. Under review.
5. 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.
6. Y. Kim, Z. Duric, and D. Richards. Modified matrix encoding technique for minimal distortion 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 314–327, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.
7. J. Kodovský. *Steganalysis of Digital Images Using Rich Image Representations and Ensemble Classifiers.* PhD thesis, Electrical and Computer Engineering Department, Binghamton University, NY, 2012.
8. J. Kodovský and J. Fridrich. Calibration revisited. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 63–74, Princeton, NJ, September 7–8, 2009.

9. J. Kodovský and J. Fridrich. Steganalysis in high dimensions: Fusing classifiers built on random subspaces. In N. D. Memon, E. J. Delp, P. W. Wong, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia XIII*, volume 7880, pages OL 1–13, San Francisco, CA, January 23–26, 2011.
10. J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, 2012. To appear.
11. J. Kodovský, T. Pevný, and J. Fridrich. Modern steganalysis can detect YASS. In N. D. Memon, E. J. Delp, P. W. Wong, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia XII*, volume 7541, pages 02–01–02–11, San Jose, CA, January 17–21, 2010.
12. T. N. Lal, O. Chapelle, J. Weston, and A. Elisseeff. Embedded methods. In I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, editors, *Feature Extraction: Foundations and Applications, Studies in Fuzziness and Soft Computing*, pages 137–165. Physica-Verlag, Springer, 2006.
13. B. Li, Y. Q. Shi, and J. Huang. Steganalysis of YASS. In A. D. Ker, J. Dittmann, and J. Fridrich, editors, *Proceedings of the 10th ACM Multimedia & Security Workshop*, pages 139–148, Oxford, UK, 2008.
14. 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.
15. Q. Liu, A. H. Sung, M. Qiao, Z. Chen, and B. Ribeiro. An improved approach to steganalysis of JPEG images. *Information Sciences*, 180(9):1643–1655, 2010.
16. T. Pevný, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 75–84, Princeton, NJ, September 7–8, 2009.
17. T. Pevný and J. Fridrich. Merging Markov and DCT features for multi-class JPEG steganalysis. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 3 1–3 14, San Jose, CA, January 29–February 1, 2007.
18. V. Sachnev, H. J. Kim, and R. Zhang. Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 131–140, Princeton, NJ, September 7–8, 2009.
19. P. Sallee. Model-based steganography. In T. Kalker, I. J. Cox, and Y. Man Ro, editors, *Digital Watermarking, 2nd International Workshop*, volume 2939 of Lecture Notes in Computer Science, pages 154–167, Seoul, Korea, October 20–22, 2003. Springer-Verlag, New York.
20. A. Sarkar, K. Solanki, and B. Manjunath. Obtaining higher rates for steganographic schemes while maintaining the same detectability. In R. Böhme and R. Safavi-Naini, editors, *Information Hiding, 12th International Workshop*, volume 6387 of Lecture Notes in Computer Science, pages 178–192, Calgary, Canada, June 28–30, 2010. Springer-Verlag, New York.
21. A. Sarkar, K. Solanki, and B. S. Manjunath. Further study on YASS: Steganography based on randomized embedding to resist blind steganalysis. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages 16–31, San Jose, CA, January 27–31, 2008.
22. 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.
23. K. Solanki, A. Sarkar, and B. S. Manjunath. YASS: Yet another steganographic scheme that resists blind steganalysis. In T. Furon, F. Cayre, G. Doërr, and P. Bas, editors, *Information Hiding, 9th International Workshop*, volume 4567 of Lecture Notes in Computer Science, pages 16–31, Saint Malo, France, June 11–13, 2007. Springer-Verlag, New York.
24. 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.