Low Complexity Features for JPEG Steganalysis Using Undecimated DCT

Vojtěch Holub and Jessica Fridrich, *Member, IEEE*

Abstract—This article introduces a novel feature set for steganalysis of JPEG images. The features are engineered as first-order statistics of quantized noise residuals obtained from the decompressed JPEG image using 64 kernels of the discrete cosine transform (the so-called undecimated DCT). This approach can be interpreted as a projection model in the JPEG domain, forming thus a counterpart to the projection spatial rich model. The most appealing aspect of this proposed steganalysis feature set is its low computational complexity, lower dimensionality in comparison to other rich models, and a competitive performance w.r.t. previously proposed JPEG domain steganalysis features.

I. INTRODUCTION

Steganalysis of JPEG images is an active and highly relevant research topic due to the ubiquitous presence of JPEG images on social networks, image sharing portals, and in Internet traffic in general. There exist numerous steganographic algorithms specifically designed for the JPEG domain. Such tools range from easy-to-use applications incorporating quite simplistic data hiding methods to advanced tools designed to avoid detection by a sophisticated adversary. According to the information provided by Wetstone Technologies, Inc, a company that keeps an up-to-date comprehensive list of all software applications capable of hiding data in electronic files, as of March 2014 a total of 349 applications that hide data in JPEG images were available for download.[1]

Historically, two different approaches to steganalysis have been developed. One can start by adopting a model for the statistical distribution of DCT coefficients in a JPEG file and design the detector using tools of statistical hypothesis testing [30], [34], [7]. In the second, much more common approach, a representation of the image (a feature) is identified that reacts sensitively to

The authors are with the Department of Electrical and Computer Engineering, Binghamton University, NY, 13902, USA. Email: vholub1,fridrich@binghamton.edu.

[1]Personal communication by Chet Hosmer, CEO of Wetstone Tech.

embedding but does not vary much due to image content. For some simple steganographic methods that introduce easily identifiable artifacts, such as Jsteg, it is often possible to identify a scalar feature – an estimate of the payload length [32], [33], [31], [4], [19]. More sophisticated embedding algorithms usually require higher-dimensional feature representation to obtain more accurate detection. In this case, the detector is typically built using machine learning through supervised training during which the classifier is presented with features of cover as well as stego images. Alternatively, the classifier can be trained that recognizes only cover images and marks all outliers as suspected stego images [26], [28]. Recently, Ker and Pevný proposed to shift the focus from identifying stego images to identifying "guilty actors," e.g., Facebook users, using unsupervised clustering over actors in the feature space [17]. Irrespectively of the chosen detection philosophy, the most important component of the detectors is the feature space – their detection accuracy is directly tied to the ability of the features to capture the steganographic embedding changes.

Selected examples of popular feature sets proposed for detection of steganography in JPEG images are the historically first image quality metric features [1], first-order statistics of wavelet coefficients [8], Markov features formed by sample intra-block conditional probabilities [29], inter- and intra-block co-occurrences of DCT coefficients [6], the PEV feature vector [27], inter and intra-block co-occurrences calibrated by difference and ratio [23], and the JPEG Rich Model (JRM) [20]. Among the more general techniques that were identified as improving the detection performance is the calibration by difference and Cartesian calibration [23], [18]. By inspecting the literature on features for steganalysis, one can observe a general trend – the features' dimensionality is increasing, a phenomenon elicited by developments in steganography. More sophisticated steganographic schemes avoid introducing easily detectable artifacts and more information is needed to obtain better detection. To address the increased complexity of detector training, simpler machine learning tools were proposed that better scale w.r.t. feature dimensionality, such as the FLD-ensemble [21] or the perceptron [25]. Even with more efficient classifiers, however, the obstacle that may prevent practical deployment of high-dimensional features is the time needed to extract the feature [3], [13], [22], [16].

In this article, we propose a novel feature set for JPEG steganalysis, which enjoys low complexity, relatively small dimension, yet provides competitive detection perfor-

mance across all tested JPEG steganographic algorithms. The features are built as histograms of residuals obtained using the basis patterns used in the DCT. The feature extraction thus requires computing mere 64 convolutions of the decompressed JPEG image with 64 8×8 kernels and forming histograms. The features can also be interpreted in the DCT domain, where their construction resembles the PSRM with non-random orthonormal projection vectors. Symmetries of these patterns are used to further compactify the features and make them better populated. The proposed features are called DCTR features (Discrete Cosine Transform Residual).

In the next section, we introduce the undecimated DCT, which is the first step in computing the DCTR features. Here, we explain the essential properties of the undecimated DCT and point out its relationship to calibration and other previous art. The complete description of the proposed DCTR feature set as well as experiments aimed at determining the free parameters appear in Section III. In Section IV, we report the detection accuracy of the DCTR feature set on selected JPEG domain steganographic algorithms. The results are contrasted with the performance obtained using current state-of-the-art rich feature sets, including the JPEG Rich Model and the Projection Spatial Rich Model. The paper is concluded in Section V, where we discuss future directions.

A condensed version of this paper was submitted to the IEEE Workshop on Information Security and Forensics (WIFS) 2014.

II. UNDECIMATED DCT

In this section, we describe the undecimated DCT and study its properties relevant for building the DCTR feature set in the next section. Since the vast majority of steganographic schemes embed data only in the luminance component, we limit the scope of this paper to grayscale JPEG images. For easier exposition, we will also assume that the size of all images is a multiple of 8.

A. Description

Given an $M \times N$ grayscale image $\mathbf{X} \in \mathbb{R}^{M \times N}$, the undecimated DCT is defined as a set of 64 convolutions with 64 DCT basis patterns $\mathbf{B}^{(k,l)}$:

$$\mathcal{U}(\mathbf{X}) = \{\mathbf{U}^{(k,l)} | 0 \leq k, l \leq 7\} \tag{1}$$
$$\mathbf{U}^{(k,l)} = \mathbf{X} \star \mathbf{B}^{(k,l)},$$

where $\mathbf{U}^{(k,l)} \in \mathbb{R}^{(M-7) \times (N-7)}$ and $'\star'$ denotes a convolution without padding. The DCT basis patterns are 8×8 matrices, $\mathbf{B}^{(k,l)} = (B_{mn}^{(k,l)})$, $0 \leq m, n \leq 7$:

$$B_{mn}^{(k,l)} = \frac{\mathbf{w}_k \mathbf{w}_l}{4} \cos \frac{\pi k (2m+1)}{16} \cos \frac{\pi l (2n+1)}{16}, \tag{2}$$

and $\mathbf{w}_0 = 1/\sqrt{2}$, $\mathbf{w}_k = 1$ for $k > 0$.

When the image is stored in the JPEG format, before computing its undecimated DCT it is first decompressed to the spatial domain without quantizing the pixel values to $\{0, \ldots, 255\}$ to avoid any loss of information.

For better readability, from now on we will reserve the indices i, j and k, l to index DCT modes (spatial frequencies); they will always be in the range $0 \leq i, j, k, l \leq 7$.

1) Relationship to prior art: The undecimated DCT has already found applications in steganalysis. The concept of calibration, for the first time introduced in the targeted quantitative attack on the F5 algorithm [9], formally consists of computing the undecimated DTC, subsampling it on an 8×8 grid shifted by four pixels in each direction, and computing a reference feature vector from the subsampled and quantized signal. Liu [23] made use of the entire transform by computing 63 inter- and intra-block 2D co-occurrences from all possible JPEG grid shifts and *averaging* them to form a more powerful reference feature that was used for calibration by difference and by ratio. In contrast, in this paper we avoid using the undecimated DCT to form a reference feature, and, instead keep the statistics collected from all shifts *separated*.

B. Properties

First, notice that when subsampling the convolution $\mathbf{U}^{(i,j)} = \mathbf{X} \star \mathbf{B}^{(i,j)}$ on the grid $\mathcal{G}_{8 \times 8} = \{0, 7, 15, \ldots, M - 9\} \times \{0, 7, 15, \ldots, N - 9\}$ (circles in Figure 1 on the left), one obtains all unquantized values of DCT coefficients for DCT mode (i, j) that form the input into the JPEG representation of $\mathbf{X}$.

We will now take a look at how the values of the undecimated DCT $\mathcal{U}(\mathbf{X})$ are affected by changing one DCT coefficient of the JPEG representation of $\mathbf{X}$. Suppose one modifies a DCT coefficient in mode (k, l) in the JPEG file corresponding to $(m, n) \in \mathcal{G}_{8 \times 8}$. This change will affect all 8×8 pixels in the corresponding block and an entire 15×15 neighborhood of values in $\mathbf{U}^{(i,j)}$ centered at $(m, n) \in \mathcal{G}_{8 \times 8}$. In particular, the values will be modified by what we call the "unit response"

$$\mathbf{R}^{(i,j)(k,l)} = \mathbf{B}^{(i,j)} \otimes \mathbf{B}^{(k,l)}, \tag{3}$$

where $\otimes$ denotes the full cross-correlation. While this unit response is not symmetrical, its absolute values are symmetrical by both axes: $|\mathbf{R}_{a,b}^{(i,j)(k,l)}| = |\mathbf{R}_{-a,b}^{(i,j)(k,l)}|$, $|\mathbf{R}_{a,b}^{(i,j)(k,l)}| = |\mathbf{R}_{a,-b}^{(i,j)(k,l)}|$ for all $0 \leq a, b \leq 7$ when indexing $\mathbf{R} \in \mathbb{R}^{15 \times 15}$ with indices in $\{-7, \ldots, -1, 0, 1, \ldots, 7\}$.

Figure 2 shows two examples of unit responses. Note that the value at the center $(0, 0)$ is zero for the response on the left and 1 for the response on the right. This central value equals to 1 only when $i = k$ and $j = l$.

We now take a closer look at how a particular value $u \in \mathbf{U}^{(i,j)}$ is computed. First, we identify the four neighbors from the grid $\mathcal{G}_{8 \times 8}$ that are closest to u (follow Figure 1 where the location of u is marked by a triangle). We will capture the position of u w.r.t. to its four closest neighbors from $\mathcal{G}_{8 \times 8}$ using relative coordinates. With respect to the upper left neighbor (A), u is at position (a, b), $0 \leq a, b, \leq 7$ $((a, b) = (3, 2)$ in Figure 1). The relative positions w.r.t.
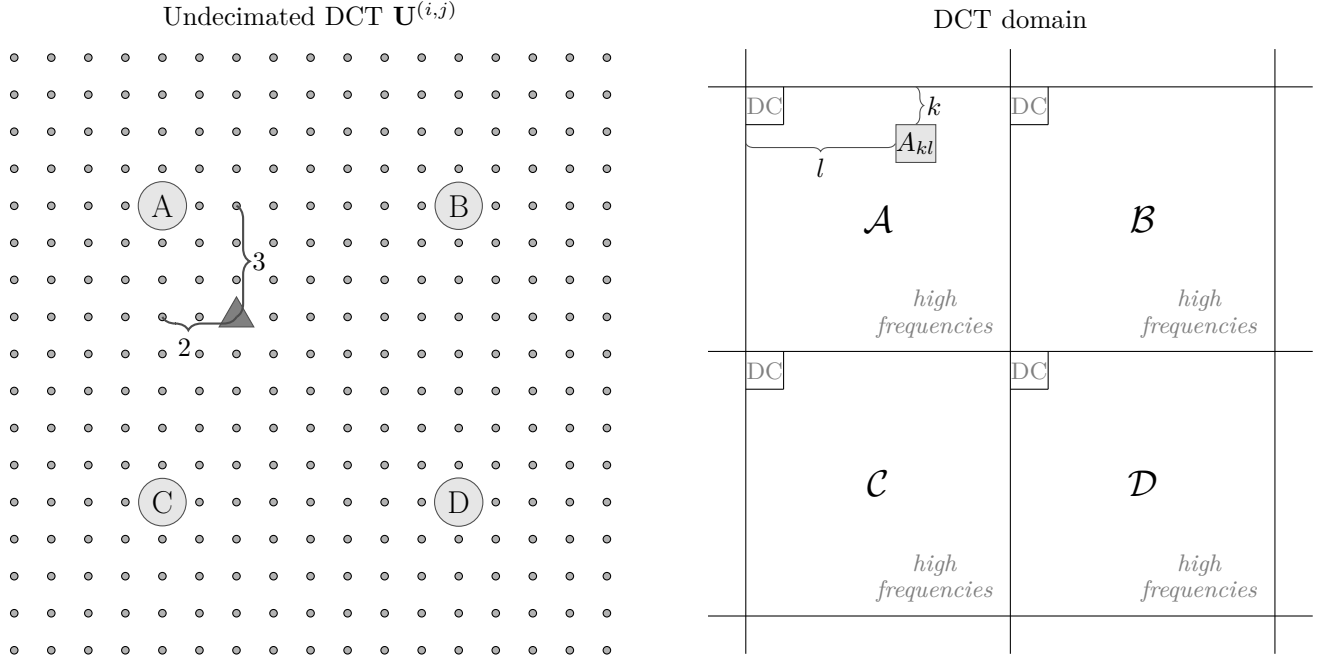
Figure 1. Left: Dots correspond to elements of $\mathbf{U}^{(i,j)} = \mathbf{X} \star \mathbf{B}^{(i,j)}$, circles correspond to grid points from $\mathcal{G}_{8\times8}$ (DCT coefficients in the JPEG representation of $\mathbf{X}$). The triangle is an element $u \in \mathbf{U}^{(i,j)}$ with relative coordinates $(a,b) = (3,2)$ w.r.t. its upper left neighbor (A) from $\mathcal{G}_{8\times8}$. Right: JPEG representation of $\mathbf{X}$ when replacing each 8×8 pixel block with a block of quantized DCT coefficients.

$$\mathbf{R}^{(1,3)(2,2)} \qquad \mathbf{R}^{(1,2)(1,2)}$$
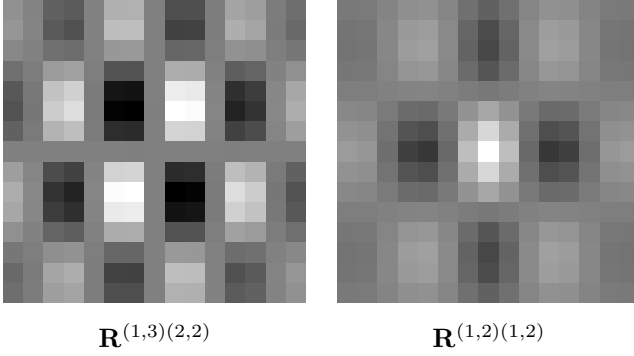
Figure 2. Examples of two unit responses scaled so that medium gray corresponds to zero.

the other three neighbors (B–D) are, correspondingly, $(a, b-8)$, $(a-8, b)$, and $(a-8, b-8)$. Also recall that the elements of $\mathbf{U}^{(i,j)}$ collected across all (i,j), $0 \leq i, j \leq 7$, at A, form all non-quantized DCT coefficients corresponding to the 8×8 block $\mathcal{A}$ (see, again Figure 1).

Arranging the DCT coefficients from the neighboring blocks $\mathcal{A}$–$\mathcal{D}$ into 8×8 matrices A_{kl}, B_{kl}, C_{kl} and D_{kl}, where k and l denote the horizontal and vertical spatial frequencies in the 8×8 DCT block, respectively, $u \in \mathbf{U}^{(i,j)}$ can be expressed as

$$u = \sum_{k=0}^{7} \sum_{l=0}^{7} Q_{kl} \left[A_{kl} R_{a,b}^{(i,j)(k,l)} + B_{kl} R_{a,b-8}^{(i,j)(k,l)} \right. $$
$$\left. + C_{kl} R_{a-8,b}^{(i,j)(k,l)} + D_{kl} R_{a-8,b-8}^{(i,j)(k,l)} \right], \quad (4)$$

where the subscripts in $R_{a,b}^{(i,j)(k,l)}$ capture the position of u

w.r.t. its upper left neighbor and Q_{kl} is the quantization step of the (k,l)-th DCT mode. This can be written as a projection of 256 dequantized DCT coefficients from four adjacent blocks from the JPEG file with a projection vector $\mathbf{p}_{a,b}^{(i,j)}$

$$u = \begin{pmatrix} Q_{00}A_{00} \\ \vdots \\ Q_{77}A_{77} \\ Q_{00}B_{00} \\ \vdots \\ Q_{77}B_{77} \\ \vdots \\ Q_{00}D_{00} \\ \vdots \\ Q_{77}D_{77} \end{pmatrix}^{T} \cdot \underbrace{\begin{pmatrix} R_{a,b}^{(i,j)(0,0)} \\ \vdots \\ R_{a,b}^{(i,j)(7,7)} \\ R_{a,b-8}^{(i,j)(0,0)} \\ \vdots \\ R_{a,b-8}^{(i,j)(7,7)} \\ \vdots \\ R_{a-8,b-8}^{(i,j)(0,0)} \\ \vdots \\ R_{a-8,b-8}^{(i,j)(7,7)} \end{pmatrix}}_{\mathbf{p}_{a,b}^{(i,j)}}. \quad (5)$$

It is proved in Appendix A that the projection vectors form an orthonormal system satisfying for all (a,b), (i,j), and (k,l)

$$\mathbf{p}_{a,b}^{(i,j)T} \cdot \mathbf{p}_{a,b}^{(k,l)} = \delta_{(i,j),(k,l)}, \quad (6)$$

where δ is the Kronecker delta. Projection vectors that are too correlated (in the extreme case, linearly dependent) would lead to undesirable redundancy (near duplication)

of feature elements. Orthonormal (uncorrelated) projection vectors increase features' diversity and provide better dimensionality-to-detection ratio.

The projection vectors also satisfy the following symmetry

$$\left|\mathbf{p}_{a,b}^{(i,j)}\right| = \left|\mathbf{p}_{a,b-8}^{(i,j)}\right| = \left|\mathbf{p}_{a-8,b}^{(i,j)}\right| = \left|\mathbf{p}_{a-8,b-8}^{(i,j)}\right| \qquad (7)$$

for all i, j and a, b when interpreting the arithmetic operations on indices as mod 8.

III. DCTR Features

The DCTR features are built by quantizing the absolute values of all elements in the undecimated DCT and collecting the first-order statistic separately for each mode (k, l) and each relative position (a, b), $0 \leq a, b \leq 7$. Formally, for each (k, l) we define the matrix[2] $\mathbf{U}_{a,b}^{(k,l)} \in \mathbb{R}^{(M-8)/8 \times (N-8)/8}$ as a submatrix of $\mathbf{U}^{(k,l)}$ with elements whose relative coordinates w.r.t. the upper left neighbor in the grid $\mathcal{G}_{8 \times 8}$ are (a, b). Thus, each $\mathbf{U}^{(k,l)} = \cup_{a,b=0}^{7} \mathbf{U}_{a,b}^{(k,l)}$ and $\mathbf{U}_{a,b}^{(k,l)} \cap \mathbf{U}_{a',b'}^{(k,l)} = \emptyset$ whenever $(a, b) \neq (a', b')$. The feature vector is formed by normalized histograms for $0 \leq k, l \leq 7$, $0 \leq a, b \leq 7$:

$$\mathbf{h}_{a,b}^{(k,l)}(r) = \frac{1}{\left|\mathbf{U}_{a,b}^{(k,l)}\right|} \sum_{u \in \mathbf{U}_{a,b}^{(k,l)}} [Q_T(|u|/q) = r], \qquad (8)$$

where Q_T is a quantizer with integer centroids $\{0, 1, \ldots, T\}$, q is the quantization step, and $[P]$ is the Iverson bracket equal to 0 when the statement P is false and 1 when P is true. We note that q could potentially depend on a, b as well as the DCT mode indices k, l, and the JPEG quality factor (see Section III-D for more discussions).

Because $\mathbf{U}^{(k,l)} = \mathbf{X} \star \mathbf{B}^{(k,l)}$ and the sum of all elements of $\mathbf{B}^{(k,l)}$ is zero (they are DCT modes (2)) each $\mathbf{U}^{(k,l)}$ is an output of a high-pass filter applied to $\mathbf{X}$. For natural images $\mathbf{X}$, the distribution of $u \in \mathbf{U}_{a,b}^{(k,l)}$ will thus be approximately symmetrical and centered at 0 for all a, b, which allows us to work with absolute values of $u \in \mathbf{U}_{a,b}^{(k,l)}$ giving the features a lower dimension and making them better populated.

Due to the symmetries of projection vectors (7), it is possible to further decrease the feature dimensionality by adding together the histograms corresponding to indices (a, b), $(a, 8-b)$, $(8-a, b)$, and $(8-a, 8-b)$ under the condition that these indices stay within $\{0, \ldots, 7\} \times \{0, \ldots, 7\}$ (see Table I). Note that for $(a, b) \in \{1, 2, 3, 5, 6, 7\}^2$, we merge four histograms. When exactly one element of (a, b) is in $\{0, 4\}$, only two histograms are merged, and when both a and b are in $\{0, 4\}$ there is only one histogram. Thus, the total dimensionality of the symmetrized feature vector is $64 \times (36/4 + 24/2 + 4) \times (T + 1) = 1600 \times (T + 1)$.

In the rest of this section, we provide experimental evidence that working with absolute values and symmetrizing

---

[2] Since $\mathbf{U}^{(k,l)} \in \mathbb{R}^{(M-7) \times (N-7)}$, the height (width) of $\mathbf{U}_{a,b}^{(k,l)}$ is larger by one when $a = 0$ ($b = 0$).

Table I

HISTOGRAMS $\mathbf{h}_{a,b}$ TO BE MERGED ARE LABELED WITH THE SAME LETTER. ALL 64 HISTOGRAMS CAN THUS BE MERGED INTO 25. LIGHT SHADING DENOTES MERGING OF FOUR HISTOGRAMS, MEDIUM SHADING TWO HISTOGRAMS, AND DARK SHADING DENOTES NO MERGING.

$a \backslash b$	0	1	2	3	4	5	6	7
0	a	b	c	d	e	d	c	b
1	e	f	g	h	i	h	g	f
2	j	k	l	m	n	m	l	k
3	o	p	q	r	s	r	q	p
4	t	u	v	w	x	w	v	u
5	o	p	q	r	s	r	q	p
6	j	k	l	m	n	m	l	k
7	e	f	g	h	i	h	g	f

the features indeed improves the detection accuracy. We also experimentally determine the proper values of the threshold T and the quantization step q, and evaluate the performance of different parts of the DCTR feature vector w.r.t. the DCT mode indices k, l.

A. Experimental setup

All experiments in this section are carried out on BOSSbase 1.01 [2] containing 10,000 grayscale 512×512 images. All detectors were trained as binary classifiers implemented using the FLD ensemble [21] with default settings available from http://dde.binghamton.edu/download/ensemble. As described in the original publication [21], the ensemble by default minimizes the total classification error probability under equal priors P_E. The random subspace dimensionality and the number of base learners is found by minimizing the out-of-bag (OOB) estimate of the testing error, E_{OOB}, on bootstrap samples of the training set. We also use E_{OOB} to report the detection performance since it is an unbiased estimate of the testing error on unseen data [5]. For experiments in Sections III-B–III-E, the steganographic method was J-UNIWARD at 0.4 bit per non-zero AC DCT coefficient (bpnzAC) with JPEG quality factor 75. We selected this steganographic method as an example of a state-of-the-art data hiding method for the JPEG domain.

B. Symmetrization validation

In this section, we experimentally validate the feature symmetrization. We denote by $E_{\text{OOB}}(X)$ the OOB error obtained when using features X. The histograms concatenated over the DCT mode indices will be denoted as

$$\mathbf{h}_{a,b} = \bigvee_{k,l=0}^{7} \mathbf{h}_{a,b}^{(k,l)}. \qquad (9)$$

For every combination of indices $a, b, c, d \in \{0, \ldots, 7\}^2$, we computed three types of error (the symbol '&' means feature concatenation):

1) $E_{a,b}^{\text{Single}} \triangleq E_{\text{OOB}}(\mathbf{h}_{a,b})$

Table II

$E_{a,b}^{\text{Single}}$ is the detection OOB error when steganalyzing with $\mathbf{h}_{a,b}$.

$a\backslash b$	0	1	2	3	4	5	6	7
0	0.427	0.343	0.298	0.336	0.304	0.335	0.298	0.345
1	0.366	0.409	0.349	0.367	0.340	0.370	0.352	0.408
2	0.335	0.372	0.338	0.345	0.327	0.344	0.343	0.371
3	0.358	0.378	0.339	0.347	0.326	0.356	0.336	0.377
4	0.334	0.348	0.319	0.328	0.310	0.325	0.323	0.351
5	0.358	0.379	0.335	0.350	0.326	0.352	0.340	0.379
6	0.335	0.374	0.340	0.347	0.324	0.346	0.340	0.372
7	0.369	0.404	0.348	0.365	0.334	0.361	0.348	0.404

Table III

$E_{(a,b),(c,d)}^{\text{Merged}} - E_{(a,b),(c,d)}^{\text{Concat}}$ for (a,b) as a function of (c,d).

$(a,b) = (1,2)$

$c\backslash d$	0	1	2	3	4	5	6	7
0	0.039	0.054	0.031	0.067	0.046	0.063	0.030	0.048
1	0.059	0.050	0	0.058	0.035	0.059	0.001	0.046
2	0.074	0.067	0.033	0.071	0.057	0.071	0.032	0.065
3	0.055	0.053	0.030	0.061	0.044	0.059	0.019	0.050
4	0.055	0.045	0.024	0.060	0.044	0.058	0.024	0.050
5	0.059	0.058	0.023	0.060	0.044	0.064	0.022	0.055
6	0.070	0.064	0.021	0.068	0.048	0.067	0.025	0.057
7	0.052	0.049	0.002	0.056	0.037	0.056	0.000	0.043

Table IV

$E_{\text{OOB}}(\mathbf{h}^{(k,l)})$ as a function of k,l.

	0	1	2	3	4	5	6	7
0	0.483	0.473	0.449	0.411	0.370	0.387	0.395	0.414
1	0.479	0.455	0.427	0.394	0.365	0.385	0.395	0.421
2	0.459	0.440	0.4220	0.398	0.392	0.397	0.405	0.424
3	0.446	0.420	0.414	0.421	0.426	0.428	0.427	0.431
4	0.419	0.403	0.406	0.423	0.432	0.443	0.438	0.438
5	0.407	0.399	0.407	0.428	0.445	0.453	0.451	0.440
6	0.406	0.402	0.410	0.428	0.448	0.460	0.446	0.427
7	0.402	0.422	0.423	0.434	0.435	0.439	0.434	0.433

for J-UNIWARD the histograms corresponding to high frequency modes provide the same or better distinguishing power than those of low frequencies.

D. Feature quantization and normalization

In this section, we investigate the effect of quantization and feature normalization on the detection performance.

We carried out experiments for two quality factors, 75 and 95, and studied the effect of the quantization step q on detection accuracy (the two top charts in Figure 3). Additionally, we also investigated whether it is advantageous, prior to quantization, to normalize the features by the DCT mode quantization step, Q_{kl}, and by scaling $\mathbf{U}^{(l)}$ to a zero mean and unit variance (the two bottom charts in Figure 3).

Figure 3 shows that the effect of feature normalization is quite weak and it appears to be slightly more advantageous to not normalize the features and keep the feature design simple. The effect of the quantization step q is, however, much stronger. For quality factor 75 (95), the optimal quantization steps were 4 (0.8). Thus, we opted for the following linear fit[3] to obtain the proper value of q for an arbitrary quality factor in the range $50 \le K \le 99$:

$$q_K = 8 \times \left(2 - \frac{K}{50} \right). \qquad (10)$$

E. Threshold

As Table V shows, the detection performance is quite insensitive to the threshold T. Although the best performance is achieved with $T = 6$, the gain is negligible compared to the dimensionality increase. Thus, in this paper we opted for $T = 4$ as a good compromise between performance and detectability.

To summarize, the final form of DCTR features includes the symmetrization as explained in Section III, no normalization, quantization according to (10), and $T = 4$. This gives the DCTR set the dimensionality of 8,000.

---

[3]Coincidentally, the term in the bracket corresponds to the multiplier used for computing standard quantization matrices.

2) $E_{(a,b),(c,d)}^{\text{Concat}} \triangleq E_{\text{OOB}}(\mathbf{h}_{a,b} \vee \mathbf{h}_{c,d})$

3) $E_{(a,b),(c,d)}^{\text{Merged}} \triangleq E_{\text{OOB}}(\mathbf{h}_{a,b} + \mathbf{h}_{c,d})$

to see the individual performance of the features across the relative indices (a,b) as well as the impact of concatenating and merging the features on detectability. In the following experiments, we fixed $q = 4$ and $T = 4$. This gave each feature $\mathbf{h}_{a,b}$ the dimensionality of $64 \times (T+1) = 320$ (the number of JPEG modes, 64, times the number of quantization bins $T + 1 = 5$).

Table II informs us about the individual performance of features $\mathbf{h}_{a,b}$. Despite the rather low dimensionality of 320, every $\mathbf{h}_{a,b}$ achieves a decent detection rate by itself (c.f., Figure 4 in Section IV).

The next experiment was aimed at assessing the loss of detection accuracy when merging histograms corresponding to different relative coordinates as opposed to concatenating them. When this drop of accuracy is approximately zero, both feature sets can be merged. Table III shows the detection drop $E_{(a,b),(c,d)}^{\text{Merged}} - E_{(a,b),(c,d)}^{\text{Concat}}$ when merging $\mathbf{h}_{1,2}$ with $\mathbf{h}_{c,d}$ as a function of c,d. The results clearly show which features should be merged; they are also consistent with the symmetries analyzed in Section II-B.

C. Mode performance analysis

In this section, we analyze the performance of the DCTR features by DCT modes when steganalyzing with the merger $\mathbf{h}^{(k,l)} \triangleq \sum_{a,b=0}^{7} \mathbf{h}_{a,b}^{(k,l)}$ of dimension $25 \times (T+1) = 125$. Table I explains why the total number of histograms can be reduced from 64 to 25 by merging histograms for different shifts a, b. Interestingly, as Table IV shows,
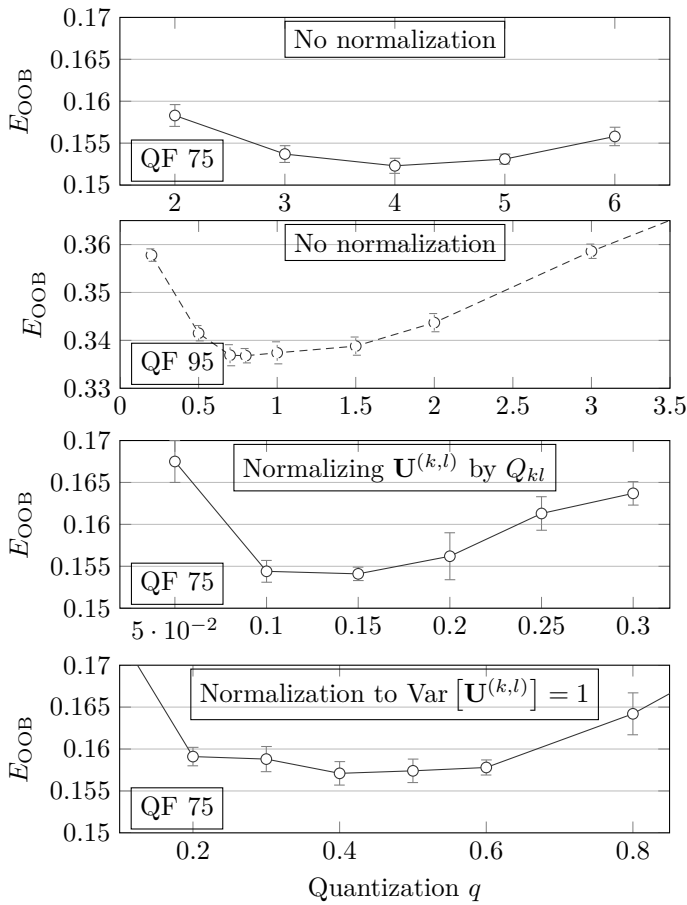
Figure 3. The effect of feature quantization without normalization (top charts) and with normalization (bottom charts) on detection accuracy.

Table V
E_{OOB} of the entire DCTR feature set with dimensionality $1600 \times (T+1)$ as a function of the threshold T for J-UNIWARD at 0.4 bpnzAC.

T	3	4	5	6
E_{OOB}	0.1545	0.1523	0.1524	0.1519

IV. Experiments

In this section, we subject the newly proposed DCTR feature set to tests on selected state-of-the-art JPEG steganographic schemes as well as examples of older embedding schemes. Additionally, we contrast the detection performance to previously proposed feature sets. Each time a separate classifier is trained for each image source, embedding method, and payload to see the performance differences.

Figures 4, 5 and 6 show the detection error E_{OOB} for J-UNIWARD [14], ternary-coded UED (Uniform Embedding Distortion) [12], and nsF5 [11] achieved using the proposed DCTR, the JPEG Rich Model (JRM) [20] of dimension 22,510, the 12,753-dimensional version of the Spatial Rich Model called SRMQ1 [10], the merger of JRM
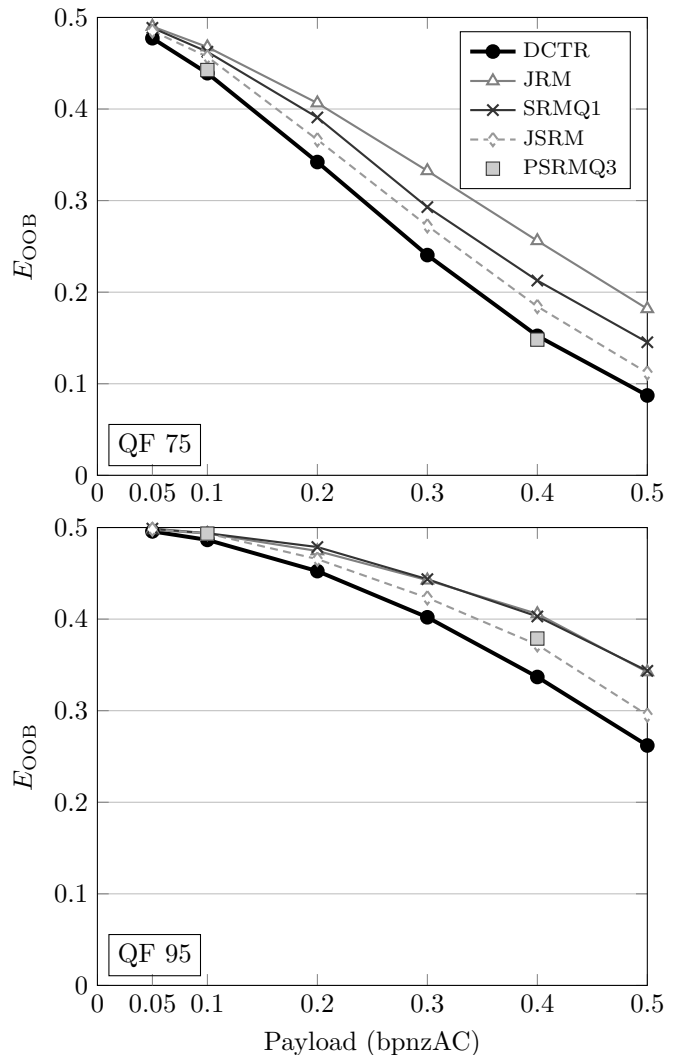

Figure 4. Detection error E_{OOB} for J-UNIWARD for quality factors 75 and 95 when steganalyzed with the proposed DCTR and other rich feature sets.

and SRMQ1 abbreviated as JSRM (dimension 35,263), and the 12,870 dimensional Projection Spatial Rich Model with quantization step 3 specially designed for the JPEG domain (PSRMQ3) [13]. When interpreting the results, one needs to take into account the fact that the DCTR has by far the lowest dimensionality and computational complexity of all tested feature sets.

The most significant improvement is seen for J-UNIWARD, even though it remains very difficult to detect. Despite its compactness and a significantly lower computational complexity, the DCTR set is the best performer for the higher quality factor and provides about the same level of detection as PSRMQ3 for quality factor 75. For the ternary UED, the DCTR is the best performer for the higher JPEG quality factor for all but the largest tested payload. For quality factor 75, the much larger 35,263-dimensional JSRM gives a slightly better detection. The DCTR also provides quite competitive detection for nsF5. The detection accuracy is roughly at the same level
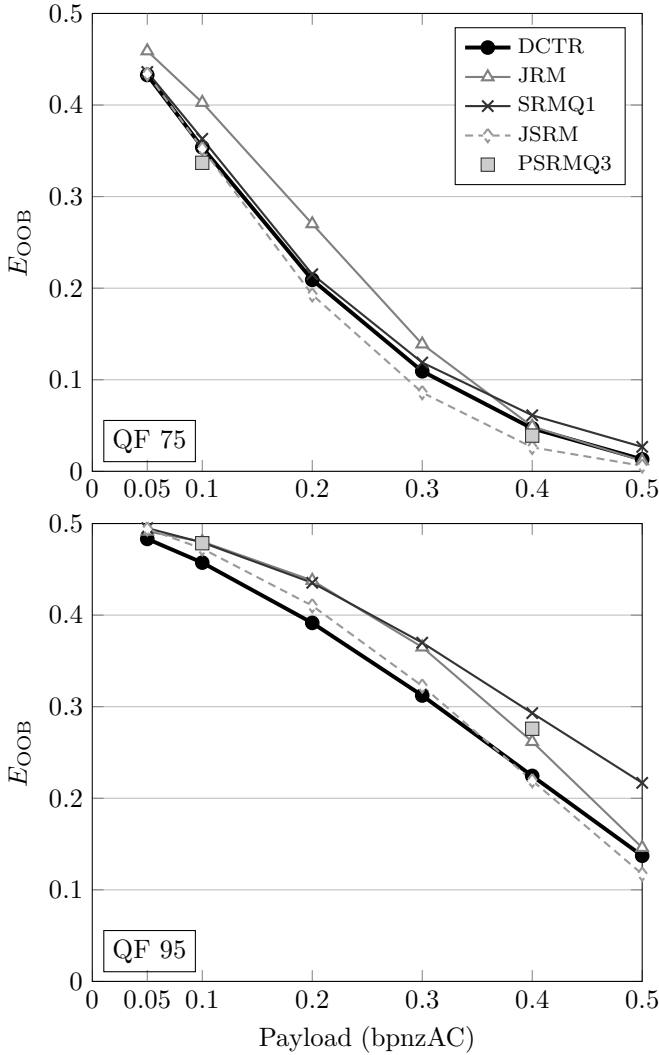
Figure 5. Detection error E_{OOB} for UED with ternary embedding for quality factors 75 and 95 when steganalyzed with the proposed DCTR and other rich feature sets.
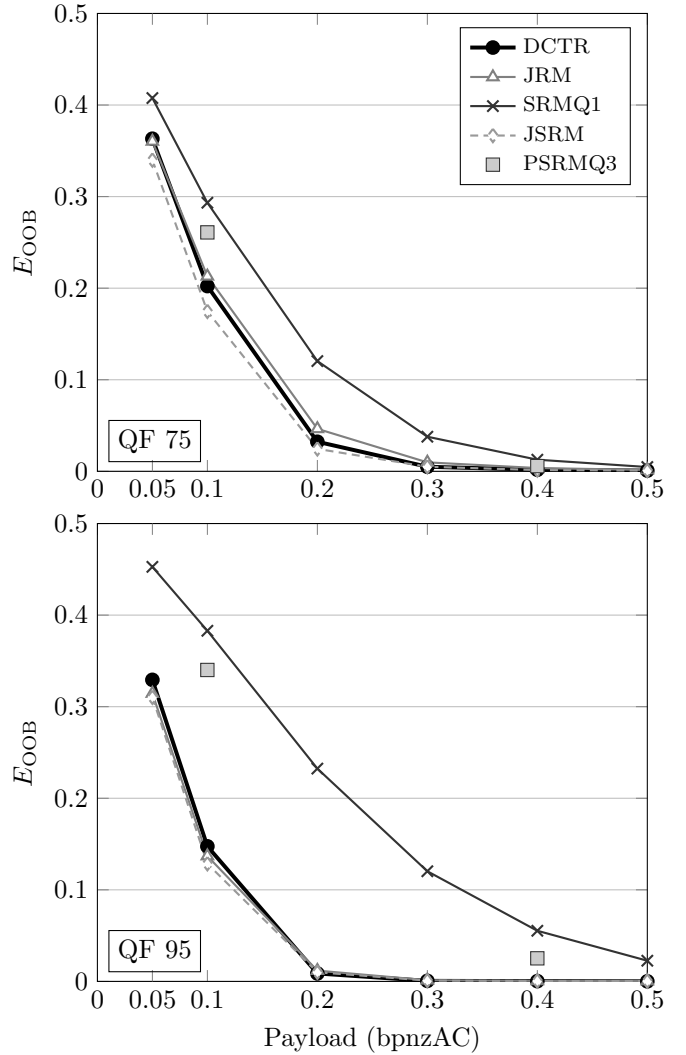

Figure 6. Detection error E_{OOB} for nsF5 for quality factors 75 and 95 when steganalyzed with the proposed DCTR and other rich feature sets.

as for the 22,510-dimensional JRM.

The DCTR feature set is also performing quite well against the state-of-the-art side-informed JPEG algorithm SI-UNIWARD [14] (Figure 7). On the other hand, JSRM and JRM are better suited to detect NPQ [15] (Figure 8). This is likely because NPQ introduces (weak) embedding artifacts into the statistics of JPEG coefficients that are easier to detect by the JRM, whose features are entirely built as co-occurrences of JPEG coefficients. We also point out the saturation of the detection error below 0.5 for quality factor 95 and small payloads for both schemes. This phenomenon, which was explained in [14], is caused by the tendency of both algorithms to place embedding changes into four specific DCT coefficients.

In Table VI, we take a look at how complementary the DCTR features are in comparison to the other rich models. This experiment was run only for J-UNIWARD at 0.4 bpnzAC. The DCTR seems to well complement PSRMQ3 as this 20,870-dimensional merger achieves so far the best detection of J-UNIWARD, decreasing E_{OOB} by more than 3% w.r.t. the PSRMQ3 alone. Next, we report on the computational complexity when extracting the feature vector using a Matlab code. The extraction of the DCTR feature vector for one BOSSbase image is twice as fast as JRM, ten times faster than SRMQ1, and almost 200 times faster than the PSRMQ3. Furthermore, a C++ (Matlab MEX) implementation takes only between 0.5–1 sec.

V. Conclusion

This paper introduces a novel feature set for steganalysis of JPEG images. Its name is DCTR because the features are computed from noise residuals obtained using the 64 DCT bases. Its main advantage over previous art is its relatively low dimensionality (8,000) and a significantly lower computational complexity while achieving a competitive detection across many JPEG algorithms. These qual-
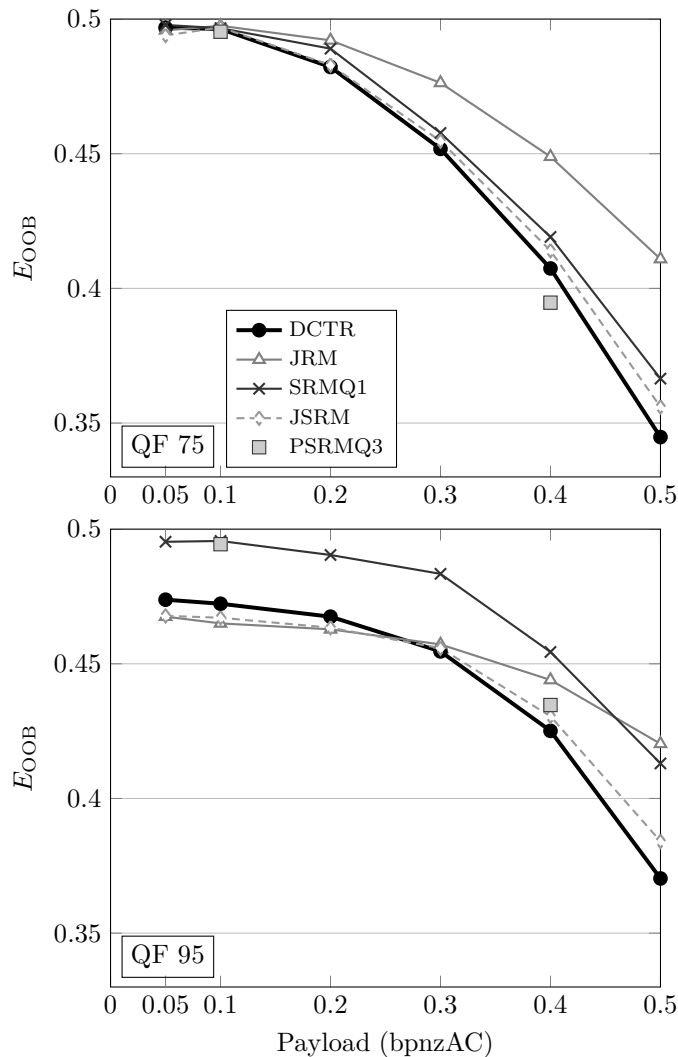
Figure 7. Detection error E_{OOB} for the side-informed SI-UNIWARD for quality factors 75 and 95 when steganalyzed with the proposed DCTR and other rich feature sets. Note the different scale of the y axis.
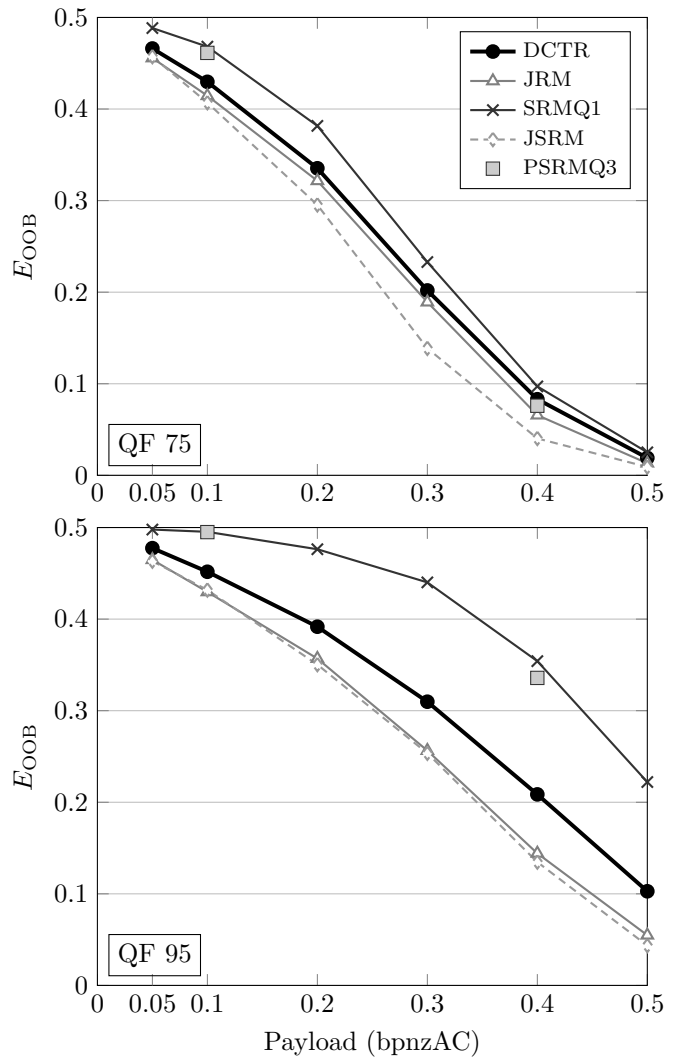

Figure 8. Detection error E_{OOB} for the side-informed NPQ for quality factors 75 and 95 when steganalyzed with the proposed DCTR and other rich feature sets.

Table VI
DETECTION OF J-UNIWARD AT PAYLOAD 0.4 BPNZAC WHEN MERGING VARIOUS FEATURE SETS. THE TABLE ALSO SHOWS THE FEATURE DIMENSIONALITY AND TIME REQUIRED TO EXTRACT A SINGLE FEATURE FOR ONE BOSSBASE IMAGE ON AN INTEL I5 2.4 GHZ COMPUTER PLATFORM.

DCTR (8000)	JRM (22510)	SRMQ1 (12753)	PSRMQ3 (12870)	E_{OOB}	Dim.	Time(s) (Matlab)
●				**0.1523**	8,000	3
	●			0.2561	22,510	6
		●		0.2127	12,753	30
			●	0.1482	12,870	520
●	●			0.1431	30,510	9
●		●		0.1407	20,753	33
●			●	**0.1146**	20,870	523
●	●	●		0.1316	43,263	39
●	●	●	●	0.1252	43,380	529
	●	●		0.1844	35,263	36
●		●	●	0.1429	35,380	526

ities make DCTR a good candidate for building practical steganography detectors and in steganalysis applications where the detection accuracy and the feature extraction time are critical.

The DCTR feature set utilizes the so-called undecimated DCT. This transform has already found applications in steganalysis in the past. In particular, the reference features used in calibration are essentially computed from the undecimated DCT subsampled on an 8×8 grid shifted w.r.t. the JPEG grid. The main point of this paper is the discovery that the undecimated DCT contains much more information that is quite useful for steganalysis.

In the spatial domain, the proposed feature set can be interpreted as a family of one-dimensional co-occurrences (histograms) of noise residuals obtained using kernels formed by DCT bases. Furthermore, the feature set can also be viewed in the JPEG domain as a projection-type model with orthonormal projection vectors. Curiously, we were unable to improve the detection performance by

forming two-dimensional co-occurrences instead of first-order statistics. This is likely because the neighboring elements in the undecimated DCT are qualitatively different projections of DCT coefficients, making the neighboring elements essentially independent.

We contrast the detection accuracy and computational complexity of DCTR with four other rich models when used for detection of five JPEG steganographic methods, including two side-informed schemes. The code for the DCTR feature vector is available from http://dde.binghamton.edu/download/feature_extractors/ (note for the reviewers: the code will be posted upon acceptance of this manuscript).

Finally, we would like to mention that it is possible that the DCTR feature set will be useful for forensic applications, such as [24], since many feature sets originally designed for steganalysis found applications in forensics. We consider this as a possible future research direction.

APPENDIX

Here, we provide the proof of orthonormality (6) of vectors $\mathbf{p}_{a,b}^{(k,l)}$ defined in (5). It will be useful to follow Figure 9 for easier understanding. For each a, b, $0 \leq a, b \leq 7$, the (i, j)th DCT basis pattern $\mathbf{B}^{(i,j)}$ positioned so that its upper left corner has relative index (a, b) is split into four 8×8 subpatterns: κ stands for cirκle, μ stands for diaμond, τ for τriangle, and σ for σtar:

$$\kappa_{mn}^{(i,j)} = \begin{cases} B_{m-a,n-b}^{(i,j)} & \begin{aligned} a \leq m \leq 7 \\ b \leq n \leq 7 \end{aligned} \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_{mn}^{(i,j)} = \begin{cases} B_{m-a,8+n-b}^{(i,j)} & \begin{aligned} a \leq m \leq 7 \\ 0 \leq n < b \end{aligned} \\ 0 & \text{otherwise} \end{cases}$$

$$\tau_{mn}^{(i,j)} = \begin{cases} B_{8+m-a,n-b}^{(i,j)} & \begin{aligned} 0 \leq m < a \\ b \leq n \leq 7 \end{aligned} \\ 0 & \text{otherwise.} \end{cases}$$

$$\sigma_{mn}^{(i,j)} = \begin{cases} B_{8+m-a,8+n-b}^{(i,j)} & \begin{aligned} 0 \leq m < a \\ 0 \leq n < b \end{aligned} \\ 0 & \text{otherwise} \end{cases}$$

In Figure 9 top, the four patterns are shown using four different markers. The light-color markers correspond to zeros. The first 64 elements of $\mathbf{p}_{a,b}^{(i,j)}$ are simply projections of $\kappa_{mn}^{(i,j)}$ onto the 64 patterns forming the DCT basis. The next 64 elements are projections of $\mu_{mn}^{(i,j)}$ onto the DCT basis, the next 64 are projections of $\tau_{mn}^{(i,j)}$, and the last 64 are projections of $\sigma_{mn}^{(i,j)}$. We will denote these projections with the same Greek letters but with a single index instead: $(\kappa_1^{(i,j)}, \ldots, \kappa_{64}^{(i,j)})$, $(\mu_1^{(i,j)}, \ldots, \mu_{64}^{(i,j)})$, $(\tau_1^{(i,j)}, \ldots, \tau_{64}^{(i,j)})$, and
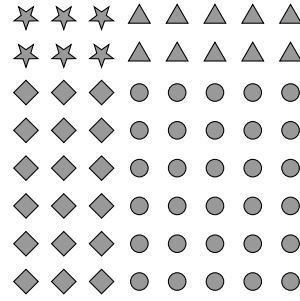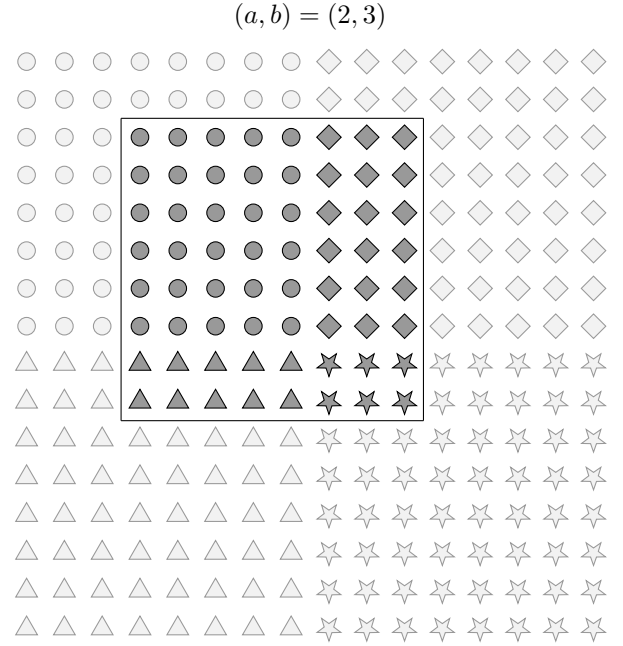
$$(a, b) = (2, 3)$$

Figure 9. Diagram showing the auxiliary patterns κ (cirκle), μ (diaμond), τ (τriangle), and σ (σtar). The black square outlines the position of the DCT basis pattern $\mathbf{B}^{(i,j)}$.

$(\sigma_1^{(i,j)}, \ldots, \sigma_{64}^{(i,j)})$. In terms of the introduced notation,

$$\mathbf{p}_{a,b}^{(i,j)T} \cdot \mathbf{p}_{a,b}^{(k,l)} = \sum_{r=1}^{64} \kappa_r^{(i,j)} \kappa_r^{(k,l)} + \sum_{r=1}^{64} \mu_r^{(i,j)} \mu_r^{(k,l)}$$

$$+ \sum_{r=1}^{64} \tau_r^{(i,j)} \tau_r^{(k,l)} + \sum_{r=1}^{64} \sigma_r^{(i,j)} \sigma_r^{(k,l)}. \quad (11)$$

Note that the sum $\kappa^{(i,j)} + \mu^{(i,j)} + \tau^{(i,j)} + \sigma^{(i,j)}$ is the entire DCT mode (i, j) split into four pieces and rearranged back together to form an 8×8 block (Figure 9 botom). For fixed a, b, due to the orthonormality of DCT modes (i, j) and (k, l), $\kappa^{(i,j)} + \mu^{(i,j)} + \tau^{(i,j)} + \sigma^{(i,j)}$ and $\kappa^{(k,l)} + \mu^{(k,l)} + \tau^{(k,l)} + \sigma^{(k,l)}$ are thus also orthonormal and so are their projections onto the DCT basis (because the

DCT transform is orthonormal):

$$\sum_{r=1}^{64}(\kappa_r^{(i,j)} + \mu_r^{(i,j)} + \tau_r^{(i,j)} + \sigma_r^{(i,j)}) \times$$

$$(\kappa_r^{(k,l)} + \mu_r^{(k,l)} + \tau_r^{(k,l)} + \sigma_r^{(k,l)}) = \delta_{(i,j),(k,l)}. \qquad (12)$$

The orthonormality now follows from the fact that the LHS of (12) and the RHS of (11) have the exact same value because the sum of every mixed term in (12) is zero (e.g., $\sum_{r=1}^{64} \kappa_r^{(i,j)}\tau_r^{(k,l)} = 0$, etc.). This is because the subpatterns $\kappa^{(i,j)}$ and $\tau^{(k,l)}$ have *disjoint supports* (their dot product in the spatial domain is 0 and thus the product in the DCT domain is also 0 because DCT is orthonormal).

REFERENCES

[1] I. Avcibas, N. D. Memon, and B. Sankur. Steganalysis using image quality metrics. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III*, volume 4314, pages 523–531, San Jose, CA, January 22–25, 2001.

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

[3] S. Bayram, A. E. Dirik, H. T. Sencar, and N. Memon. An ensemble of classifiers approach to steganalysis. In *20th International Conference on Pattern Recognition (ICPR)*, pages 4376–4379, Istanbul, Turkey, August 23 2010.

[4] R. Böhme. Weighted stego-image steganalysis for JPEG covers. In K. Solanki, K. Sullivan, and U. Madhow, editors, *Information Hiding, 10th International Workshop*, volume 5284 of Lecture Notes in Computer Science, pages 178–194, Santa Barbara, CA, June 19–21, 2007. Springer-Verlag, New York.

[5] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, August 1996.

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

[7] R. Cogranne and F. Retraint. Application of hypothesis testing theory for optimal detection of LSB Matching data hiding. *Signal Processing*, 93(7):1724–1737, July, 2013.

[8] H. Farid and L. Siwei. Detecting hidden messages using higher-order statistics and support vector machines. In F. A. P. Petitcolas, editor, *Information Hiding, 5th International Workshop*, volume 2578 of Lecture Notes in Computer Science, pages 340–354, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.

[9] J. Fridrich, M. Goljan, and D. Hogea. Steganalysis of JPEG images: Breaking the F5 algorithm. In *Information Hiding, 5th International Workshop*, volume 2578 of Lecture Notes in Computer Science, pages 310–323, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.

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

[11] 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.

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

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

[14] V. Holub and J. Fridrich. Universal distortion design for steganography in an arbitrary domain. *EURASIP Journal on Information Security, Special Issue on Revised Selected Papers of the 1st ACM IH and MMS Workshop*, 2014:1, 2014.

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

[16] A. D. Ker. Implementing the projected spatial rich features on a GPU. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014*, volume 9028, pages 1801–1810, San Francisco, CA, February 3–5, 2014.

[17] A. D. Ker and T. Pevný. Identifying a steganographer in realistic and heterogeneous data sets. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages 0N 1–13, San Francisco, CA, January 23–26, 2012.

[18] 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.

[19] J. Kodovský and J. Fridrich. Quantitative structural steganalysis of Jsteg. *IEEE Transactions on Information Forensics and Security*, 5(4):681–693, December 2010.

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

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

[22] Liyun Li, H. T. Sencar, and N. Memon. A cost-effective decision tree based approach to steganalysis. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2013*, volume 8665, pages 0P 1–7, San Francisco, CA, February 5–7, 2013.

[23] 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.

[24] Q. Liu and Z. Chen. Improved approaches to steganalysis and seam-carved forgery detection in JPEG images. *ACM Trans. Intell. Syst. Tech. Syst.*, pages 39:1–30, 2014.

[25] I. Lubenko and A. D. Ker. Going from small to large data sets in steganalysis. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages OM 1–10, San Francisco, CA, January 23–26, 2012.

[26] S. Lyu and H. Farid. Steganalysis using higher-order image statistics. *IEEE Transactions on Information Forensics and Security*, 1(1):111–119, 2006.

[27] 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 1–14, San Jose, CA, January 29–February 1, 2007.

[28] T. Pevný and J. Fridrich. Novelty detection in blind steganalysis. In A. D. Ker, J. Dittmann, and J. Fridrich, editors, *Proceedings of the 10th ACM Multimedia & Security Workshop*, pages 167–176, Oxford, UK, September 22–23, 2008.

[29] 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.

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

[31] A. Westfeld. Generic adoption of spatial steganalysis to transformed domain. In K. Solanki, K. Sullivan, and U. Madhow, editors, *Information Hiding, 10th International Workshop*, vol-

ume 5284 of Lecture Notes in Computer Science, pages 161–177, Santa Barbara, CA, June 19–21, 2007. Springer-Verlag, New York.

[32] A. Westfeld and A. Pfitzmann. Attacks on steganographic systems. In A. Pfitzmann, editor, *Information Hiding, 3rd International Workshop*, volume 1768 of Lecture Notes in Computer Science, pages 61–75, Dresden, Germany, September 29–October 1, 1999. Springer-Verlag, New York.

[33] T. Zhang and X. Ping. A fast and effective steganalytic technique against Jsteg-like algorithms. In *Proceedings of the ACM Symposium on Applied Computing*, pages 307–311, Melbourne, FL, March 9–12, 2003.

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

Vojtěch Holub is currently an R&D engineer at Digimarc Corporation, Beaverton, OR. He received his Ph.D. in 2014 at the department of Electrical and Computer Engineering at Binghamton University, New York. The main focus of his dissertation was on steganalysis and steganography. He received his M.S. degree in Software Engineering from the Czech Technical University in Prague in 2010.

Jessica Fridrich holds the position of Professor of Electrical and Computer Engineering at Binghamton University (SUNY). She has received her PhD in Systems Science from Binghamton University in 1995 and MS in Applied Mathematics from Czech Technical University in Prague in 1987. Her main interests are in steganography, steganalysis, digital watermarking, and digital image forensic. Dr. Fridrich's research work has been generously supported by the US Air Force and AFOSR. Since 1995, she received 19 research grants totaling over $9 mil for projects on data embedding and steganalysis that lead to more than 160 papers and 7 US patents. Dr. Fridrich is a member of IEEE and ACM.