

Steganography with Multiple JPEG Images of the Same Scene

Tomáš Denemark, *Student Member, IEEE* and Jessica Fridrich, *Fellow, IEEE*

Abstract—It is widely recognized that incorporating side-information at the sender can significantly improve steganographic security in practice. Currently, most side-informed schemes utilize a high quality “precover” image that is subsequently processed and then jointly quantized and embedded with a secret. In this paper, we investigate an alternative form of side-information – a set of multiple JPEG images of the same scene – for applications when the sender does not have access to a precover. The additional JPEG images are used to determine the preferred polarity of embedding changes to modulate the costs of changing individual DCT coefficients in an existing embedding scheme. Tests on real images with synthesized acquisition noise and on real multiple acquisitions obtained with a tripod-mounted and hand-held digital camera show a rather significant improvement in empirical security with respect to steganography utilizing a single JPEG image. The proposed empirically determined modulation of embedding costs is justified using Monte Carlo simulations by showing that qualitatively the same modulation minimizes the Bhattacharyya distance between a quantized generalized Gaussian model of cover and stego DCT coefficients corrupted by AWG acquisition noise.

Index Terms—Steganography, side-information, precover, acquisition, security, steganalysis, JPEG

I. INTRODUCTION

Steganography is typically cast using three characters – Alice and Bob, who communicate by hiding their messages in cover objects, and the steganalyst, the Warden, whose goal is to discover the presence of secrets. Since empirical cover sources [1], such as digital media, are too complex to be exhaustively described using tractable statistical models, both the steganographer and the Warden have to work with approximations. This has fundamental consequences for the steganographer, who is unable to achieve perfect security, as well as for the Warden, who inevitably builds sub-optimal detectors.

The steganographer seems to have a fundamental advantage because she may have access to more information than the Warden and thus partially compensate for the

lack of the cover model. For example, Alice may have a high quality representation of the cover image called precover [2] and embed her secret while processing the precover and/or converting it to a different format. The first example of this technique is the embedding-while-dithering steganography [3], which embeds secrets when converting a true-color image to a palette format. By far the most common side-informed steganography today hides in JPEG images using non-rounded DCT coefficients [4], [5], [6], [7], [8], [9], [10].

Most consumer electronic devices, such as cell phones, tablets, and low-end digital cameras, however, save their images only in the JPEG format and thus do not give the user access to non-rounded DCT coefficients. In this case, Alice can utilize a different type of side-information – she can take multiple JPEG images of the same scene. This research direction has not been developed as much mostly due to the difficulty of acquiring the required imagery and modeling the differences between acquisitions. Prior work on this topic includes [11], [12], [13] where the authors made multiple scans of the same printed image on a flat-bed scanner and then attempted to model the acquisition noise. Unfortunately, this requires acquiring a potentially large number of scans, which makes this approach rather labor intensive. Moreover, differences in the movement of the scanner head between individual scans lead to slight spatial misalignment that complicates using this type of side-information properly. Because this problem is especially pronounced when embedding in the pixel domain, in this paper we work with multiple images acquired in the JPEG format as we expect quantized DCT coefficients to be naturally more robust to small differences between acquisitions. Since our intention is to design a practical method, we avoid the difficult and potentially extremely time consuming task of modeling the differences between acquisitions [11], [12], [13] and make the approach work well even when mere *two* images are available to Alice. In another relevant prior art [14], the authors proposed embedding by stitching patches from multiple acquisitions in a predefined pattern. The individual patches are not modified and are therefore statistically indistinguishable from the original images. However, as the authors discussed in their paper there are likely going to be detectable differences between individual patches and inconsistencies at their boundaries. Furthermore, the required number of acquisitions quickly grows with the length of the secret message. By using 150 acquisitions of the same scene (scans), the authors were able to embed only 0.157 bits per non-zero AC coefficient on average.

The following authors are with the Department of Electrical and Computer Engineering, Binghamton University, NY, 13902, USA. Email: {tdenema1,fridrich}@binghamton.edu.

The work on this paper was partially supported by NSF grant No. 1561446 and by Air Force Office of Scientific Research under the research grant number FA9950-12-1-0124. 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.

In the next section, we introduce background information and notation used throughout the paper. Section III contains a brief summary of existing side-informed steganography with a high quality precover. In Section IV, the new steganographic method that uses two or more JPEG images at the sender is described. Starting with the embedding costs of an existing cost-based JPEG steganography, they are modulated based on the preferred direction deduced from the second JPEG image of the same scene. The method is first subjected to tests on BOSSbase images with simulated acquisition noise in Section V to see the gain in the ideal case with a simple acquisition noise. To gain insight about the security of the proposed scheme in real-life conditions, in Section VI we describe two new datasets called BURSTbase and BURSTbaseH with images obtained with a tripod-mounted and hand-held digital camera, respectively. Evidence is provided that the differences between the two closest exposures in BURSTbase are due to heteroscedastic acquisition noise. In Section VII, we first report the results of experiments on BURSTbase for J-UNIWARD costs [9] across a wide range of quality factors and payloads and contrasted with J-UNIWARD and SI-UNIWARD to see the gain w.r.t. using only a single JPEG image and the comparison to other type of side-information. We also investigate how the gain in security decreases with increased differences between exposures. This section continues with a summary of experiments on BURSTbaseH images with hand-held camera on both J-UNIWARD and UED-JC [8]. Although the security gain is smaller than for BURSTbase, when the steganographer rejects bad bursts, a significant security gain is still observed w.r.t. steganography with a single JPEG. Finally, the appendix contains analysis that explains the shape of the experimentally determined modulation of costs. The paper is concluded in Section VIII.

This manuscript is an expanded version of an abbreviated version of this work published at IEEE ICASSP [15]. In particular, this 13-page manuscript extends the 4+1-page conference paper in the following important aspects:

- 1) The proposed method is introduced in a more general setting applicable to any cost-based embedding scheme operating in the JPEG domain. Likewise, it is implemented and tested for other embedding schemes besides J-UNIWARD, such as the UED-JC steganography [8].
- 2) The qualitative dependence of the modulation factor for adjusting the costs of DCT coefficients on the JPEG quality factor is explained with Monte Carlo simulations by employing a generalized Gaussian model of DCT coefficients.
- 3) The database used in the main bulk of experiments, the BURSTbase, is analyzed in detail to put forward evidence that the two closest images from BURSTbase indeed differ primarily in the acquisition noise with heteroscedastic properties.
- 4) The experimental section was substantially expanded with a) experiments on images taken with

a hand-held camera to show the practicality of the proposed method, b) experiments on simulated acquisition noise to show that in this ideal case the proposed method can outperform even side-informed steganography with a single high-quality precover (this gain is explained by contrasting steganography with precover and with two JPEGs w.r.t. the number of correctly and incorrectly determined directions of changes to be modulated), c) experiments on the UED-JC embedding algorithm to show the generality of the proposed methodology, and d) experiments showing that by rejecting bad bursts the steganographer can retain a rather significant advantage of embedding with two JPEGs w.r.t. a single JPEG.

- 5) Specific ideas for technology transfer of the proposed method are put forward.

II. PRELIMINARIES

In this section, we introduce basic terminology, notation, and concepts used throughout the paper.

For simplicity and WLOG, we will work with 8-bit $M \times N$ grayscale images with pixels $\mathbf{z} = (z_{ij}) \in \mathcal{R}^{M \times N}$, $\mathcal{R} = \{0, \dots, 255\}$, with both M and N multiples of 8. During JPEG compression, \mathbf{z} is divided into disjoint blocks of 8×8 pixels, $z_{ij}^{(u,v)}$, $1 \leq i, j \leq 8$, $1 \leq u \leq M/8$, $1 \leq v \leq N/8$, where (u, v) is the block index. Discrete cosine transform (DCT) is then applied to each block, resulting in 8×8 blocks of DCT coefficients $d_{ij}^{(u,v)}$, $\mathbf{d}^{(u,v)} = \text{DCT}(\mathbf{z}^{(u,v)})$, where $\mathbf{d}^{(u,v)}$ and $\mathbf{z}^{(u,v)}$ are 8×8 matrices of DCT coefficients and pixels in the (u, v) th block, respectively. The next step in JPEG compression involves dividing $d_{ij}^{(u,v)}$ by quantization steps q_{ij} , $c_{ij}^{(u,v)} = d_{ij}^{(u,v)}/q_{ij}$, and rounding to integers $x_{ij}^{(u,v)} = Q_1(c_{ij}^{(u,v)})$, where $Q_1(\cdot)$ quantizes to $\{-1023, \dots, 1024\}$ and $\mathbf{q} = (q_{ij})$ is the luminance quantization matrix. The quantized DCT coefficients $x_{ij}^{(u,v)}$ are then losslessly encoded, appended with a header, and saved as a JPEG file.

Throughout this paper, we will use indices i, j to index DCT coefficients in an image as well as in a specific (u, v) th block. Thus, in x_{ij} , the range of indices i, j is over the entire $M \times N$ image while in $x_{ij}^{(u,v)}$ it is restricted to $1 \leq i, j \leq 8$. We believe that this switching from global to block-based indexing is natural, it simplifies the language, and should not become a source of confusion.

A generalized Gaussian distribution with density

$$f_{\text{GG}}(x; \mu, \alpha, b) = \frac{\alpha}{2b\Gamma(1/\alpha)} \exp\left(-\left|\frac{x-\mu}{b}\right|^\alpha\right), \quad (1)$$

where μ, α, b are the mean, shape, and width parameters, will be denoted $\mathcal{G}(\mu, \alpha, b)$.

Images acquired using an imaging sensor are noisy measurements of the true scene \mathbf{r} by which we understand the image rendered by the camera lens. The randomness in the form of noise or imperfections is introduced by several separate mechanisms [16], which include the shot noise (photon noise), dark current, and electronic and readout noise. Note that defective pixels and the photo-response

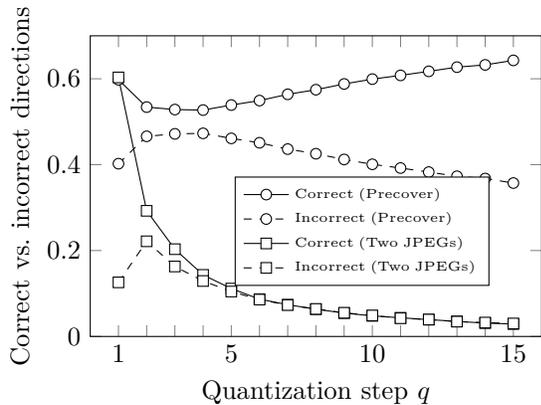


Figure 1. Relative number of correctly and incorrectly determined embedding directions for steganography informed by the values of non-rounded DCT coefficients (precover) and by two JPEG images. See Section V for details.

non-uniformity are deterministic imperfections that are fixed for a given camera. Formally, $\mathbf{z} = \mathbf{r} + \boldsymbol{\xi} \in \mathcal{R}^{M \times N}$, where $\boldsymbol{\xi}$ is the acquisition noise and \mathbf{r} is a parameter that is unknown to both Alice and the Warden but technically not random. An additive white Gaussian (AWG) model $\xi_{ij} \sim \mathcal{N}(0, \sigma_a^2)$ is rather accurate for RAW sensor capture of a uniformly lit scene but only an approximation for images with natural content where the variance is a linear function of pixel intensity (the heteroscedastic noise [17], [18]). For a sensor capable of registering color, color interpolation and correction introduce dependencies among neighboring values of ξ_{ij} and across color channels. Additional local dependencies are introduced by filtering that may be applied inside the camera, such as denoising and sharpening, and by lens distortion correction, making the statistical properties of the random field ξ_{ij} extremely complicated.

III. STEGANOGRAPHY WITH PRECOVER

With the exception of YASS [19], all modern embedding schemes for JPEG images, whether or not they use precover, are implemented within the paradigm of distortion minimization. The steganographer first specifies the cost of modifying each cover element (DCT coefficient) and then embeds the payload so that the expected value of the total induced distortion (the sum of costs of all changed cover elements) is as small as possible. Syndrome-trellis codes [20] can achieve this goal near the corresponding rate-distortion bound.

The costs of changing the quantized JPEG coefficient $x_{ij}^{(u,v)}$ by $+1$ and -1 will be denoted $\rho_{ij}^{(u,v)}(+1)$ and $\rho_{ij}^{(u,v)}(-1)$, respectively. The total cost (distortion) of embedding is $D(\mathbf{x}, \mathbf{y}) = \sum_{x_{ij} \neq y_{ij}} \rho_{ij}(y_{ij} - x_{ij})$, where $y_{ij} \in \{x_{ij} - 1, x_{ij}, x_{ij} + 1\}$ are quantized DCT coefficients from the stego image. An embedding scheme operating at the rate-distortion bound (with minimal D) embeds a payload of R bits by modifying the DCT coefficients with

probabilities [20]:

$$\beta_{ij}^{\pm} = \mathbb{P}\{y_{ij} = x_{ij} \pm 1\} = \frac{e^{-\lambda \rho_{ij}(\pm 1)}}{1 + e^{-\lambda \rho_{ij}(+1)} + e^{-\lambda \rho_{ij}(-1)}} \quad (2)$$

where λ is determined from the payload constraint

$$R = \sum_{ij} h_3(\beta_{ij}^+, \beta_{ij}^-), \quad (3)$$

with $h_3(x, y) = -x \log_2 x - y \log_2 y - (1 - x - y) \log_2(1 - x - y)$ the ternary entropy function in bits.

One of the most secure schemes for JPEG images called J-UNIWARD [9] uses symmetric costs $\rho_{ij}(+1) = \rho_{ij}(-1)$ for all i, j . Alice can prohibit the embedding from modifying x_{ij} , e.g., by $+1$, by setting $\rho_{ij}(+1) = C_{\text{wet}}$, where C_{wet} is a very large number, the so-called “wet cost” [21].

Side-informed steganography relates to embedding schemes where the sender has some additional information that is used to adjust the costs. For JPEG steganography, the side-information may be in the form of an uncompressed image or, equivalently, the unquantized precover values c_{ij} . Since c_{ij} are not available to the Warden, Alice has a fundamental advantage. As shown in [22], c_{ij} partially compensates for the lack of knowledge of the cover model when it is highly non-stationary.

While it is currently not known how to use side-information in an optimal fashion for embedding, numerous heuristic schemes were proposed in the past [5], [23], [7], [8], [9], [10], [6]. Typically, the rounding error $e_{ij} = c_{ij} - x_{ij}$, $-1/2 \leq e_{ij} \leq 1/2$, is used to modulate the embedding costs ρ_{ij} by $1 - 2|e_{ij}| \in [0, 1]$. In SI-UNIWARD [9], for example, the costs are:

$$\rho_{ij}(\text{sign}(e_{ij})) = (1 - 2|e_{ij}|)\rho_{ij}^{(J)} \quad (4)$$

$$\rho_{ij}(-\text{sign}(e_{ij})) = C_{\text{wet}}, \quad (5)$$

where $\rho_{ij}^{(J)}$ are J-UNIWARD costs. In other words, SI-UNIWARD is a binary embedding scheme that either leaves a DCT coefficient unmodified (rounds c_{ij} to x_{ij}) or rounds it to the “other side” in the direction of x_{ij} , in which case the J-UNIWARD cost associated with this change is modulated. The intuition behind the modulation is clear: when $|e_{ij}| \approx 1/2$, a small perturbation could cause c_{ij} to be rounded to the other side. Such coefficients are thus assigned a proportionally smaller cost. On the other hand, the costs are unchanged when $e_{ij} \approx 0$, as it takes a larger perturbation to change the rounded value.

In [10], a ternary version of SI-UNIWARD was studied where the authors argued that, as the rounding error e_{ij} becomes small, the embedding rule should be allowed to change the coefficient both ways. This ternary version of SI-UNIWARD uses the following costs:

$$\rho_{ij}(\text{sign}(e_{ij})) = (1 - 2|e_{ij}|)\rho_{ij}^{(J)} \quad (6)$$

$$\rho_{ij}(-\text{sign}(e_{ij})) = \rho_{ij}^{(J)}. \quad (7)$$

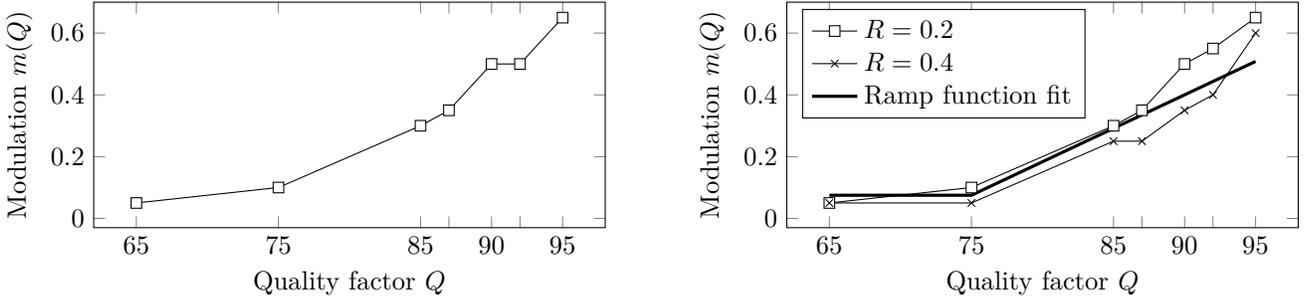


Figure 2. Optimal modulation factor $m(Q)$ as a function of the JPEG quality factor Q . Left: BOSSbase 1.01 images with simulated acquisition noise. Right: BURSTbase.

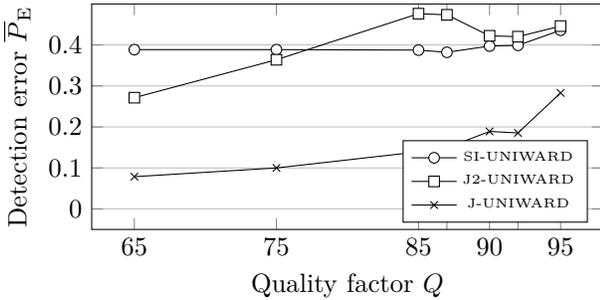


Figure 3. Empirical security, \bar{P}_E , as a function of the JPEG quality factor for relative payload $R = 0.4$ bpnzac for J2-UNIWARD, J-UNIWARD, and SI-UNIWARD. BOSSbase with simulated acquisition noise, low-complexity linear classifier trained with GFR.

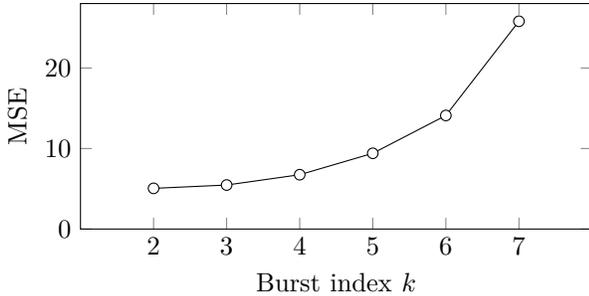


Figure 4. MSE between $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(k)}$, $k = 2, \dots, 7$ from each burst averaged over all 9,310 bursts from BURSTbase. See Section VI for notation and further details.

IV. STEGANOGRAPHY WITH MULTIPLE JPEGs

In this section, we describe the proposed scheme for embedding in JPEG images when the sender possesses more than one acquisition of (approximately) the same scene. We start with the embedding algorithm for two acquisitions and then discuss the possibilities for its generalization to more than two acquisitions. The main embedding algorithm is explained with a pseudo-code to allow faster understanding of the main concept and ease the implementation for practitioners.

Before we start, we wish to discuss some important philosophical issues. In reality, it is in principle impossible to obtain two independent samplings of one object (Heraclitus' "You could not step twice into the same river" by Plato in *Cratylus*, 402a) because of small differences

in exposure time, physical shaking of the camera, and small differences in the scene itself, e.g., due to wind and the amount and direction of illumination. In this article, for brevity we nevertheless abuse the language a little while being aware of the fact that in reality the images will inevitably contain differences other than those due to acquisition noise. One mission of this paper is to investigate whether, despite these obvious limitations, it is possible to make use of the other acquisitions to improve steganographic security.

The proposed method can be applied to any cost-based scheme that embeds in quantized DCT coefficients of a JPEG file. In fact, it is not limited to the JPEG format and could be applied to other lossy formats, such as the JPEG 2000. We restrict ourselves to JPEG images in this article because it is by far the most ubiquitous image format in current use.

A. Two exposures

First, we describe the embedding algorithm when two JPEG versions of the cover image are available. We denote the quantized DCT coefficients in both images by $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$ and pronounce, for example, the first image as the cover JPEG and consider $x_{ij}^{(2)}$ as side-information.

Pronouncing $x_{ij}^{(1)}$ as cover and $x_{ij}^{(2)}$ as side-information, the sender first computes from $x_{ij}^{(1)}$ the costs of changing the ij th DCT coefficient by -1 and $+1$: $\rho_{ij}^{(0)}(-1)$ and $\rho_{ij}^{(0)}(+1)$. The costs can be computed using, e.g., an existing cost-based embedding scheme, such as J-UNIWARD or one of the versions of UED. The proposed embedding scheme keeps these costs when $x_{ij}^{(1)} = x_{ij}^{(2)}$ and modulates the costs otherwise. This can be explained by finding the new costs $\rho_{ij}(\pm 1)$ via the following two-step procedure:

$$\text{Step 1: set } \rho_{ij}(\pm 1) = \rho_{ij}^{(0)}(\pm 1) \quad (8)$$

$$\text{Step 2: } x_{ij}^{(1)} \neq x_{ij}^{(2)} \Rightarrow \rho_{ij}(s_{ij}) = m(Q)\rho_{ij}^{(0)}(s_{ij}), \quad (9)$$

$$\text{where } s_{ij} = \text{sign}(x_{ij}^{(2)} - x_{ij}^{(1)}) \quad (10)$$

where $m(Q) \in [0, 1]$ is a modulation factor that depends on the quality factor $1 \leq Q \leq 100$. To ease the understanding of the embedding method and its implementation,

Algorithm 1 shows the pseudo-code for the embedding algorithm.

The value of the modulation factor $m(Q)$ will be determined experimentally for each tested quality factor Q and cover source by a search over $m(Q) \in [0, 1]$ to obtain the smallest minimal total probability of error, $P_E = \min_{P_{FA}} (P_{MD} + P_{FA})/2$, where P_{MD} and P_{FA} are missed-detection and false-alarm rates of a detector implemented using a low-complexity linear classifier [24] with the Gabor Filter Residual (GFR) features [25] on the training set. The GFR features were selected for the design because they are known to be highly effective against modern JPEG steganography, including J-UNIWARD and all versions of UED [26], [8]. Experiments show that $m(Q)$ should generally be increasing in Q . The experimental Sections V and VII and the appendix contain further details on the specific form $m(Q)$.

Our final note of this section concerns a naming convention. An embedding scheme with two JPEGs with J-UNIWARD (UED-JC) costs will be abbreviated as J2-UNIWARD (and UED2-JC).

B. Multiple exposures

In this section, we discuss several possibilities for extending the embedding algorithm to the case when Alice acquires $k > 2$ JPEG images of the same scene, $x_{ij}^{(1)}, \dots, x_{ij}^{(k)}$.

With increased k , it may become possible to obtain a more accurate estimate of the noise-free scene r_{ij} (Section II), for example, as a maximum-likelihood $\hat{r}_{ij}^{(ML)} = (x_{ij}^{(1)} + \dots + x_{ij}^{(k)})/k$ or a MAP estimate by leveraging a prior on $x_{ij}^{(u,v)}$, $1 \leq i, j \leq 8$, with u, v the 8×8 block index, estimated for the given source. The estimates, however, will likely be biased since spatial misalignment between exposures and differences other than due to the acquisition noise will likely increase with k , making it not clear whether the additional exposures are an asset.

Moreover, it is not clear how the embedding should incorporate such estimates. Using \hat{r}_{ij} as a “high-quality precover” and applying standard side-informed steganography, such as SI-UNIWARD, is questionable because the rounded values $[\hat{r}_{ij}]$ form a different source with a suppressed acquisition noise. On the other hand, using \hat{r}_{ij} as a “high-quality precover” for one of the JPEGs, e.g., $x_{ij}^{(1)}$, would lead to “rounding errors” $e_{ij} = \hat{r}_{ij} - x_{ij}^{(1)}$ out of the range $[-1/2, 1/2]$ and would thus require a revisit of the established cost modulation (4) and (6).

In the end, and based on our experiments in Sections VII-A and VII-B, it appears that the best way to use multiple images in practice is to simply select a pair of two closest images among the k exposures and apply the algorithm described in the previous section.

V. STUDY WITH SIMULATED ACQUISITION NOISE

Our first experimental evaluation involves tests on images with simulated acquisition noise. These are included

Algorithm 1 Pseudo-code for side-informed embedding with two JPEGs.

- 1: **Input:** Two quality factor Q JPEG images with quantized DCT coefficients $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$, $1 \leq i \leq M$, $1 \leq j \leq N$
 - 2: **Output:** Stego JPEG image with DCT coefficients $y_{ij}^{(1)}$
 - 3: **Compute costs** $\rho_{ij}^{(0)}(-1), \rho_{ij}^{(0)}(+1)$ **of DCT coefficients from JPEG** $x_{ij}^{(1)}$ **(the cover)**
 - 4: **for** $i = 1, \dots, M$ **do**
 - 5: **for** $j = 1, \dots, N$ **do**
 - 6: $\rho_{ij}(\pm 1) = \rho_{ij}^{(0)}(\pm 1)$
 - 7: $s_{ij} = \text{sign}(x_{ij}^{(2)} - x_{ij}^{(1)})$
 - 8: **IF** $x_{ij}^{(1)} \neq x_{ij}^{(2)}$ **THEN** $\rho_{ij}(s_{ij}) = m(Q)\rho_{ij}^{(0)}(s_{ij})$
 - 9: **end for**
 - 10: **end for**
 - 11: **Embed message in** $x_{ij}^{(1)}$ **using costs** ρ_{ij} **using STCs to obtain stego JPEG file with DCT coefficients** y_{ij}
 - 12: **Recipient reads the secret message using STCs from the stego JPEG file** y_{ij}
-

because they constitute the “ideal” (and unachievable in practice) situation when no other differences between the exposures exist besides a very simple form of the acquisition noise. These results will be contrasted with real multiple exposures.

The mother database was BOSSbase 1.01 [27] containing 10,000 8-bit grayscale 512×512 PGM images. Two different realizations of Gaussian noise $\mathcal{N}(0, 1)$ were added to the images, producing two simulated acquisitions $z_{ij}^{(l)}$, $l = 1, 2$, which were subsequently compressed with a range of JPEG quality factors to obtain the values of rounded DCT coefficients $x_{ij}^{(l)}$, $l = 1, 2$, for each image in the database. Each JPEG image $x_{ij}^{(1)}$ was then embedded with relative payload $R = 0.4$ bits per non-zero AC DCT coefficient (bpnzac) using J2-UNIWARD. The values of the optimal modulation factor $m(Q)$ as a function of Q for this source are shown in Figure 2 left.

Figure 3 shows \bar{P}_E , which is the detection error P_E averaged over ten random splits of the database into training and testing parts as a function of the JPEG quality factor. We do not show the statistical spread of the detection error as it is very small and in most cases covered by the markers. In all experiments in this manuscript, the largest encountered standard deviation of the detection error was 0.0122 and the average was 0.0042. The classifier was a low-complexity linear classifier [24] and the feature set is the Gabor Filter Residual (GFR) [25] rich model known to be highly effective against modern steganographic schemes. For comparison, the figure also contains the detection error for J-UNIWARD (with $x_{ij}^{(1)}$ as covers) and SI-UNIWARD (with $c_{ij}^{(1)}$ as side-information). For a simulated acquisition noise, the side-information in the form of two JPEG images significantly increases

empirical security w.r.t. embedding with a single JPEG (J-UNIWARD). It seems even more valuable for quality factors $Q \gtrsim 80$ than non-rounded DCT coefficients (SI-UNIWARD). We next shed some light on why this is the case.

The value $x_{ij}^{(2)}$ can only be useful to Alice when $x_{ij}^{(2)} \neq x_{ij}^{(1)}$, which will happen increasingly more often with smaller quantization steps q_{ij} (larger JPEG quality). This type of side-information is different from the non-rounded values $c_{ij}^{(1)}$. It informs Alice about the direction along which the costs should be modulated and less about the magnitude of the rounding error $e_{ij}^{(1)} = c_{ij}^{(1)} - x_{ij}^{(1)}$. To better understand the difference between these two types of side-information, we conducted the following experiment.

A generalized Gaussian model $\mathcal{G}(0, 0.4, 0.1)$ was adopted for the distribution of DCT coefficients of the noise-free scene \mathbf{r} . These parameters roughly correspond to medium spatial frequencies in BOSSbase 1.01 [27] images. Then, we generated $2 \times N_{\text{MC}}$ independent realizations from $\mathcal{G}(0, 0.4, 0.1)$, $r_k^{(1)}$ and $r_k^{(2)}$, $k \in \{1, \dots, N_{\text{MC}}\}$, $N_{\text{MC}} = 10^6$. Next, N_{MC} independent realizations from $\mathcal{N}(0, 1)$ were added to both vectors,¹ divided by $q \in \{1, \dots, 15\}$ and rounded to integers, $c_k^{(l)} = (r_k^{(l)} + \xi_k^{(l)})/q$, $x_k^{(l)} = \lfloor c_k^{(l)} \rfloor$, $l = 1, 2$. We then counted how often the different side-information correctly informed us about the sign of the rounding error (direction of the stego changes).

We will say that side-information $c_k^{(1)}$ correctly determines the direction of steganographic changes with respect to the noise-free scene if the embedding modifies the quantized cover value $x_k^{(1)}$ towards the noise-free scene $r_k^{(1)}$, which will happen when the rounding error $e_k^{(1)} = c_k^{(1)} - x_k^{(1)}$ has the same sign as $r_k^{(1)}/q - x_k^{(1)}$, or when $(c_k^{(1)} - x_k^{(1)})(r_k^{(1)}/q - x_k^{(1)}) > 0$. It determines the direction incorrectly if this product is negative.² Similarly, we will say that side-information $x_k^{(2)}$ determines the correct direction with respect to the noise-free scene if $(x_k^{(2)} - x_k^{(1)})(r_k^{(1)}/q - x_k^{(1)}) > 0$. When this product is negative, it determines the direction incorrectly. When it is zero ($x_k^{(2)} = x_k^{(1)}$), the side-information is not useful.

Figure 1 shows the relative number of correctly and incorrectly determined embedding directions based on side-information in the form of one non-quantized DCT coefficient $c_k^{(1)}$ (Precover) and two quantized coefficients $x_k^{(1)}$ and $x_k^{(2)}$ (Two JPEGs). The most interesting part of the figure is for small values of q . Two quantized images are much more conservative in the sense that they determine the direction incorrectly much less frequently than from one non-rounded value. On the other hand, with increasing q , the two quantized images find fewer correct directions. For small values of $q = 1, 2, 3$ (more generally, for large values of σ_a/q), two JPEG images provide more useful side-information about the preferred changes compared to the non-rounded DCTs. This is in qualitative agreement

¹ $\sigma_a = 1$ approximately corresponds to acquisition noise with 1/60th sec. exposure at 100 ISO with Canon 6D.

²We can ignore the zero-probability event $r_k^{(1)}/q = x_k^{(1)}$.

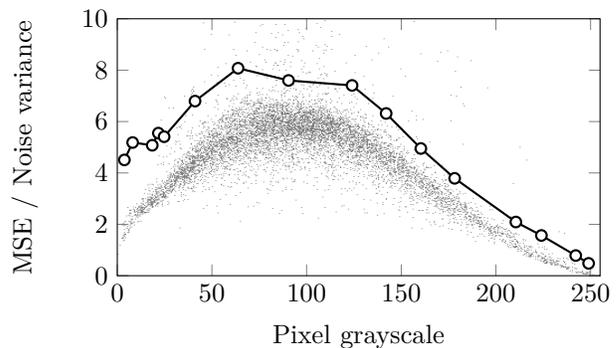


Figure 5. Gray dots: $\text{MSE}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)})$ vs. average grayscale of $\mathbf{z}^{(1)}$ across images from BURSTbase. Circles: acquisition noise variance estimated from images of gray wall. Both at ISO 200.

with Figure 3 that shows that J2-UNIWARD indeed outperforms SI-UNIWARD for high quality factors (small q).

Note that for side-information with non-rounded values $c_k^{(1)}$, the sum of the relative number of correctly and incorrectly determined directions is one while this is not the case for two quantized coefficients because “ties” $x_k^{(1)} = x_k^{(2)}$ occur with non-zero probability.

VI. DATASETS FOR EXPERIMENTS

In general, it is difficult to acquire two images of the same scene because the camera position may slightly change between the exposures even when mounted on a tripod due to vibrations caused by the shutter. Another potential source of differences is slightly varying exposure time and changing light conditions between exposures. To test the real-life performance of the proposed side-informed steganography in Section VII, we prepared two new datasets: BURSTbase with images obtained with a camera mounted on a tripod and BURSTbaseH with images shot from hand.

A. BURSTbase

To eliminate possible impact of flicker of artificial lights, all images were acquired in daylight, both indoor and outdoor, and without a flash. Canon 6D, a DSLR camera with a full-frame 20 MP CMOS sensor, set to ISO 200 was used in a burst mode. The shutter was operated with a two-second self-timer to further minimize vibrations due to operating the camera. To prevent the camera from changing the settings during the burst, it was used in manual mode. All images were acquired in the RAW CR2 format and then exported from Lightroom 5.7 to 24-bit TIFF format with no other processing applied.

We acquired 133 bursts, each containing 7 images. To increase the number of images for experiments, the 5472×3648 TIFF images were cropped into 10×7 equidistantly positioned tiles with 512×512 pixels. This required a slight overlap between neighboring tiles (7 pixels horizontally and 35 pixels vertically). These $70 \times 133 = 9,310$ smaller images were then converted to grayscale in Matlab using

'rgb2gray' and saved in a lossless raster format to facilitate experiments with a range of JPEG quality factors. We call this database of $7 \times 9, 310$ uncompressed grayscale images 'BURSTbase'.

For each pair of different images from each burst, we computed the mean square error (MSE) between them and then selected the pair with the smallest MSE, denoting one of them randomly as $z_{ij}^{(1)}$ and the other $z_{ij}^{(2)}$. The remaining five images from the burst were denoted $z_{ij}^{(k)}$, $k = 3, \dots, 7$, so that the MSE between $z_{ij}^{(1)}$ and $z_{ij}^{(k)}$ forms a non-decreasing sequence in k . We analyzed images from BURSTbase sorted in this manner to determine how much the differences between images are due to acquisition noise or slight spatial misalignment. Figure 4 shows the MSE between $z_{ij}^{(1)}$ and $z_{ij}^{(k)}$, $k = 2, \dots, 7$, averaged over the entire BURSTbase. For the closest pairs, $\text{MSE}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) \approx 5$, which would correspond to $\sigma_a^2 = 5$ if the differences were solely due to AWG noise with variance σ_a^2 . This closely matches the variance estimated from a single image of content-less scenes with medium gray. This reasoning indicates that $\mathbf{z}^{(2)}$ and $\mathbf{z}^{(3)}$ are on average reasonably well aligned with $\mathbf{z}^{(1)}$ while $\mathbf{z}^{(k)}$, $k \geq 4$, are increasingly more affected by small spatial shifts.

To obtain additional evidence that the differences between the two closest images from each burst are due to acquisition noise rather than slight spatial misalignment, we conducted another experiment in which we studied the MSE as a function of luminance. This was done to capture the dependence of the acquisition noise variance on luminance – it follows the heteroscedastic model further modified by tonal curve adjustment. To map out the dependence, we took RAW images of a uniform gray wall in the exposure priority mode with a wide range of exposures while all other settings were kept unchanged (at ISO 200). These flat-field images were then exported from Lightroom to 24-bit TIFF images, converted to grayscale using Matlab's 'rgb2gray', and cropped to the central 512×512 region. To isolate only the acquisition noise, a third-degree polynomial fit for each pixel on a sliding 32×32 block was subtracted from the pixels to remove any leftover gradual fall-off of luminance towards the image edges due to vignetting. Figure 5 shows the MSE as a function of the average image grayscale across BURSTbase, with the circles corresponding to variance–grayscale pairs from images of gray wall. The data is in qualitative agreement with the maximum variance for pixels with grayscale around 100. The decreased variance for grayscales below 100 is most likely due to the tonal adjustment done by cameras to avoid magnifying noise in underexposed areas.

B. BURSTbaseH

Since most casual photographers do not shoot from a tripod, we prepared a second dataset with images shot from hand to see whether the proposed modulation of costs still provides a boost under this more realistic and less ideal conditions. A different set of images was acquired using the same Canon 6D camera on a different day, this

Table I
MAXIMUM AND AVERAGE MSE BETWEEN TWO CLOSEST EXPOSURES FROM EACH BURST IN BURSTBASEH WHEN CONSTRAINING IT TO A FRACTION γ OF BEST BURSTS.

γ	1	0.5	0.2	0.1
max MSE	3790	100.1	25.42	12.94
avg MSE	254.23	39.10	13.32	7.81

time with the camera being hand-held instead of mounted on a tripod. A total of 154 bursts of 7–13 images were obtained that were processed and then cropped into 10,780 smaller 512×512 images in the same manner as described in the previous section. To distinguish this source from BURSTbase, we call this database BURSTbaseH (H as in **Hand-held**).

The average MSE between the two closest images from each burst was 254.23, which is significantly larger than for BURSTbase (5.05). This tells us that the images are on average misaligned by a large amount, which is likely to have a significant impact on the security of the proposed scheme. The steganographer, however, can reject bad bursts and/or take another one and only embed in images from bursts that are not grossly misaligned. In fact, many mobile devices today are capable of taking bursts, such as for HDR photography or to reduce high-ISO noise. The authors envision a mobile app that would leverage this capability for the purpose of increasing the security of steganographic communication. Another possibility to obtain well-aligned multiple exposures is to extract consecutive frames from short M-JPEG video clips. This, too, could be achieved with a mobile app.

Based on the considerations spelled out in the previous paragraph, in the next section we experiment with subsets of BURSTbaseH consisting of a fraction $\gamma \in [0, 1]$ of images with the smallest MSE for the closest pair. For example, in BURSTbaseH with $\gamma = 0.5$, we selected $10,780/2 = 5,390$ bursts with the smallest MSE, eliminating thus half of the bursts with the worst misalignment. Table I shows the average MSE between the closest pair of images when constraining BURSTbaseH to the fraction of $\gamma \in \{0.1, 0.2, 0.5, 1\}$ best bursts. Note that the average MSE between the two closest exposures from each burst in BURSTbaseH with $\gamma = 0.1$ is rather close to the MSE between the closest images of BURSTbase.

VII. EXPERIMENTS

In this section, we first study the empirical security of J2-UNIWARD on BURSTbase across a range of quality factors and payloads and contrast it with J-UNIWARD and SI-UNIWARD. We also assess how the security boost of the second exposure changes with increased differences between exposures. In the second round of experiments, we assess the performance of the proposed scheme in more realistic conditions when the bursts are taken with a hand-held camera instead of mounted on a tripod (BURSTbaseH). On tests with J2-UNIWARD and UED2-JC, we show that when bad bursts are rejected embedding with

two JPEGs still provides a significant performance boost with respect to embedding in single JPEGs despite rather large spatial misalignments.

Since the feedback from a detector utilizing the GFR feature set was used to determine the modulation factor, it is essential that we test J2-UNIWARD with other feature sets to evaluate its security. Thus, all experiments in this section were executed with a low-complexity linear classifier trained with the merger of the GFR features, the spatial rich model (SRM) [28], and the cartesian-calibrated JPEG Rich Model (ccJRM) [29].

A. BURSTbase

The modulation factor $m(Q)$ (10) found experimentally as described in Section IV is shown in Figure 2 right. All our experiments in this subsection were executed with $m(Q)$ approximated by a following ramp function:

$$m(Q) = \max\{0.075, 0.02167 \times Q - 1.55\}. \quad (11)$$

The appendix contains a simple qualitative argument explaining why the modulation factor follows a ramp function.

Figure 6 left shows \bar{P}_E as a function of the JPEG quality factor for payload 0.2 bpnzac together with the results for J-UNIWARD (with $x_k^{(1)}$ as covers) and SI-UNIWARD (with $c_k^{(1)}$ as side-information). For real acquisitions, the side-information in the form of two JPEG images significantly increases empirical security w.r.t. embedding with a single JPEG (J-UNIWARD). In contrast with the experiments with simulated acquisition noise, however, the empirical security is not better than when non-rounded DCT coefficients are used as side-information (SI-UNIWARD). For completeness, in Figure 6 right we report the detection error as a function of the quality factor for five payloads and in Table II we report all numerical values, including the results obtained with STCs with constraint height $h = 10$ rather than with an embedding simulator to see the coding loss.

To assess how sensitive J2-UNIWARD is w.r.t. small differences between exposures, we implemented it with $x_{ij}^{(1)}$ as cover and $x_{ij}^{(k)}$, $k = 3, \dots, 7$ as side-information, essentially using the second closest ($k = 3$), the third closest ($k = 4$), etc., image instead of the closest image. As apparent from Figure 4, with increasing k the MSE increases and thus the security boost should start diminish. Figure 7 shows \bar{P}_E as a function of the quality factor across $k = 2, \dots, 7$ together with the value of J-UNIWARD. While the gain of the second image indeed decreases with increased MSE, this decrease is rather gradual and very small for higher quality factors. This experiment proves that the second exposure provides useful side-information even when small spatial shifts are present opening thus the possibility to improve steganography even when multiple exposures are acquired with a hand-held camera rather than mounted on a tripod, a topic studied in the next section.

Table II
EMPIRICAL SECURITY \bar{P}_E OF EMBEDDING SCHEMES, \mathcal{M} , J-UNIWARD (J), J2-UNIWARD (J2), J2-UNIWARD IMPLEMENTED USING STCs (J2c), AND SI-UNIWARD (SI) ON BURSTBASE FOR A RANGE OF PAYLOADS, R , AND QUALITY FACTORS.

R	\mathcal{M}	Quality factor Q						
		65	75	85	87	90	92	95
0.1	SI	0.4991	0.4973	0.4897	0.4892	0.4952	0.4984	0.4525
	J	0.3508	0.3541	0.3766	0.4892	0.4121	0.4087	0.4421
	J2	0.4897	0.4659	0.4610	0.4633	0.4560	0.4523	0.4433
	J2c	0.4550	0.4591	0.4326	0.4289	0.4149	0.4138	0.4155
0.2	SI	0.4815	0.4811	0.4761	0.4753	0.4812	0.4811	0.4498
	J	0.1946	0.1953	0.2258	0.2301	0.2840	0.2787	0.3622
	J2	0.4620	0.4275	0.4178	0.4128	0.4161	0.4100	0.3796
	J2c	0.4146	0.4186	0.4179	0.4119	0.4103	0.3959	0.3695
0.3	SI	0.4501	0.4456	0.4406	0.4437	0.4506	0.4520	0.4200
	J	0.1010	0.0975	0.1179	0.1256	0.1771	0.1660	0.2647
	J2	0.4245	0.3827	0.3729	0.3723	0.3733	0.3560	0.3196
	J2c	0.3740	0.3709	0.3626	0.3524	0.3569	0.3346	0.2990
0.4	SI	0.4056	0.3989	0.3976	0.3963	0.4118	0.4037	0.4201
	J	0.0528	0.0469	0.0592	0.0627	0.0980	0.0906	0.1776
	J2	0.3734	0.3394	0.3144	0.3084	0.3218	0.2932	0.2647
	J2c	0.3356	0.3244	0.2949	0.2862	0.2976	0.2649	0.2380
0.5	SI	0.3552	0.3446	0.3392	0.3361	0.3571	0.3491	0.3779
	J	0.0280	0.0234	0.0289	0.0291	0.0506	0.0444	0.1076
	J2	0.3062	0.2989	0.2501	0.2383	0.2569	0.2168	0.2043
	J2c	0.2777	0.2815	0.2210	0.1991	0.2231	0.1848	0.1779

Table III
EMPIRICAL SECURITY \bar{P}_E OF EMBEDDING SCHEMES, \mathcal{M} , J-UNIWARD (J), J2-UNIWARD (J2) AND SI-UNIWARD (SI) ON BURSTBASEH FOR A RANGE OF PAYLOADS, R , AND QUALITY FACTORS FOR $\gamma = 0.1$.

R	\mathcal{M}	Quality factor Q						
		65	75	85	87	90	92	95
0.2	SI	0.4788	0.4706	0.4744	0.4697	0.4736	0.4739	0.4541
	J	0.2596	0.2600	0.2729	0.2769	0.2769	0.2996	0.3887
	J2	0.3786	0.3963	0.4084	0.4163	0.4250	0.4176	0.4260
0.4	SI	0.4372	0.4305	0.4186	0.4275	0.4442	0.4541	0.4363
	J	0.1267	0.1131	0.1000	0.1043	0.1356	0.3887	0.2399
	J2	0.2583	0.0075	0.3020	0.2956	0.3274	0.4260	0.3518

B. BURSTbaseH

To investigate the security of the proposed technique under more realistic setting, we experimented with J2-UNIWARD and UED2-JC on BURSTbaseH with $\gamma \in \{0.1, 0.2, 0.5, 1\}$ for a range of quality factors and payloads. For J2-UNIWARD, we reused the modulation factor $m(Q)$ determined on BURSTbase (Eq. (11)). Although we did perform a search for the best modulation factor for UED2-JC, the detection error was rather insensitive to $m(Q)$ as long as it was sufficiently small. In all our experiments with UED2-JC, the modulation factor was

Table IV
EMPIRICAL SECURITY \bar{P}_E OF EMBEDDING SCHEMES, \mathcal{M} , UED-JC (U), UED2-JC (U2), AND SI-UED-JC (SI) ON BURSTBASEH FOR TWO PAYLOADS AND TWO JPEG QUALITY FACTORS FOR $\gamma = 0.1$.

R	\mathcal{M}	Quality factor Q	
		75	95
0.2	SI	0.2185	0.2893
	U	0.0462	0.1318
	U2	0.1995	0.3547
0.4	SI	0.0970	0.2477
	U	0.0250	0.0706
	U2	0.1032	0.1884

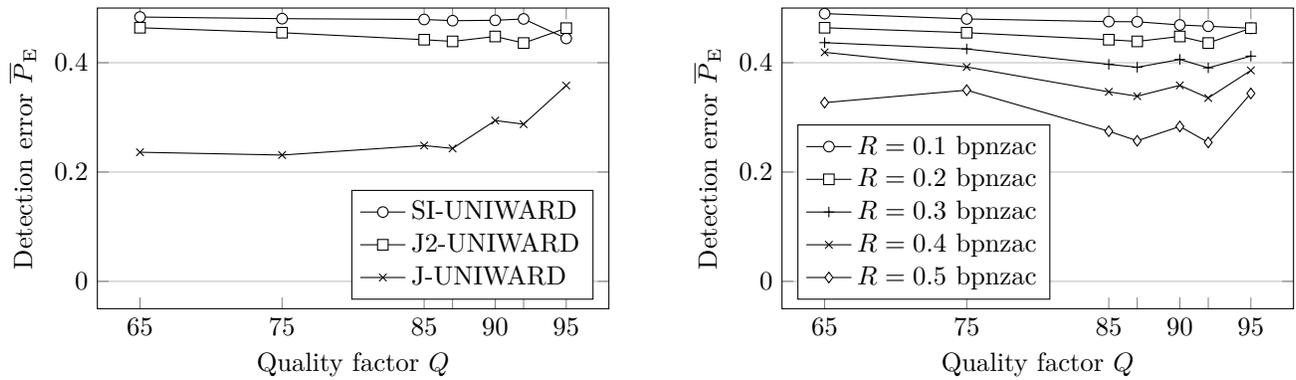


Figure 6. Empirical security \bar{P}_E of J2-UNIWARD as a function of the JPEG quality factor Q on BURSTbase. Left: Comparison with previous art for $R = 0.2$ bpnzac. Right: J2-UNIWARD \bar{P}_E for $R \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ bpnzac, embedding simulated at rate-distortion bound.

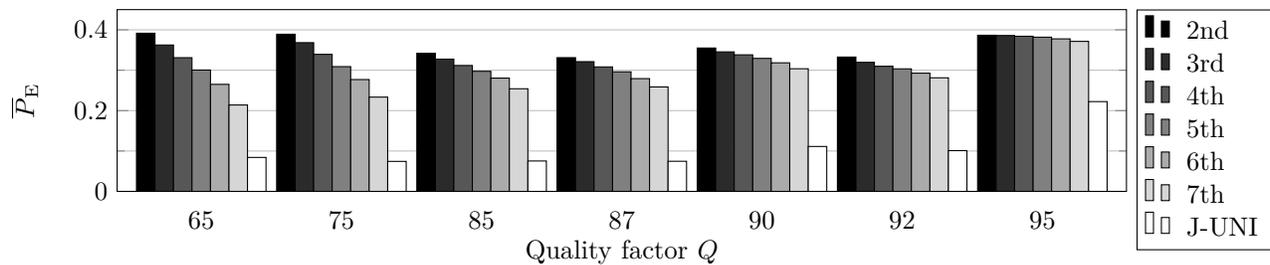


Figure 7. Empirical security \bar{P}_E of J2-UNIWARD when the k th closest image from each burst from BURSTbase was used as side-information. Payload $R = 0.4$ bpnzac.

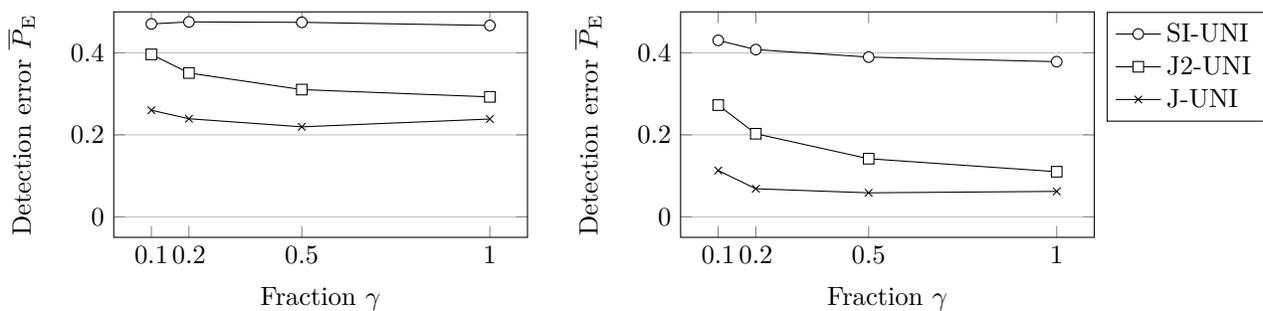


Figure 8. Empirical security \bar{P}_E of J-UNIWARD, J2-UNIWARD, and SI-UNIWARD as a function of γ best bursts from BURSTbaseH. JPEG quality factor 75, left column 0.2 bpnzac, right column 0.4 bpnzac.

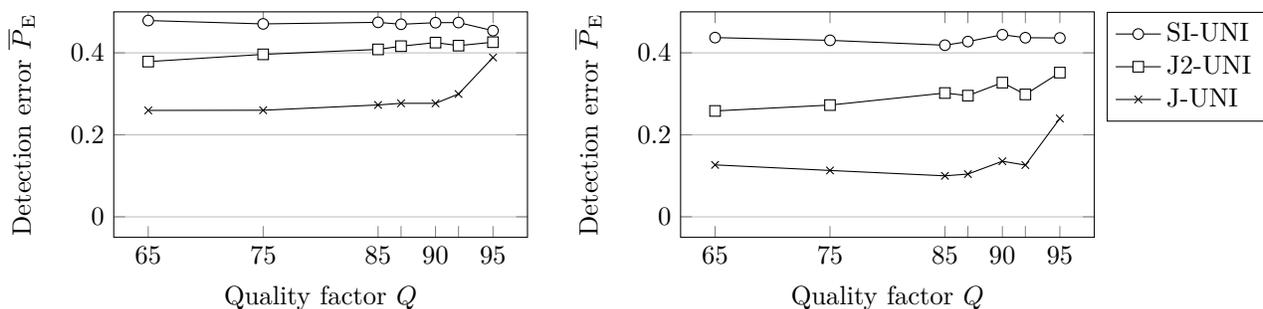


Figure 9. Empirical security \bar{P}_E of J-UNIWARD, J2-UNIWARD, and SI-UNIWARD as a function of JPEG quality factor Q for $\gamma = 0.1$ best bursts from BURSTbaseH. Left column 0.2 bpnzac, right column 0.4 bpnzac.

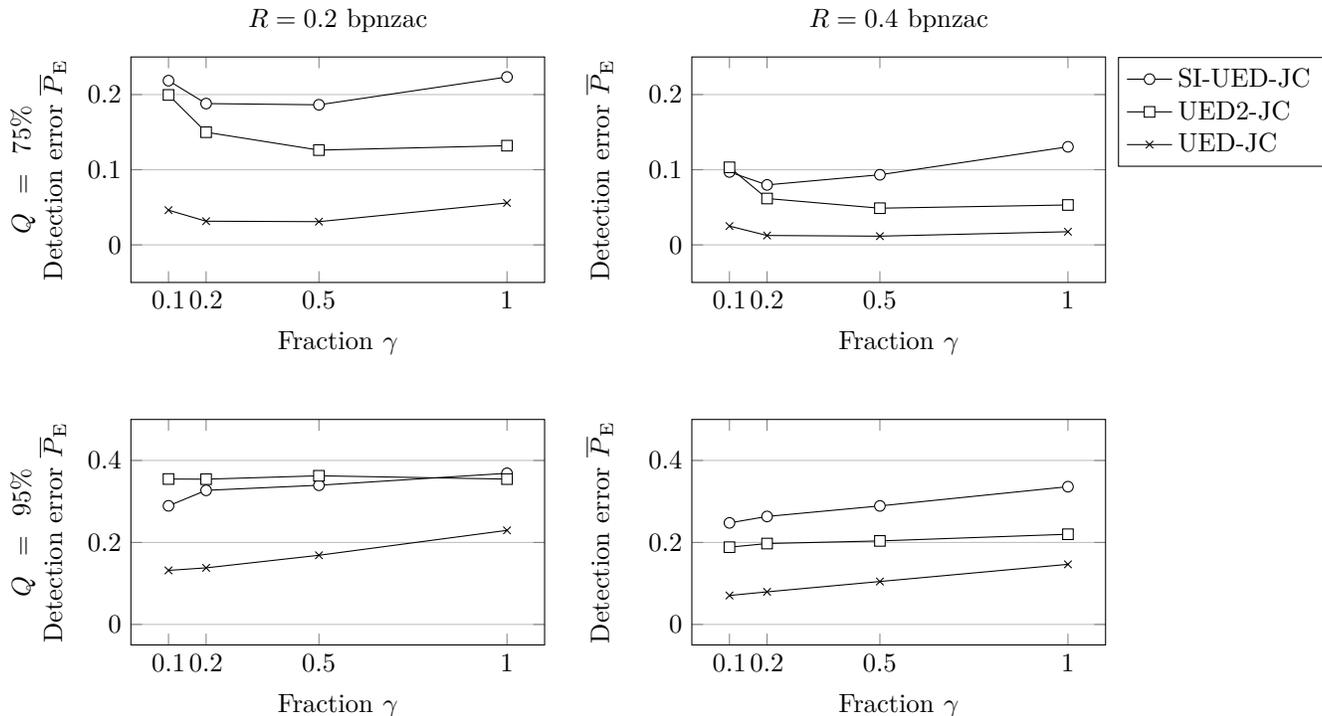


Figure 10. Empirical security \bar{P}_E of UED-JC, UED2-JC, and SI-UED-JC as a function of γ best bursts from BURSTbaseH for two JPEG quality factors and two payloads.

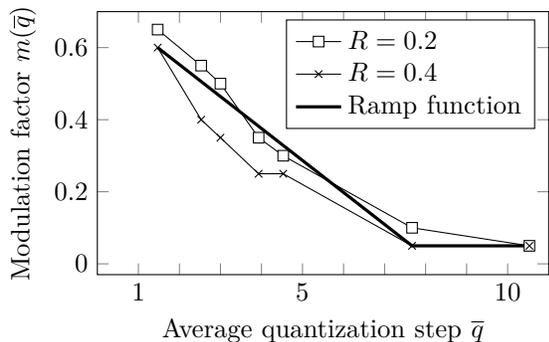


Figure 11. Modulation factor versus average quantization step \bar{q} (real acquisitions).

set as $m(Q) = 0.01$ for all tested payloads and quality factors.

Figure 8 shows the detection error \bar{P}_E for two payloads for JPEG quality factor 75 for all four values of γ for J-UNIWARD, J2-UNIWARD, and SI-UNIWARD with the same steganalysis detector as in the previous section. Figure 9 contains the detection error for the same three embedding schemes as a function of JPEG quality factor for $\gamma = 0.1$. Both figures demonstrate a substantial gain in security of J2-UNIWARD w.r.t. J-UNIWARD. While this gain is understandably smaller for the images of BURSTbaseH, it becomes substantial in comparison with embedding with a single JPEG image as the number of rejected bursts increases. The numerical values of \bar{P}_E of all experiments are provided in Table III.

In Figure 10, we display the detection error as a function of γ for two payloads and two quality factors for the UED-JC embedding algorithm. Here, the bad burst rejection is even more effective than for J2-UNIWARD. For quality factor 95, UED2-JC even outperforms UED informed by the precover (SI-UED-JC) for all $\gamma < 1$. Substantial security gain is observed even for $\gamma = 0.5$, e.g., when every other burst is rejected on average, across all payloads and quality factors.

VIII. CONCLUSIONS

We introduce a novel steganographic method with side-information at the sender in the form of a second JPEG image of the same scene. The second exposure is used to infer the preferred direction of steganographic embedding changes in the first exposure (cover). This information is incorporated in any cost-based steganography by decreasing the embedding costs of such preferred changes with a multiplicative modulation factor.

The proposed methodology is first studied on J-UNIWARD costs with multiple exposures simulated by adding AWG noise to BOSSbase 1.01 images. This experiment revealed that, under such ideal conditions, the proposed method with two JPEG images of the same scene exhibits empirical security comparable with and sometimes even better than SI-UNIWARD informed by the uncompressed precover. This observation was attributed to the fact that for larger quality factors two JPEGs better inform the sender about the preferred embedding change direction than one uncompressed image.

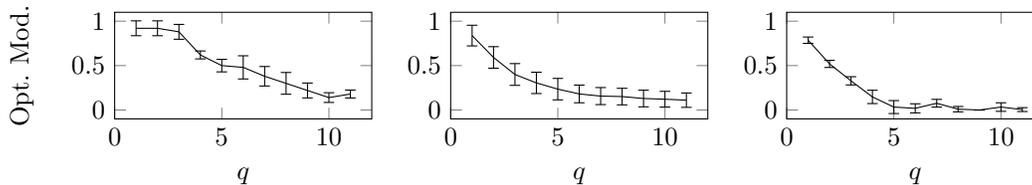


Figure 12. Optimal modulation factor $m_{ij}(q, R)$ as a function of the quantization step q for relative payload $R = 0.4$ determined by minimizing the Bhattacharyya distance between cover and stego distributions on generalized Gaussian models of DCT coefficients. Left: low frequency DCT modes (i, j) , $3 \leq i + j \leq 4$ (second and third minor diagonal), Middle: medium frequency DCT modes (i, j) , $5 \leq i + j \leq 10$, Right: high frequency DCT modes (i, j) , $11 \leq i + j \leq 16$.

To evaluate the proposed method in real-life conditions, we created two new datasets: BURSTbase with multiple exposures obtained by a tripod-mounted camera and BURSTbaseH with images shot with a hand-held camera. Detailed analysis of the differences between the two closest exposures from BURSTbase confirmed that they differ mostly by the acquisition noise, while images from BURSTbaseH are generally significantly much more spatially misaligned due to camera shake.

For BURSTbase, we observed a quite significant increase in empirical security with respect to steganography with a single cover image that gracefully decreased with increased spatial misalignment between images. On the other hand, because of the comparatively larger misalignments between images shot with a hand-held camera the security improvement on BURSTbaseH was understandably smaller. However, we demonstrated for both J-UNIWARD and UED-JC, that the sender can still significantly gain on empirical security by rejecting a portion of “bad bursts”, which testifies about the practicality of the proposed embedding scheme.

Finally, the dependence of the experimentally determined modulation factor on the quality factor is justified using Monte Carlo simulations by adopting generalized Gaussian model for DCT coefficients and measuring the impact of cost modulation on statistical detectability in terms of the Bhattacharyya distance between cover and stego distributions. Optimal modulation derived from this model qualitatively matches the modulation obtained experimentally on real multiple exposures.

Further improvement is likely possible by optimizing the embedding cost modulation for the average grayscale of the DCT block because the acquisition noise amplitude depends on luminance. We plan to further study how the embedding should utilize more than two (quantized and unquantized) acquisitions of the same scene, possibly by extending the approach proposed in [30]. We anticipate that the proposed methodology will also work with multiple exposures obtained as consecutive frames from video clips. Finally, we note that the proposed approach is not limited to JPEG domain and will likely work for side-informed embedding in other domains [10].

APPENDIX

In this appendix, we provide some insight into why the experimentally-found optimal modulation factor follows

the ramp function (11) depicted in Figure 2. First, in Figure 11 we redraw the modulation factor shown in Figure 2 right as a function of the average quantization step $\bar{q} = 1/15 \sum_{i+j \leq 5} q_{ij}$ instead of the quality factor Q . We only average the first five diagonals of the quantization matrix because this is where the vast majority of differences between two JPEG files occur ($x_{ij}^{(1)} \neq x_{ij}^{(2)}$). *This figure tells us that the modulation factor should be smaller for larger quantization steps and vice versa.* This important observation is validated via the following experiment.

A total of 100 random images from BOSSbase 1.01 were selected. A generalized Gaussian distribution (1) was fitted using the method of moments [31] to each AC DCT mode (i, j) across all 100 images, obtaining thus 63 values of the shape and width parameters α_{ij} , b_{ij} , $1 \leq i, j \leq 8$, $i + j > 2$. For each AC DCT mode (i, j) and for each quantization step q , we twice generated $N_{MC} = 10^8$ independent realizations from $\mathcal{G}(0, \alpha_{ij}, b_{ij})$, denoting them $r_k^{(1)}$ and $r_k^{(2)}$, $k \in \{1, \dots, N_{MC}\}$, and N_{MC} independent realizations $\xi_k^{(1)}$ and $\xi_k^{(2)}$ from $\mathcal{N}(0, 1)$, the acquisition noise. The non-rounded DCT coefficients and their rounded values were computed and denoted $c_k^{(l)} = (r_k^{(l)} + \xi_k^{(l)})/q$ and $x_k^{(l)} = \lfloor c_k^{(l)} \rfloor$, $l = 1, 2$. Next, we simulated J2-UNIWARD with $x_k^{(1)}$ as the cover and $x_k^{(2)}$ as the side-information with $\rho_{ij}^{(1)} = 1$ for all i, j modulated as in (10). The embedding was simulated with change probabilities as explained in Section III for a fixed relative payload $R = 0.4$ measured w.r.t. the number of non-zero coefficients, $N_0 = |\{k | x_k^{(1)} \neq 0\}|$, giving us the stego object $y_k \in \{x_k^{(1)} - 1, x_k^{(1)}, x_k^{(1)} + 1\}$. The impact of embedding on the cover model was measured by computing the complement of the Bhattacharyya coefficient³ between the sample cover and stego distributions, $\mathbf{p}^{(x)}, \mathbf{p}^{(y)}$:

$$B(\mathbf{p}^{(x)}, \mathbf{p}^{(y)}) = 1 - \sum_r \sqrt{p_r^{(x)} p_r^{(y)}} \quad \text{where} \quad (12)$$

$$p_r^{(x)} = \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} [x_k^{(1)} = r], r \in \mathbb{Z} \quad (13)$$

$$p_r^{(y)} = \frac{1}{N_{MC}} \sum_{k=1}^{N_{MC}} [y_k = r], r \in \mathbb{Z}. \quad (14)$$

³Since the Bhattacharyya distance is $B_{\text{dist}} = -\log(1 - B)$, B reaches its minimum exactly when B_{dist} does.

Above, $[P]$ denotes the Iverson bracket, $[P] = 1$ when P is true and 0 when P is false. The exact range of index r depends on the specific realizations generated. The Bhattacharyya coefficient was selected for its good numerical stability w.r.t. unpopulated bins.

Since the quantized cover and stego DCT coefficients $x_k^{(1)}$ and y_k depend on the DCT mode (i, j) , the quantization step q , and relative payload R , the sample distributions $\mathbf{p}^{(x)}$, $\mathbf{p}^{(y)}$ and thus $B(\mathbf{p}^{(x)}, \mathbf{p}^{(y)})$ also depend on these parameters. The optimal value of the modulation parameter, $m_{ij}(q, R)$, was determined for each DCT mode (i, j) by minimizing $B(\mathbf{p}^{(x)}, \mathbf{p}^{(y)})$ over $m \in [0, 1]$:

$$m_{ij}(q, R) = \arg \min_{m \in [0, 1]} B(\mathbf{p}^{(x)}, \mathbf{p}^{(y)}). \quad (15)$$

The optimal values of the modulation parameter as a function of the quantization step q are shown in Figure 12 for low, mid, and high-frequency DCT modes for payload $R = 0.4$. The error bars are across the DCT modes from the frequency band. We observe that the modulation mainly depends on q and stays approximately constant over DCT modes for each frequency band. The dependence on the quantization step q is qualitatively and quantitatively similar to Figure 11, validating thus our design choice.

REFERENCES

- [1] R. Böhme, *Advanced Statistical Steganalysis*, Springer-Verlag, Berlin Heidelberg, 2010.
- [2] A. D. Ker, “A fusion of maximal likelihood and structural steganalysis,” in *Information Hiding, 9th International Workshop*, T. Furon, F. Cayre, G. Doërr, and P. Bas, Eds., Saint Malo, France, June 11–13, 2007, vol. 4567 of Lecture Notes in Computer Science, pp. 204–219, Springer-Verlag, Berlin.
- [3] J. Fridrich and R. Du, “Secure steganographic methods for palette images,” in *Information Hiding, 3rd International Workshop*, A. Pfitzmann, Ed., Dresden, Germany, September 29–October 1, 1999, vol. 1768 of Lecture Notes in Computer Science, pp. 47–60, Springer-Verlag, New York.
- [4] J. Fridrich, M. Goljan, and D. Soukal, “Perturbed quantization steganography using wet paper codes,” in *Proceedings of the 6th ACM Multimedia & Security Workshop*, J. Dittmann and J. Fridrich, Eds., Magdeburg, Germany, September 20–21, 2004, pp. 4–15.
- [5] Y. Kim, Z. Duric, and D. Richards, “Modified matrix encoding technique for minimal distortion steganography,” in *Information Hiding, 8th International Workshop*, J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, Eds., Alexandria, VA, July 10–12, 2006, vol. 4437 of Lecture Notes in Computer Science, pp. 314–327, Springer-Verlag, New York.
- [6] V. Sachnev, H. J. Kim, and R. Zhang, “Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding,” in *Proceedings of the 11th ACM Multimedia & Security Workshop*, J. Dittmann, S. Craver, and J. Fridrich, Eds., Princeton, NJ, September 7–8, 2009, pp. 131–140.
- [7] F. Huang, J. Huang, and Y.-Q. Shi, “New channel selection rule for JPEG steganography,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1181–1191, August 2012.
- [8] L. Guo, J. Ni, and Y. Q. Shi, “Uniform embedding for efficient JPEG steganography,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 814–825, May 2014.
- [9] V. Holub, J. Fridrich, and T. Denemark, “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*, vol. 2014:1, 2014.
- [10] T. Denemark and J. Fridrich, “Side-informed steganography with additive distortion,” in *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19, 2015.
- [11] E. Franz, “Steganography preserving statistical properties,” in *Information Hiding, 5th International Workshop*, F. A. P. Petitcolas, Ed., Noordwijkerhout, The Netherlands, October 7–9, 2002, vol. 2578 of Lecture Notes in Computer Science, pp. 278–294, Springer-Verlag, New York.
- [12] E. Franz and A. Schneidewind, “Pre-processing for adding noise steganography,” in *Information Hiding, 7th International Workshop*, M. Barni, J. Herrera, S. Katzenbeisser, and F. Pérez-González, Eds., Barcelona, Spain, June 6–8, 2005, vol. 3727 of Lecture Notes in Computer Science, pp. 189–203, Springer-Verlag, Berlin.
- [13] E. Franz, “Embedding considering dependencies between pixels,” in *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, Eds., San Jose, CA, January 27–31, 2008, vol. 6819, pp. D 1–12.
- [14] K. Petrowski, M. Kharrazi, H. T. Sencar, and N. Memon, “PSTEG: steganographic embedding through patching [image steganography],” in *Proc. IEEE ICASSP*, Philadelphia, PA, March 18–23, 2005.
- [15] T. Denemark and J. Fridrich, “Side-informed steganography with two JPEGs,” in *Proc. IEEE ICASSP*, New Orleans, LA, March 5–8, 2017.
- [16] J. R. Janesick, *Scientific Charge-Coupled Devices*, vol. Monograph PM83, Washington, DC: SPIE Press - The International Society for Optical Engineering, January 2001.
- [17] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, “Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data,” *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1737–1754, October 2008.
- [18] Thanh Hai Thai, R. Cogranne, and F. Retraint, “Camera model identification based on the heteroscedastic noise model,” *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 250–263, January 2014.
- [19] A. Sarkar, K. Solanki, and B. S. Manjunath, “Further study on YASS: Steganography based on randomized embedding to resist blind steganalysis,” in *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, Eds., San Jose, CA, January 27–31, 2008, vol. 6819, pp. 16–31.
- [20] T. Filler, J. Judas, and J. Fridrich, “Minimizing additive distortion in steganography using syndrome-trellis codes,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, September 2011.
- [21] J. Fridrich, M. Goljan, D. Soukal, and P. Lisoněk, “Writing on wet paper,” in *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VII*, E. J. Delp and P. W. Wong, Eds., San Jose, CA, January 16–20, 2005, vol. 5681, pp. 328–340.
- [22] J. Fridrich, “On the role of side-information in steganography in empirical covers,” in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2013*, A. Alattar, N. D. Memon, and C. Heitzenrater, Eds., San Francisco, CA, February 5–7, 2013, vol. 8665, pp. 0I 1–11.
- [23] C. Wang and J. Ni, “An efficient JPEG steganographic scheme based on the block-entropy of DCT coefficients,” in *Proc. of IEEE ICASSP*, Kyoto, Japan, March 25–30, 2012.
- [24] R. Cogranne, V. Sedighi, T. Pevný, and J. Fridrich, “Is ensemble classifier needed for steganalysis in high-dimensional feature spaces?,” in *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19, 2015.
- [25] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, “Steganalysis of adaptive JPEG steganography using 2D Gabor filters,” in *3rd ACM IH&MMSec. Workshop*, P. Comesana, J. Fridrich, and A. Alattar, Eds., Portland, Oregon, June 17–19, 2015.
- [26] 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.
- [27] P. Bas, T. Filler, and T. Pevný, “Break our steganographic system – the ins and outs of organizing BOSS,” in *Information*

- Hiding*, 13th International Conference, T. Filler, T. Pevný, A. Ker, and S. Craver, Eds., Prague, Czech Republic, May 18–20, 2011, vol. 6958 of Lecture Notes in Computer Science, pp. 59–70, Springer, Berlin Heidelberg.
- [28] J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, June 2011.
- [29] J. Kodovský and J. Fridrich, “Steganalysis of JPEG images using rich models,” in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, A. Alattar, N. D. Memon, and E. J. Delp, Eds., San Francisco, CA, January 23–26, 2012, vol. 8303, pp. 0A 1–13.
- [30] T. Denmark and J. Fridrich, “Model based steganography with precover,” in *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2017*, A. Alattar and N. D. Memon, Eds., San Francisco, CA, January 29 – February 2, 2017.
- [31] S. Meignen and H. Meignen, “On the modeling of DCT and subband image data for compression,” *IEEE Transactions on Image Processing*, vol. 4, no. 2, pp. 186–193, February 1995.



Tomas Denmark received his M.S. in mathematical modeling from the Czech Technical University in Prague in 2012 and is currently pursuing the Ph.D. degree at the Thomas J. Watson School of Engineering and Applied Science in the Department of Electrical and Computer Engineering at Binghamton University (SUNY) under the lead of Jessica Fridrich. His

research focuses on steganography, steganalysis, machine learning and deep learning.



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 21 research grants totaling over \$11 mil for projects on data embedding and steganalysis that lead to more than 170 papers and 7 US patents. Dr. Fridrich is an IEEE Fellow and an ACM member.