STEGANOGRAPHY WITH TWO JPEGS OF THE SAME SCENE

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

Binghamton University
Department of ECE
Binghamton, NY

ABSTRACT

It is widely recognized that incorporating side-information at the sender can significantly improve steganographic security in practice. Currently, most side-informed schemes for digital images 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 in the form of two JPEG images of the same scene. The second JPEG image is used to determine the preferred polarity of embedding changes and to modulate their costs. Tests on real imagery show a very significant improvement in empirical security with respect to steganography utilizing a single JPEG image.

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

1. INTRODUCTION

Steganography is a private communication tool in which secrets are embedded in cover objects to hide the presence of the message itself. In side-informed steganography, the sender utilizes information that is unavailable to the steganalyst (and the recipient) to improve security. For example, the embedding can take place while processing (compressing) a higher quality representation of the cover image called precover [1]. The most common example of this type of steganography uses non-rounded DCT coefficients when saving an uncompressed image as JPEG [2, 3, 4, 5, 6, 7, 8].

Most consumer-end electronic devices, such as cell phones, tablets, and low-end digital cameras, however, can save images only in the JPEG format and thus do not give the user access to the uncompressed image. In this case, one can utilize a different type of side-information – multiple JPEG images of the same scene. This research direction has not been developed much mostly due to the difficulty of acquiring the required imagery and modeling the differences between acquisitions. The first work on this topic includes [9, 10, 11] where the authors made multiple scans of the same printed image and then modeled the differences between scans and among neighboring pixels. 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 scans lead to misalignment that complicates using this type of side-information properly.

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 imperfections during acquisition. 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 [9, 10, 11] and make the approach work well even when mere *two* images are available to the sender. In particular, we modulate the embedding costs of J-UNIWARD [7] based on the preferred direction inferred from two JPEG images of the same scene. The method is tested on real-life multiple exposures obtained using a tripod-mounted digital camera. The proposed embedding with two JPEG images is substantially more secure than when only a single JPEG is available to the steganographer.

In the next section, we review existing side-informed steganography with a high quality precover. The new steganographic method for embedding with two JPEGs is detailed in Section 3. In Section 4, we describe and analyze the image source used for experiments in Section 5. The same section contains a comparison with J-UNIWARD and SI-UNIWARD as well as a study of how the security gain due to the second JPEG changes with differences between exposures. The paper is concluded in Section 6.

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.

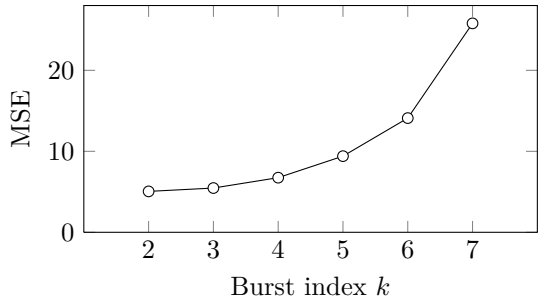

Fig. 1. MSE between $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(k)}$, $k = 2, \ldots, 7$ from each burst averaged over all 9,310 bursts from BURSTbase. See the main text for notation.

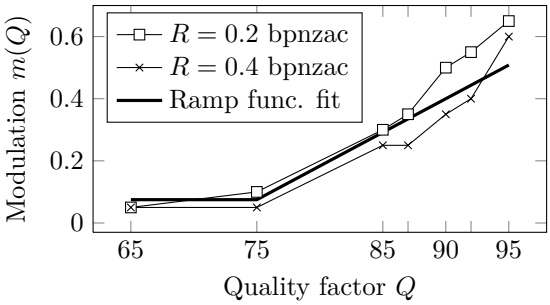

Fig. 2. Modulation factor $m(Q)$ as a function of the JPEG quality factor Q for images from BURSTbase.

2. STEGANOGRAPHY WITH PRECOVER

Virtually all modern embedding schemes for JPEG images, whether or not they use side-information, are implemented within the paradigm of distortion minimization. The sender first defines the cost of modifying each cover element (DCT coefficient) and then embeds the payload so that the expected value of the total cost is as small as possible. Syndrome-trellis codes [12] can be used to implement the embedding in practice.

For simplicity, we work with 8-bit $M \times N$ grayscale images with M and N multiples of 8. The non-rounded and rounded values of DCT coefficients from (u, v)th 8×8 block will be denoted $c_{ij}^{(u,v)} \in \mathbb{R}$ and $x_{ij}^{(u,v)} \in \{-1023, \ldots, 1024\}$, $1 \le i, j \le 8$, $1 \le u \le M/8, 1 \le v \le N/8$, respectively. The cost of changing $x_{ij}^{(u,v)}$ by 1 and -1 is $\rho_{ij}^{(u,v)}(1)$ and $\rho_{ij}^{(u,v)}(-1)$, respectively. When the symbols $x_{ij}, c_{ij}, \rho_{ij}$ are used without superscripts, the range of i, j spans the entire $M \times N$ image. 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 $y_{ij} \in \{x_{ij} - 1, x_{ij}, x_{ij} + 1\}$ is the stego image. An embedding scheme operating at the rate–distortion bound (with minimal D) would embed a payload of R bits by modifying the DCT coefficients with probabilities:

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

where λ is determined from the payload constraint $R = \sum_{ij} h_3(\beta_{ij}^{+}, \beta_{ij}^{-})$, 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. One of the most secure schemes for JPEG images called J-UNIWARD [7] computes the costs from the decompressed JPEG image. The costs are symmetric $\rho_{ij}(1) = \rho_{ij}(-1)$ for all i, j.

While it is currently an open problem how to use side-information (c_{ij}) in an optimal fashion for embedding [13], numerous heuristic schemes have been proposed in the past [3, 5, 6, 7, 8, 4]. In a nut shell, these schemes use the rounding error $e_{ij} = c_{ij} - x_{ij}$, $-1/2 \le e_{ij} \le 1/2$, to modulate the embedding costs ρ_{ij} by $1 - 2|e_{ij}| \in [0, 1]$. In SI-UNIWARD [7], for example, the costs are:

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

$$\rho_{ij}(-\text{sign}(e_{ij})) = C_{\text{wet}}, \tag{3}$$

where $\rho_{ij}^{(\text{J})}$ are J-UNIWARD costs and C_{wet} is some large number ("wet cost"). In [8], 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 $\rho_{ij}(-\text{sign}(e_{ij})) = \rho_{ij}^{(\text{J})}$ instead of (3).

3. STEGANOGRAPHY WITH TWO JPEGS

Let us consider a situation when the sender acquires two JPEG images of the same scene, $x_{ij}^{(1)}$ and $x_{ij}^{(2)}$, while pronouncing, e.g., the first image as cover JPEG and considering $x_{ij}^{(2)}$ as side-information. The value $x_{ij}^{(2)}$ can only be useful to the sender when $x_{ij}^{(2)} \neq x_{ij}^{(1)}$, which happens increasingly more often with smaller quantization steps (larger JPEG quality). This type of side-information is different from the non-rounded values $c_{ij}^{(1)}$. In particular, it informs the sender more about the direction along which the costs should be modulated and less about the magnitude of the rounding error $e_{ij}^{(1)} = c_{ij}^{(1)} - x_{ij}^{(1)}$.

The proposed embedding scheme, which we call J2-UNIWARD, uses J-UNIWARD costs [7] when $x_{ij}^{(1)} = x_{ij}^{(2)}$ and modulated costs otherwise:

$$\rho_{ij}(\text{sign}(x_{ij}^{(2)} - x_{ij}^{(1)})) = \begin{cases} \rho_{ij}^{(\text{J})} & \text{if } x_{ij}^{(1)} = x_{ij}^{(2)} \\ m(Q)\rho_{ij}^{(\text{J})} & \text{if } x_{ij}^{(1)} \neq x_{ij}^{(2)}, \end{cases} \tag{4}$$

with the modulation factors $m(Q) \in [0, 1]$ to be determined experimentally for each JPEG quality factor $1 \le Q \le 100$.

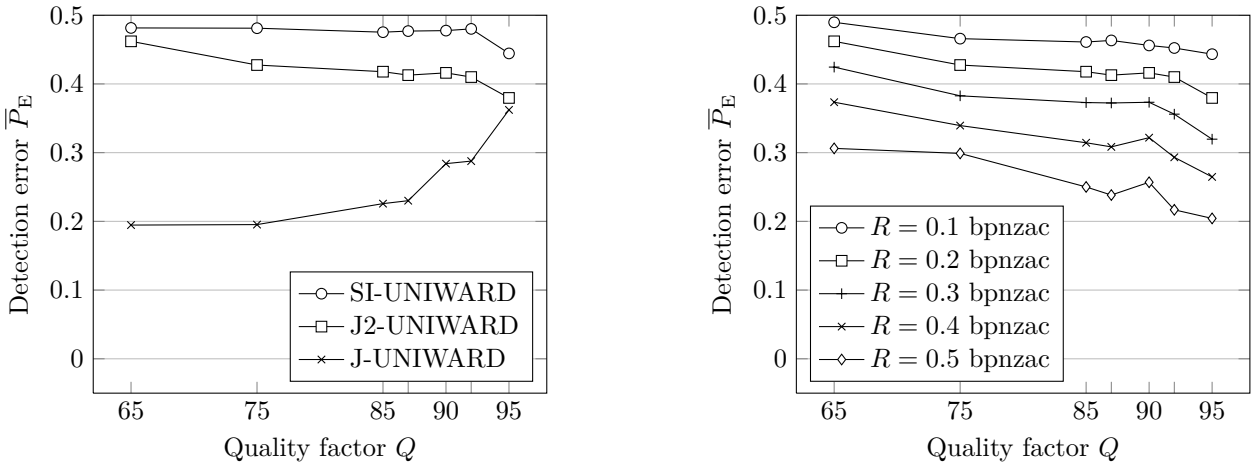

Fig. 3. Empirical security of J2-UNIWARD as a function of the JPEG quality factor Q with the merger of GFR, SRM, and ccJRM features. Left: Comparison with previous art for $R = 0.2$ bpnzac. Right: $\overline{P}_{\rm E}$ for $R \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ bpnzac, embedding simulated at rate–distortion bound.

4. THE BURSTBASE DATASET

It is generally difficult to acquire two images of the exact 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 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 a fixed ISO of 200 was used in a burst mode. The shutter was operated using a cable release 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.

A total of 133 bursts were acquired, 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'. The images were further JPEG compressed with different quality factors for all experiments in this paper.

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, randomly denoting one 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. Next, we analyzed images from BURSTbase sorted in this manner to determine how much the differences between the images are due to acquisition noise. The MSE between $z_{ij}^{(1)}$ and $z_{ij}^{(k)}$, $k = 2, \dots, 7$, averaged over the BURSTbase is plotted in Figure 1. For the closest pair, $\mathrm{MSE}(\mathbf{z}^{(1)}, \mathbf{z}^{(2)}) \approx 5$, 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, such as blue sky. 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 affected by small spatial shifts.

5. EXPERIMENTS

In this section, the security of J2-UNIWARD is studied across a range of quality factors and payloads and contrasted with the same scheme utilizing a single JPEG image and a scheme utilizing a single high-quality precover. We also investigate the security boost of the second exposure with increased differences between exposures.

The modulation factor $m(Q)$ (4) was determined for each quality factor Q to minimize $P_{\rm E} = \min_{P_{\rm FA}}(P_{\rm MD} + P_{\rm FA})/2$, the minimal total probability of error on the training set, where $P_{\rm MD}$, $P_{\rm FA}$ are missed-detection and false-alarm rates of a detector implemented using the

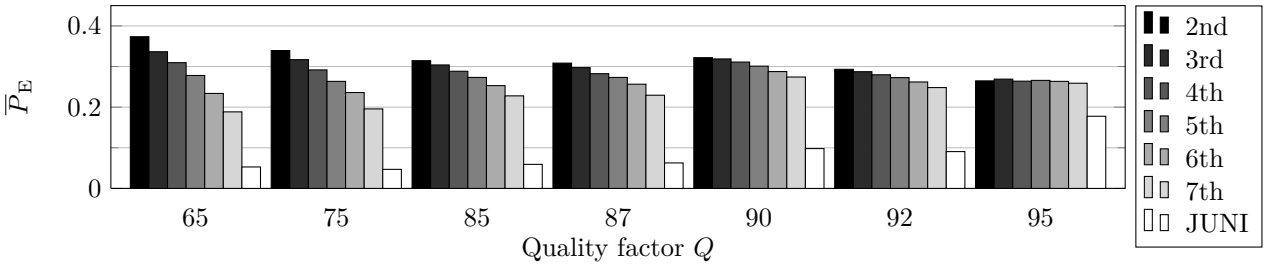

Fig. 4. Security of J2-UNIWARD when kth closest image from each burst is used as side-information, 0.4 bpnzac.

ensemble classifier [14] with GFR (Gabor Filter Residual) features [15] when splitting BURSTbase into equally sized training and testing set. The GFR features were selected because they are known to be highly effective against modern JPEG steganography, including J-UNIWARD and SI-UNIWARD. The optimal modulation factor determined experimentally and shown in Figure 2 can be well approximated by a ramp function:

$$m(Q) = \max\{0.075, 0.02167 \times Q - 1.55\}. \quad (5)$$

The ramp function can be justified when adopting a generalized Gaussian model of precover JPEG DCT coefficients and an AWG model of the acquisition noise. This argument is omitted here due to space limitations and will appear in the journal version of this paper [16]. The largest observed loss in P_{E} due to replacing optimal values of $m(Q)$ with the ramp function was about 0.01.

Because the feedback from detection with GFR features was used to design the embedding scheme, all detectors in this section were implemented with a diverse feature set that is a merger of the spatial rich model (SRM) [17], Cartesian-calibrate JPEG Rich Model (ccJRM) [18], and GFR to make sure the embedding does not have a fatal weakness with respect to older features. Figure 3 left shows P_{E} averaged over ten splits of BURSTbase into training and testing sets (denoted $\overline{P}_{\mathrm{E}}$) 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). The side-information in the form of two JPEG images significantly increases empirical security w.r.t. embedding with a single JPEG (J-UNIWARD) especially for large payloads and small quality factors. The empirical security is however not better than when non-rounded DCT coefficients are used as side-information (SI-UNIWARD). Figure 3 right shows the detection error as a function of the quality factor for five payloads. Since the statistical spread of P_{E} over the splits ranged between $0.0010 - 0.0075$, we do not show the error bars in the figure as it would be hard to discern them visually.

To assess how sensitive J2-UNIWARD is w.r.t. small differences between exposures, we implemented the scheme with $x_{ij}^{(1)}$ as cover and $x_{ij}^{(k)}$, $k = 3, \ldots, 7$ 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 1, with increasing k, the boost should start decreasing. Figure 4 shows $\overline{P}_{\mathrm{E}}$ as a function of the quality factor across $k = 2, \ldots, 7$ together with the value of J-UNIWARD (JUNI). While the gain of the second image indeed decreases with increased MSE, this decrease is gradual and rather 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 the multiple exposures are acquired with a hand-held camera rather than mounted on a tripod. This possibility is left as part of future research.

6. CONCLUSIONS

We study steganography with side-information at the sender in the form of a second JPEG image of the same scene that is used to infer the preferred direction of steganographic embedding changes. This information is incorporated into the embedding algorithm by decreasing (modulating) the embedding costs of such preferred changes. Experiments with real multiple acquisitions show a quite significant increase in empirical security of with respect to steganography with a single cover image (J-UNIWARD). The boost in empirical security appears fairly insensitive to small differences between the two acquisitions, which makes the proposed method practical and opens up the possibility to use multiple exposures obtained using a hand-held camera or acquiring multiple exposures from short video clips.

Further improvement is likely possible by optimizing the embedding cost modulation for each DCT mode, quantization step, and the average grayscale of the DCT block because the acquisition noise amplitude depends on luminance. Finally, we plan to study how to utilize more than two (quantized and unquantized) acquisitions.

7. REFERENCES

[1] 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 *LNCS*, pp. 204–219, Springer-Verlag, Berlin.

[2] 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.

[3] 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 *LNCS*, pp. 314–327, Springer-Verlag, New York.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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 *LNCS*, pp. 278–294, Springer-Verlag, New York.

[10] 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 *LNCS*, pp. 189–203, Springer-Verlag, Berlin.

[11] 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.

[12] 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.

[13] 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.

[14] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, April 2012.

[15] 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.

[16] T. Denemark and J. Fridrich, "Steganography with multiple JPEG images of the same scene," *IEEE Transactions on Information Forensics and Security*, 2016, in preparation.

[17] 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.

[18] 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.