PROCEEDINGS

IS&T / SPIE
Electronic
Imaging
SCIENCE AND TECHNOLOGY

Media Watermarking, Security, and Forensics 2014

Adnan M. Alattar
Nasir D. Memon
Chad D. Heitzenrater
Editors

3–5 February 2014
San Francisco, California, United States

Sponsored by
IS&T—The Society for Imaging Science and Technology
SPIE

Volume 9028

IS&T
imaging.org

SPIE

PROCEEDINGS

IS&T / SPIE
Electronic
Imaging
SCIENCE AND TECHNOLOGY

Media Watermarking, Security, and Forensics 2014

Adnan M. Alattar
Nasir D. Memon
Chad D. Heitzenrater
Editors

3–5 February 2014
San Francisco, California, United States

Sponsored
IS&T—The Society for Imaging Science and Technology
SPIE

Published by
SPIE

Volume 9028

Proceedings of SPIE 0277-786X, V. 9028

The papers included in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. The papers published in these proceedings reflect the work and thoughts of the authors and are published herein as submitted. The publishers are not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," in *Media Watermarking, Security, and Forensics 2014*, edited by Adnan M. Alattar, Nasir D. Memon, Chad D. Heitzenrater, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 9028, Article CID Number (2014)

ISSN: 0277-786X
ISBN: 9780819499455

Copublished by
SPIE
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time) · Fax +1 360 647 1445
SPIE.org
and
IS&T—The Society for Imaging Science and Technology
7003 Kilworth Lane, Springfield, Virginia, 22151 USA
Telephone +1 703 642 9090 (Eastern Time) · Fax +1 703 642 9094
imaging.org

Printed in the United States of America.

Paper Numbering: Proceedings of SPIE follow an e-First publication model, with papers published first online and then in print and on CD-ROM. Papers are published as they are submitted and meet publication criteria. A unique, consistent, permanent citation identifier (CID) number is assigned to each article at the time of the first publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online, print, and electronic versions of the publication. SPIE uses a six-digit CID article numbering system in which:

- The first four digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc.

The CID Number appears on each page of the manuscript. The complete citation is used on the first page, and an abbreviated version on subsequent pages. Numbers in the index correspond to the last two digits of the six-digit CID Number.

Conference Committee

Symposium Chair

Sergio R. Goma, Qualcomm Inc. (United States)

Symposium Co-chair

Sheila S. Hemami, Northeastern University (United States)

Conference Chairs

Adnan M. Alattar, Digimarc Corporation (United States)
Nasir D. Memon, Polytechnic Institute of New York University (United States)
Chad D. Heitzenrater, Air Force Research Laboratory (United States)

Conference Program Committee

Mauro Barni, Università degli Studi di Siena (Italy)
Sebastiano Battiato, Università degli Studi di Catania (Italy)
Jeffrey A. Bloom, Sirius XM Satellite Radio (United States)
Scott A. Craver, Binghamton University (United States)
Edward J. Delp III, Purdue University (United States)
Jana Dittmann, Otto-von-Guericke-Universität Magdeburg (Germany)
Gwenaël Doërr, Technicolor Research and Development (France)
Tomas Filler, Digimarc Corporation (United States)
Jessica Fridrich, Binghamton University (United States)
Anthony T. S. Ho, University of Surrey (United Kingdom)
Jiwu Huang, Sun Yat-Sen University (China)
Ton Kalker, DTS, Inc. (United States)
Andrew D. Ker, University of Oxford (United Kingdom)
Alex Chichung Kot, Nanyang Technological University (Singapore)
Chang-Tsun Li, The University of Warwick (United Kingdom)
Pierre Moulin, University of Illinois at Urbana-Champaign (United States)
Regunathan Radhakrishnan, Pivotal Systems (United States)
Husrev Taha Sencar, TOBB University of Economics and Technology (Turkey)
Gaurav Sharma, University of Rochester (United States)
Yun Qing Shi, New Jersey Institute of Technology (United States)
Ashwin Swaminathan, Qualcomm Inc. (United States)
Claus Vielhauer, Fachhochschule Brandenburg (Germany)
Svyatoslav V. Voloshynovskiy, University of Geneva (Switzerland)
Chang Dong Yoo, KAIST (Korea, Republic of)

Introduction

It is our pleasure to bring you the papers presented at the 2014 Media Watermarking, Security, and Forensics Conference. This year's conference was a great success, as it maintained and strengthened its status as the premier conference in the field. It was held at the Hilton, in the popular tourist destination of downtown San Francisco. This convenient location let the attendees enjoy their San Francisco stay while participating in the outstanding technical exchange this forum is well known for.

For three full days, researchers from all over the globe in the field of watermarking, steganography, and forensics presented state-of-the-art research results in a lecture style. Thirty-one top-quality, original research papers were presented in a well-attended single track. Many great papers were submitted, but unfortunately not all could be accommodated by the conference. Attendees from academic, industrial, and governmental establishments enjoyed the presentations of latest research results. After the presentations and during the coffee breaks, presenters enjoyed discussing their research approaches and results with colleagues. The valuable feedback and comments they received planted the seeds for more research to advance the state-of-the-art of watermarking, security, and forensics.

Attendees also enjoyed the strong industrial flavor of this conference. This year, three distinguished keynote speakers from industry and academia addressed recent technological developments in their field of expertise:

- Sunil Jain from Intel Corporation gave an excellent speech on digital wallet and mobile payment, which included a description of his "Dream Wallet."

- Markus Jakobsson from ZapFraud gave a wonderful speech titled, "Authenticate or Not." He exposed an array of mistaken beliefs relating to authentication and authenticity.

- Hany Farid from Dartmouth College gave a great talk titled, "Photo Forensics from Shadows and Shading." He presented a clever way to detect inconsistency in lighting from shading and shadows in an image.

All three speeches were informative, illuminating, and well-received by the audience.

In addition to the keynote speeches, representatives from the industry gave three demos of their innovative products and solutions of related technologies:

- Digimarc Corporation gave a video demo of their Digimarc Barcode technology, which is an outstanding application of watermarking. This demo was a recording of Digimarc breaking the Guinness World Record for fastest scanning and bagging of 50 items at the National Retail Forum in NYC.

- Konica Minolta demoed their self-verifiable paper documents and automatic content verification system. Verification is done using a 2D color barcode printed on the document. A compressed version of the information on the document is stored in this barcode.

- Hewlett Packard demoed their watermark-based Weekend GoGuide application that allows a participant to receive printed promotions on his or her Internet-connected printer. The user can view more information about this promotion by snapping a photo of it using his or her mobile device.

All three demos showed the great potential of the technology being developed by the attendees of the conference.

New this year, to acknowledge excellent contributions to the field, the program committee of the conference decided to offer a Best Paper award. The committee sought input from the session chairs of the conference, and in response they received the following nominations:

1. "Feature-based watermark localization in digital capture systems," by Vojtech Holub, Binghamton Univ. (United States); Tomas Filler, Digimarc, (United States).

2. "A Mishmash of Methods to Mitigate the Model Mess Mismatch," by Andrew Ker, Univ. of Oxford (United Kingdom); Tomas Pevny, Czech Technical Univ. in Prague (Czech Republic).

3. "Cover Estimation and Payload Location using Markov Random Fields" by Tu-Thach Quach, Sandia National Labs (United States).

4. "A framework for fast and secure packaging identification on mobile phones," by Maurits Diephuis, Svyatoslav V. Voloshynovskiy, Taras Holotyak, Nabil Stendardo, Univ. of Geneva (Switzerland); Bruno Keel, U-NICA Group (Switzerland).

5. "Content identification: binary content fingerprinting versus binary content encoding," by Sohrab Ferdowsi, Univ. of Geneva (Switzerland); Svyatoslav V. Voloshynovskiy, Univ. of Geneva (Switzerland); Dimche Kostadinov, Univ. of Geneva (Switzerland).

All nominated papers will be reviewed by the technical committee of the conference. The papers will be judged based on their originality, creativity, clarity, and potential impact on the field of watermarking, security, and forensics. The winner will be announced in the next year's call for papers or in a special event during next year's conference.

Finally, the conference chairs express appreciation for all the hard work and for the enthusiastic participation of the researchers, scientists, practitioners, industry and government representatives, keynote speakers, and the organization committee of the conference. The chairs also congratulate all of you on the success of this conference. Without your dedicated effort, little could have been achieved. Together, we will have another great conference next year. Thank you.

Adnan Alattar
Nasir Memon
Chad Heitzenrater

Session Chairs

1 Steganography
Jessica Fridrich, Binghamton University (United States)

2 Biometrics and Watermarking
Svyatoslav V. Voloshynovskiy, University of Geneva (Switzerland)

Keynote Session I
Tomas Filler, Digimarc Corporation (United States)

3 Watermarking
Gwenaël Doërr, Technicolor Research and Development (France)

4 Steganalysis
Andrew D. Ker, University of Oxford (United Kingdom)

Keynote Session II
Nasir D. Memon, Polytechnic Institute of New York University (United States)

5 Identification
Gaurav Sharma, University of Rochester (United States)

6 Authentication
Sebastiano Battiato, Università degli Studi di Catania (Italy)

Keynote Session III
Edward J. Delp III, Purdue University (United States)

7 Forensics
Mauro Barni, Università degli Studi di Siena (Italy)

Contents

SESSION 1 STEGANOGRAPHY

SESSION 2 BIOMETRICS AND WATERMARKING

SESSION 3 WATERMARKING

SESSION 4 STEGANALYSIS

VIDEO/DEMO II: KONICA MINOLTA LABORATORY

SESSION 5 IDENTIFICATION

SESSION 6 AUTHENTICATION

SESSION 7 FORENSICS

Challenging the Doctrines of JPEG Steganography

Vojtěch Holub and Jessica Fridrich

Department of ECE, SUNY Binghamton, NY, USA

ABSTRACT

The design of both steganography and steganalysis methods for digital images heavily relies on empirically justified principles. In steganography, the domain in which the embedding changes are executed is usually the preferred domain in which to measure the statistical impact of embedding (to construct the distortion function). Another principle almost exclusively used in steganalysis states that the most accurate detection is obtained when extracting the steganalysis features from the embedding domain. While a substantial body of prior art seems to support these two doctrines, this article challenges both principles when applied to the JPEG format. Through a series of targeted experiments on numerous older as well as current steganographic algorithms, we lay out arguments for why measuring the embedding distortion in the spatial domain can be highly beneficial for JPEG steganography. Moreover, as modern embedding algorithms avoid introducing easily detectable artifacts in the statistics of quantized DCT coefficients, we demonstrate that more accurate detection is obtained when constructing the steganalysis features in the spatial domain where the distortion function is minimized, challenging thus both established doctrines.

Keywords: Steganalysis, steganography, JPEG, rich models, distortion, security

1. INTRODUCTION

It is an obvious fact that if the sender executes the embedding changes uniformly pseudo-randomly across the cover image, a scheme that on average introduces the fewest number of embedding changes ought to be more secure than its competitors. This reasoning provided a bridge between the theory of covering codes and steganography[2,3,10] responsible for an avalanche of papers on matrix embedding and a suite of more secure steganographic algorithms, such as the F5 algorithm[24] and its improved version called nsF5.[9]

Measuring the embedding distortion by counting the embedding changes, however, fails to take into account the fact that modifications of quantized DCT coefficients from the same 8 × 8 block strongly interact and that the embedding changes may have different "costs" depending on the associated quantization step and the local image content. Moreover, DCT coefficients that are adjacent either in the frequency or spatial domain exhibit complex dependencies that are not well understood. While discernible objects and their orientation are easily identifiable in the spatial domain, it is harder to determine them by inspecting DCT coefficients. From this perspective, it appears that it might be advantageous to abandon the doctrine that requires measuring distortion in the embedding domain as it is more manageable to design distortion functions that correlate with statistical detectability in the spatial domain. This thesis seems to be in agreement with recent developments in steganography that we discuss below.

The authors of BCHopt[20] were the first to recognize that a good distortion measure needs to consider the effect of the quantization step associated with the modified coefficients. Barring some unimportant details, the distortion function was basically designed to minimize the embedding distortion w.r.t. the uncompressed cover image (the precover). The minimized quantity was the square of the product of the quantization step and the change in the DCT coefficient w.r.t. the precover. Such an embedding distortion, however, could equivalently be defined as an L_2 norm in the spatial domain due to Parseval equality because the DCT is orthonormal. The more recent Entropy Block Steganography (EBS)[23] improved significantly upon BCHopt using a similar distortion function by replacing the BCH codes with the much more powerful Syndrome–Trellis Codes (STCs).[6]

Viewing both algorithms from the perspective of the current state of the art, both BCHopt and EBS hinted at a trend to embed in JPEG images by minimizing an embedding distortion defined in the spatial domain. This

E-mail: {vholub1,fridrich}@binghamton.edu; `http://dde.binghamton.edu`

Media Watermarking, Security, and Forensics 2014, edited by Adnan M. Alattar, Nasir D. Memon,
Chad D. Heitzenrater, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9028, 902802
© 2014 SPIE-IS&T · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2035178

development culminated in the design of the recently proposed UNIWARD distortion function,[12,14] which provides a universal method for measuring the embedding distortion independently of where the embedding changes are executed. Schemes based on UNIWARD were shown to significantly outperform prior art for steganography in JPEG images (both with and without side information at the sender). In UNIWARD, the distortion is computed as a sum of relative changes of directional residuals obtained using a Daubechies 8-tap filter bank. As shown later in this paper using experiments with the JPEG rich model,[17] minimizing a spatial-domain-based UNIWARD seems to minimize the impact on the statistics of DCT coefficients as well. UNIWARD also naturally incorporates the effect of the quantization step that other schemes need to build in, usually in some ad hoc manner (see, e.g., NPQ[15] and its improved version[4]).

We now take a closer look at the opposite problem, which is the detection of steganography (steganalysis). A doctrine has been formulated in 2004 [7] claiming that the most accurate steganalysis will naturally be achieved in the embedding domain because this is where the embedding changes are lumped and isolated. This doctrine seemed to hold true for embedding algorithms available at that time. This was mostly due to the fact that the early JPEG-domain stego algorithms, e.g., Jsteg,[22] F5, and OutGuess,[19] introduced quite detectable artifacts into the distribution of DCT coefficients (both their first-order and higher-order statistics). Furthermore, this doctrine was engraved even deeper in the minds of researchers after the BOSS competition[1] when all successful participants used steganalysis features constructed in the spatial (embedding) domain.

The fact that features computed in other domains can be useful for steganalysis is not new and it appeared already in the first papers on feature-based blind steganalysis[5] as well as in [7] (the "blockiness" feature is defined in the spatial domain). For a long time it remained true, though, that features constructed in the embedding domain provided the most accurate steganalysis results. The authors of [16] proposed the so-called Cross-Domain Features (CDFs) to improve the attack on YASS.[21] This was not surprising as YASS embeds in a key-dependent domain and thus one cannot construct features in the embedding domain. With the development of rich image models for both the spatial (SRM)[8] and DCT (JRM)[17] domains it was shown in [17] that virtually all JPEG-domain algorithms can be detected more reliably with the union of the SRM and JRM called JSRM. The size of the improvement was dependent on the algorithm and was generally larger for those embedding algorithms that were harder to detect, which were exactly those that somehow utilized the spatial domain representation in computing their distortion function. Using selected experiments, we demonstrate in this paper that the current most advanced JPEG-domain stego algorithms are better detected in the domain in which the distortion is minimized rather than the domain where the embedding changes are executed.

In the next section, we introduce the common core of all experiments and briefly describe the steganalysis features and steganographic algorithms utilized in experiments. Section 3 contains the results of all experiments and their interpretation that challenges both doctrines discussed above. Section 4 contains a brief summary.

Even though parts of this work have appeared in a scattered form in other papers, the authors believe that clearly spelling out the main message (the challenge of both doctrines) in a stand-alone paper supported with dedicated experiments is valuable for the steganographic community.

2. PRELIMINARIES

2.1 Common core of all experiments

All experiments in this paper were run on the standard database BOSSbase 1.01.[1] This source contains $10,000$ images acquired by seven digital cameras in the RAW format (CR2 or DNG) and subsequently processed by converting them to 8-bit grayscale, resizing, and central-cropping to 512×512 pixels. The script for this processing is also available from the BOSS competition web site. For JPEG experiments, the database is JPEG-compressed using the Matlab's 'dct2' command with standard quantization tables corresponding to quality factors 75 and 95.

The classifiers we use are all instances of the ensemble proposed in [18] and are available from `http://dde.binghamton.edu/download/ensemble`. They employ Fisher linear discriminants as base learners trained on random subspaces of the feature space. The out-of-bag (OOB) estimate of the testing error, E_{OOB}, on bootstrap samples of the training set is used to automatically determine the random subspace dimensionality and the number of base learners as described in [18]. The OOB error is an unbiased estimate of the minimal total detection error under equal priors, $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}\left(P_{\mathrm{FA}} + P_{\mathrm{MD}}(P_{\mathrm{FA}})\right)$. The E_{OOB} is also used to report the detection performance. Finally, we note that a separate classifier was trained for each quality factor, embedding method, and payload.

2.2 Feature sets

The feature sets used in this paper are high dimensional rich models previously developed separately for each embedding domain. To the best knowledge of the authors, these feature sets currently provide the most accurate detection across many embedding algorithms in their corresponding domains, provided the cover source is known to the Warden.

The JPEG-domain rich model (JRM)[17] with a dimensionality of $22,510$ is formed by higher-order statistics of quantized DCT coefficients. The JRM features are also Cartesian-calibrated,[16] which means that the features computed from the image are supplemented with the same features computed by decompressing the image, cropping by 4 pixels, and recompressing using the same quantization table.

The spatial-domain rich model we use in this paper is the recently proposed Projection Spatial Rich Model (PSRM) with the quantization step equal to 3 (PSRMQ3)[13] of dimensionality $12,870$. The PSRM uses the same set of image residuals as the SRM but it represents them in a different manner. Instead of four-dimensional co-occurrences used in the SRM, the PSRM projects a set of randomly selected adjacent residual samples onto a random direction and uses the first-order statistics of the projections as features. The PSRM enjoys a much improved detection-performance vs. dimensionality trade off – the PSRM can achieve the same detection accuracy with a dimensionality lower by an order of magnitude. Moreover, at the same dimensionality, the PSRM achieves a lower detection error. This improvement is generally larger for hard-to-detect highly content adaptive embedding schemes. This is primarily because, in order to control the feature dimensionality, the co-occurrences require a quite harsh truncation of the residuals and a low co-occurrence order. On the other hand, the projections can capture dependencies among a large number of adjacent residual samples and can work with non-truncated versions of the residuals.

Finally, we note that the acronym JPSRM will be used for the $35,380$-dimensional merger of JRM and PSRM.

2.3 Steganographic algorithms

In our experiments, we included both older algorithms and current state-of-the-art schemes. We also cover both algorithms that do not use any side information at the sender (no uncompressed precover available) as well as side-informed embedding schemes. The following is the list of all non side-informed algorithms:

- OutGuess as published in [19];
- Jsteg[22] with simulated optimal binary coding;
- nsF5[9] with simulated optimal binary coding;
- UED[11] with simulated optimal ternary coding;
- J-UNIWARD as published in [12] simulated at its payload–distortion bound.

The side-informed stego algorithms evaluated in this paper are:

- BCHopt as published in [20];
- NPQ as published in [15] simulated at its payload–distortion bound;
- Square cost simulated at its payload–distortion bound with the embedding cost of changing the ij-th DCT coefficient corresponding to a DCT mode k, l by ±1: $\rho_{ij}^{(kl)} = \left(q_{kl}(1 - 2|e_{ij}|)\right)^2$. Here, q_{kl} is the quantization step of the kl-th mode and e_{ij} is the quantization error when rounding the DCT coefficient obtained from the precover image during JPEG compression;

- SI-UNIWARD as published in [12] simulated at its payload–distortion bound.

With the exception of BCHopt, all side-informed embedding algorithms avoid making embedding changes to DCT coefficients with rounding error $e_{ij} = 1/2$ in DCT modes $(k, l) \in \{(0,0), (0,4), (4,0), (4,4)\}$ to avoid a singular behavior for small payloads that is especially apparent for large quality factors (see Section 5.3 in [12] for details).

We wish remark that the J-UNIWARD version published in [12] differs slightly from the newer version that will appear in [14]. The only difference is a different value of the stabilizing constant σ, which makes the newer version slightly more secure. The differences between both versions are, however, small, and the conclusions of this paper remain valid for both versions.

3. EXPERIMENTS

In this section, we interpret the results of experiments shown in Table 1. By doing so, we challenge the doctrines mentioned in the introduction. The table shows the E_{OOB} detection error obtained using the JRM, the spatial domain PSRMQ3, and the combined JPSRM on the steganographic algorithms listed in Section 2.3. The results are presented for two quality factors and one small and one large relative payload expressed in bits per non-zero AC DCT coefficient (bpnzAC). Since the coding in BCHopt does not allow embedding 0.4 bpnzAC in all images, we tested it for 0.3 bpnzAC.

Figure 1 displays the same results in a graphical form for the quality factor 75. In the figure, the algorithms are ordered by their statistical detectability obtained using the JPSRM. To give the reader a sense of the statistical significance of small changes in the E_{OOB}, we measured this error over ten runs of the ensemble classifier with different seeds for its random number generator that drives the selection of random subspaces as well as the bootstrapping for the training sets. The standard deviation of E_{OOB} was rather stable across the payloads, quality factors, as well as embedding algorithms, and it was always below 0.003. For better readability, we refrain from including this spread in the table.

When the JRM can detect a stego algorithm efficiently, one can say that the embedding disturbs important statistics of DCT coefficients. We view such algorithms as "faulty." Depending on the stego algorithm, the problem is either in the embedding operation or in the design of the distortion function that is supposed to measure the statistical detectability of embedding changes. Both the LSB replacement embedding operation of Jsteg and the operation of nsF5, which always decreases the absolute value of the DCT coefficient, predictably modify the first-order (and higher-order) statistics of coefficients. Such artifacts are understandably better detected by the JRM than the PSRM. The same is true for OutGuess, which turned out as the most detectable out of all tested algorithms. Even though it preserves the global histogram, it does so at the expense of introducing additional changes, and, in the end, disturbs the statistics of DCT coefficients even more. (Recall, that the JRM uses statistics of individual pairs of DCT modes, which are not necessarily preserved by OutGuess.)

While the ternary coded UED algorithm is markedly better than the older non side-informed algorithms, it is clearly outperformed by J-UNIWARD, which minimizes a distortion function defined in the spatial domain. This experimental fact challenges the first doctrine from Section 1 that claims that one should always minimize distortion defined in the embedding domain. The distortion function of J-UNIWARD seems to capture the impact on the statistics of DCT coefficients rather well. This finding should be taken "with a grain of salt" as it is entirely possible that better, more sophisticated distortion functions can be built in the DCT domain. The authors, however, believe that designing such functions will be rather challenging for the reasons mentioned in the introduction.

Two of the distortion-based side-informed steganographic schemes, BCHopt and NPQ, are also better detectable by the JRM than the PSRM. Their embedding operation is LSB matching, which introduces less strong artifacts in the statistics of coefficients than LSB replacement or the operation of nsF5. However, since all algorithms with the exception of nsF5, Jsteg, and OutGuess use LSB matching, the increased detectability of BCHopt and NPQ by JRM is most likely due to weaknesses in their distortion function, which does not capture the statistical dependencies among DCT coefficients well.

Payload	SI	0.1 bpnzAC			0.4 bpnzAC		
Features		JRM	PSRMQ3	JPSRM	JRM	PSRMQ3	JPSRM
Dimension		22,510	12,870	35,380	22,510	12,870	35,380
			Quality factor 75				
OutGuess		0.0010	0.0011	0.0005	0.0001	0.0003	0.0001
Jsteg		0.0578	0.1159	0.0372	0.0004	0.0007	0.0003
nsF5		0.2115	0.2609	0.1631	0.0036	0.0057	0.0008
UED ternary		0.3968	0.3369	0.3393	0.0488	0.0390	0.0202
J-UNIWARD		0.4632	0.4319	0.4350	0.2376	0.1294	0.1228
BCHopt	●	0.4122	0.4228	0.3941	0.0830*	0.1039*	0.0546*
NPQ	●	0.4139	0.4613	0.4076	0.0654	0.0760	0.0345
Square loss	●	0.4908	0.4880	0.4914	0.3656	0.3246	0.3246
SI-UNIWARD	●	0.5004	0.4952	0.4970	0.4470	0.3744	0.3755
			Quality factor 95				
OutGuess		0.0006	0.0015	0.0005	0.0001	0.0012	0.0002
Jsteg		0.0429	0.2033	0.0352	0.0001	0.0054	0.0003
nsF5		0.1354	0.3401	0.1220	0.0005	0.0252	0.0005
UED ternary		0.4750	0.4785	0.4727	0.2604	0.2759	0.2180
J-UNIWARD		0.4923	0.4943	0.4920	0.3951	0.3256	0.3246
BCHopt	●	0.3600	0.4715	0.3582	0.1172*	0.3491*	0.1144*
NPQ	●	0.4295	0.4950	0.4308	0.1471	0.3358	0.1342
Square loss	●	0.4556	0.4865	0.4554	0.3664	0.3952	0.3442
SI-UNIWARD	●	0.4654	0.4955	0.4672	0.4418	0.3909	0.3790

Table 1. Detection error E_{OOB} achieved using three different rich models for two JPEG quality factors and two payloads. The dot in the column labeled "SI" highlights those JPEG algorithms that use side information in the form of the uncompressed image. The asterisk highlights the fact that BCHopt was tested for payload 0.3 bpnzAC instead of 0.4 because its coding does not allow embedding payloads of this size in all images.

On the other hand, the most secure JPEG-domain algorithms, J-UNIWARD, and the side-informed Square Loss and SI-UNIWARD, are better detectable by the spatial-domain PSRM than by the JRM.[*] In fact, for the UNIWARD family the entire detection power seems to be coming from the PSRMQ3 as adding the JRM does not lead to any statistically significant improvement. This seems to point to two interesting facts. Reiterating and strengthening what has already been said about J-UNIWARD, since the distortion functions of the UNIWARD family are designed in the spatial domain, they naturally incorporate the effect of the quantization step and can better evaluate the impact of embedding on blockiness. What is more remarkable is that the schemes minimizing the impact in the spatial domain also seem to avoid introducing artifacts in the JPEG domain.

Moreover, with more sophisticated JPEG-domain algorithms that avoid disturbing the statistics of DCT coefficients it becomes more advantageous to steganalyze by representing the images in the domain in which the distortion is designed rather than in the embedding domain.

[*] For the small payload of 0.1 bpnzAC, they are essentially undetectable using any of the rich models.

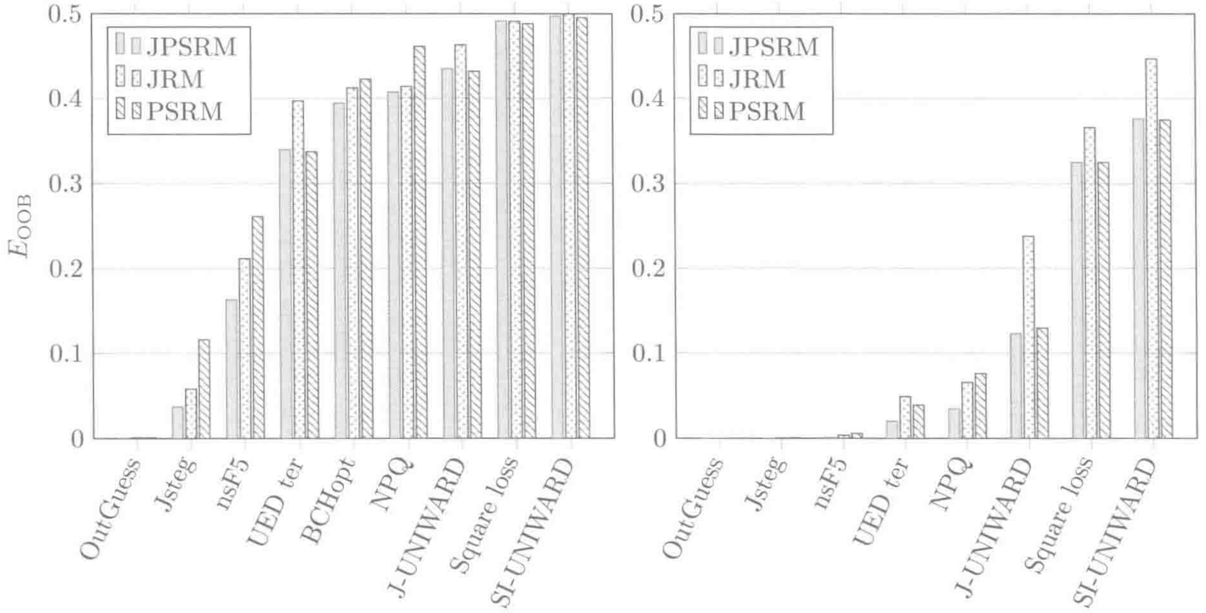

Figure 1. Detection error E_{OOB} using JPSRM, JRM, and PSRM on all tested steganographic algorithms for quality factor 75 with payloads 0.1 (left) and 0.4 (right) bits per non-zero AC coefficients. Note especially the cases when the spatial-domain features detect better than JPEG-domain features (when the right bar is smaller than the middle bar). Note that the merged JPSRM always provides the smallest detection error. This figure also nicely shows the progress made in JPEG steganography over the years.

4. CONCLUSION

Throughout the history, researchers have converged to a few empirical principles widely used when designing both steganography and steganalysis algorithms. The two most prominent doctrines concern the role of the embedding domain as the preferred domain in which to measure the impact of embedding as well as extract steganalysis features. In this paper, we provide experimental evidence that these doctrines may not be valid for embedding in JPEG images. This is mainly because the quantized DCT coefficients form 64 parallel channels that exhibit complex dependencies that are not easily quantified. On the contrary, in the spatial domain, elements that form typical objects, such as edges, segments, and textures, are easily identifiable, which allows for a simpler and more transparent design of distortion functions as well as extraction of good steganalysis features.

Experiments on older as well as modern steganographic algorithms for JPEG images point to several interesting findings:

1. Embedding algorithms that introduce easily identifiable artifacts in the statistics of DCT coefficients are better detected using features constructed in the embedding domain. This applies to older algorithms, such as OutGuess, nsF5, Jsteg, and Model-based steganography.
2. JPEG algorithms whose distortion function takes into account the impact of embedding in the spatial domain tend to exhibit higher security and avoid introducing artifacts that can be captured using the JPEG rich model.
3. Modern embedding algorithms that minimize the embedding impact computed in the spatial domain are generally better detected using the spatial rich model rather than the JPEG rich model.

These findings pose some intriguing open questions pertaining to both steganography design and detection. In particular, with modern and more secure steganographic algorithms, the domain of choice for steganalysis might shift from the embedding domain to the domain in which the distortion is minimized.