



HAL
open science

Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images.

Rémi Cogranne, Quentin Giboulot, Patrick Bas

► **To cite this version:**

Rémi Cogranne, Quentin Giboulot, Patrick Bas. Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images.. ACM Information Hiding and MultiMedia Security (IH&MMSec), Jun 2020, Denver, CO, United States. 10.1145/3369412.3395075 . hal-02542075

HAL Id: hal-02542075

<https://utt.hal.science/hal-02542075v1>

Submitted on 20 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images.

Accepted version, Final version to be published in ACM IH&MMSec proceedings, DOI: TBA.

This paper is copyrighted; personal and classroom use of this material is permitted.

Rémi Cogranne*
LM2S Lab. - ROSAS Dept., Troyes
University of Technology
Troyes, France
remi.cogranne@utt.fr

Quentin Giboulot
LM2S Lab. - ROSAS Dept., Troyes
University of Technology
Troyes, France
quentin.giboulot@utt.fr

Patrick Bas*
CRISAL Lab, CNRS, Ecole Centrale
de Lille, Univ. of Lille
Lille, France
patrick.bas@centralelille.fr

ABSTRACT

This short paper presents a novel method for steganography in JPEG-compressed images, extended the so-called MiPOD scheme based on minimizing the detection accuracy of the most-powerful test using a Gaussian model of independent DCT coefficients. This method is also applied to address the problem of embedding into color JPEG images. The main issue in such case is that color channels are not processed in the same way and, hence, a statistically based approach is expected to bring significant improvements when one needs to consider heterogeneous channels together. The results presented show that, on the one hand, the extension of MiPOD for JPEG domain, referred to as J-MiPOD, is very competitive as compared to current state-of-the-art embedding schemes. On the other hands, we also show that addressing the problem of embedding in JPEG color images is far from being straightforward and that future works are required to understand better how to deal with color channels in JPEG images.

CCS CONCEPTS

• **Mathematics of computing** → **Probabilistic reasoning algorithms**; • **Security and privacy**; • **Computing methodologies** → **Image representations**; **Image compression**;

KEYWORDS

Steganography ; Steganalysis ; JPEG compression ; Model-based embedding ; Color images

ACM Reference Format:

Rémi Cogranne, Quentin Giboulot, and Patrick Bas. 2020. Steganography by Minimizing Statistical Detectability: The cases of JPEG and Color Images. Accepted version, Final version to be published in ACM IH&MMSec proceedings, DOI: TBA. This paper is copyrighted; personal and classroom use of this material is permitted. . In *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security (IHMMSec '20)*,

* All the source codes and images datasets used to obtain the results presented will be made available by the dates of the conference on first author's webpage. This work has been funded in part by the French National Research Agency (ANR-18-ASTR-0009), ALASKA project: <https://alaska.utt.fr>, by the French ANR DEFALS program (ANR-16-DEFA-0003).

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

IHMMSec '20, June 22–24, 2020, Denver, CO, USA

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-7050-9/20/06...\$15.00

<https://doi.org/10.1145/3369412.3395075>

June 22–24, 2020, Denver, CO, USA. ACM, New York, NY, USA, 7 pages.
<https://doi.org/10.1145/3369412.3395075>

1 INTRODUCTION

Steganography is usually referred to as the techniques and methods that allow to hide information within an innocuous-like cover object. The resulting stego-object resembles, as much as possible, the original cover object. Therefore it can be sent over an unsecured communication channel that may be subject to wiretapping by an eavesdropper. Nowadays, steganography has been mostly developed for digital images because of their massive presence over the Internet, the universally adopted JPEG compression scheme and its relative simplicity to be modified. A widely used scenario to present steganography is the one of the prisoners' problem [20], in which the steganographers, referred to as Alice and Bob share digital images that may, or may not, contain hidden data. However, in this scenario, the communication channel is under surveillance of a warden named Wendy whose goal is to perform steganalysis, that is to get any information relative to possibly hidden data ; the most interesting knowledge being the very presence of hidden data.

Over the past two decades, steganography and steganalysis has been dramatically improved. On the one hand, steganography has mainly benefited from advances in coding methods that allow the same number of bits to be embedded with fewer modifications, and, more importantly, enable to associate each cover element (pixels for uncompressed images and coefficients for JPEG) with different costs of modification. Syndrome-Trellis Codes [7] are typical examples of such codes with the added benefit of closely reaching the efficiency of information-theoretical bounds.

As a result, steganography is essentially reduced to defining a function that assigns a cost to each and every element a given cover-image is composed of.

There have been several attempts to design such a "cost map" ; for spatial domain images, S-UNIWARD [14] as well as HILL [17] have been widely adopted among the community ; for JPEG-compressed images, UERD [12] and J-UNIWARD [14] are two representatives state-of-the-art approach to capture cost of modifying pixels.

On the other hand, steganalysis has been widely developed with increasingly large features sets, from 686 in 2010 [18] to 34, 671 in 2012 [8], which were able to detect hidden data more and more efficiently. Roughly speaking, most of those features sets aim at removing as much image content as possible and count adjacent pixels values to detect artefacts left by data hiding. More recently,

the Deep Learning methods have been adapted for applications in steganalysis, see for instance [2, 25]. In both cases, steganalysis is essentially a supervised machine learning method in which features are extracted from sets of cover and stego-images and fed to a learning algorithm whose goal is to find the best possible decision rule based on the provided labeled examples.

As noted several times, steganography and steganalysis have mostly been studied in academic contexts. For instance, steganalysis is often used to benchmark steganography assuming, by Kerckhoffs's principle, that the warden knows the embedding algorithm, the length of hidden data.

The present paper focuses on steganography in which academic studies differ from practical context. The most striking, example pointed out in 2013 in [15], in that most current works focus on uncompressed and grayscale images. More precisely, it was noted in [3] that, between 2016 and 2019, only "33% of the papers deal with JPEG steganography/steganalysis" and, most important, "84% deal with gray-level images (steganalysis in color images was studied for the first time in 2014 [11, 16])".

In addition, there is a general lack of scientifically-sound method for designing cost-maps in steganography. The most popular algorithms, UNIWARD [14], HILL [17] and UERD [12] are based on heuristic approaches whose parameters are often optimized to offer the best performance on a reference dataset [1].

Nevertheless, alternative approaches have been proposed based on a statistical model of pixels. The first attempt [9] was quickly followed by improvements of the estimation of the statistical model parameters [19, 24] which brought this approach up to state-of-the-art performances. Recently, methods based on a statistical model of the whole image processing pipeline have also been leveraged to design steganography that is either able to capture local correlations [10] or that mimics natural image noise [23].

The present paper aims at exploiting this approach based on a statistical model of cover images in order to design distortion functions that are relevant for JPEG colored images. As opposed to prior works that require a precise knowledge of the image processing pipeline [10, 23], this paper aims at providing practical solutions in the most general case in which the steganographer is only given a digital image already compressed (typically out-of-camera picture).

The contribution of the present paper is mainly two-fold:

- (1) It presents an extension of MiPOD to design statistically-founded distortion function for JPEG-compressed images;
- (2) It proposes a method to address the problem of steganography for colored JPEG images.

Numerical results contrast dramatically with prior works and show possible important gains obtained by the proposed methods.

The present paper is organized as follows. Section 2 summarizes the foundations of the MiPOD embedding scheme and present its extension for JPEG compressed images. The problem of steganography on color images is then stated in Section 3 which also details the proposed approaches. Section 4 describes the experimental setup and discusses the numerical results. Section 5 eventually concludes the present paper and sketches possible future works.

2 J-MIPOD: STATISTICALLY-BASED DISTORTION FUNCTION TO JPEG IMAGES

2.1 Notations and Reminders on MiPOD

Throughout this paper, we will use let x for images in spatial domain and letter u for DCT coefficients. We will use boldface letters \mathbf{x} for vectors and boldface capitals \mathbf{U} for matrices. Grayscale images will be denoted $\mathbf{X} = (x_{m,n})$, $m \in \{1, \dots, M\}$, $n \in \{1, \dots, N\}$ and color images will be represented with a superscript in bracket $\mathbf{U}^{(c)}$, $c \in \{Y, Cb, Cr\}$ that represents the index of color channel.

The methodology presents in the present paper extends the statistically-based design of distortion function from [19]. Therefore, let us recall briefly that the MiPOD [19] embedding method is based on a statistical model under which pixels are assumed to be statistically independent, each following a Gaussian distribution $x_{m,n} \sim \mathcal{N}(\mu_{m,n}, \sigma_{m,n}^2)$. Based on this statistical model of pixels, the embedding method essentially consists in finding the probability of embedding in each pixel $\beta_{m,n}$ while minimizing the power (true positive probability) of the most powerful likelihood ratio test (LRT) which is can be expressed as:

$$\varrho^2 = \sum_{m,n} \frac{\beta_{m,n}^2}{\sigma_{m,n}^4}. \quad (1)$$

Assuming that the steganographer can estimate the variance $\sigma_{m,n}^2$, minimizing the power of the most powerful LRT consists in setting the change rates β_n in order to minimize the so-called deflection coefficient ϱ (1) under the constraint the desire payload R can be embedded.

Solving such an optimization problem can easily be carried out, see details in [9, 19], using the method of Lagrange multipliers which states that the solution must satisfy:

$$\beta_{m,n}^2 \sigma_{m,n}^{-4} = \frac{1}{2\lambda} \ln \frac{1 - 2\beta_{m,n}}{\beta_{m,n}}, \quad (2)$$

$$R = \sum_{m,n} H(\beta_{m,n}), \quad (3)$$

where $\lambda > 0$ is the Lagrange multiplier and $H(\cdot)$ represents the ternary entropy .

Academic works usually simulate embedding using information-theoretical bounds in which case the previous description is enough. In practice, if one wishes to embed real data, embedding probabilities $\beta_{m,n}$ must be converted into costs $\rho_{m,n}$ from which it can use STC. This is described in [19] but not considered in the present paper in which embedding is simulated in all cases we studied.

2.2 Extending MiPOD for JPEG Images

The main strength that makes MiPOD [19] lies in the estimation of the pixels variances. In the present paper, we extend this method for JPEG compressed images by simply (1) using the linearity of Discrete Cosine Transform (DCT) and (2) exploiting the law of linear transformation of Gaussian multivariate random variables. First, let us recall that JPEG compression is mostly based on DCT, especially for grayscale images. Pixels values are gathered into blocks of 8×8 pixels, represented as vectors of 64 elements denoted \mathbf{x}_i , $i \in \{1, \dots, MN/64\}$, over which the DCT is applied to obtain

the DCT coefficients denoted \mathbf{u}_i :

$$\mathbf{u}_i = \tilde{\mathbf{D}}\mathbf{x}_i \quad (4)$$

where matrix $\tilde{\mathbf{D}}$ of size 64×64 is an orthonormal matrix whose vectors represent each mode of the DCT.

Since pixels in spatial domain are modeled as Gaussian random variables, the block of 64 pixels \mathbf{x}_i can be modeled as a multivariate normal distribution:

$$\mathbf{x}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (5)$$

Raw pixels can modeled as independ realizations and for simpliity, we will adopt this model in present paper which leads to consider that $\boldsymbol{\Sigma}$ is a diagonal matrix [22]. In practice, one should note that after developement, various image processing operations (demaicing, denoising, sharpening, etc. ...) make the pixels correlated with each other but this phenomenon is neglected in the paper work since it would make the assessment of most-powerfull test statistical performance hardly possible [10, 23].

Thanks to the law of linear transformation of Gaussian random vectors, the DCT coefficients \mathbf{u}_i can be modeled as:

$$\mathbf{u}_i \sim \mathcal{N}(\tilde{\mathbf{D}}\boldsymbol{\mu}, \tilde{\mathbf{D}}\boldsymbol{\Sigma}\tilde{\mathbf{D}}^\top). \quad (6)$$

The last step in JPEG compression is the quantization of the DCT coefficient which essentially consists of a division, by a quantization step that depends on the DCT mode, followed by rounding:

$$\bar{\mathbf{u}}_i = \text{round}(\mathbf{u}_i \oslash \mathbf{q}) \quad (7)$$

where \oslash represents the element wise division by the vector that contains quantization steps \mathbf{q} . Obviously, the distribution of $\bar{\mathbf{u}}_i$ is not straightforward to express in a close form due to the rounding operation.

In the present paper we extend the MiPOD approach by minimizing the detectability through the deflection coefficient ϱ , as defined in (1), over all DCT coefficients by computing their variance. For the sake of simplicity, we will assume that the rounding operation only adds independent quantization noise from which the variance of DCT coefficients can be simplified to:

$$\tilde{\mathbf{D}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{D}}^\top \oslash \mathbf{q}^2. \quad (8)$$

Note that in the present paper we will only consider the variance of DCT coefficients, *i.e.* the diagonal elements, first, for simplicity when minimizing the power of LRT, and second because off-diagonal elements are very small. In addition, one should to take into account the quantization noise in (8). Usually, “statistical-based” approach in steganalysis focus on the case of fine quantization limit. For JPEG compressed images we noted that it is important, to ensure efficiency of the method, neither to threshold the variance (as proposed in [19]) nor to add the factor 1/12 because for low QF this would lead to allow embedding in DCT coefficients that should not be modified (with very low variance).

It follows from Equations (1)-(8) that the extension of MiPOD to grayscale images compressed using JPEG format is rather straightforward since it only requires estimating the variance of pixels in spatial domain. Such an estimation has been one of the main contribution of [19].

In the present paper, the method proposed for extending detectability-based steganography for JPEG images is summarized by Algorithm 1. In brief, it essentially consists of (1) decompressing DCT

Algorithm 1 : J-MiPOD Algorithm for Grayscale Images

- 1: $\mathbf{X} = \text{iDCT}(\mathbf{U} \oslash \mathbf{q})$ \triangleright Apply inverse DCT to obtain pixels in spatial domain
 - 2: Estimate pixel residual variances $\sigma_{m,n}^2$ using the estimator described in [19, Sec. V] ;
 - 3: Over blocks of 8×8 computes ensuing variance of DCT coefficients as in Eq. (8) ;
 - 4: Numerically solve Eqs. (2) and (3) determine the change rates β_n , $n = 1, \dots, N$ and the Lagrange multiplier λ .
-

coefficients into spatial domain (2) estimating variance of pixels (3) applying linear transformation to compute variance of DCT coefficients (4) minimize power of most powerful LRT based on these variances of DCT coefficients.

Two important comments are important. While this extension seems straightforward, it has been presented to the best of our knowledge in a single prior-work only [6] which does not focus on such extension but mostly attempt to apply the method in the context of side-informed (SI) steganography. Interestingly, it is shown in [6] that SI-J-MiPOD, that can access the pre-cover, works very well but not J-MiPOD.

We also note that extending detectability-based steganography for JPEG images open new ways to tackle the problem of how to spread the payload among several different images. This is the main contribution which proposes to apply such method for the problem of spreading message among several color channels.

3 STRATEGIES FOR EMBEDDING INTO COLOR IMAGES

The problem of steganography in color images is somewhat similar to the one of batch steganography. Roughly speaking, it can be summarized as below: let us consider a set of I images within which it is aimed to hide B bits. The problem is to find how to determine the number of bits to be hidden in each image b_1, \dots, b_I . Of course it is desirable to remain as undetectable as possible. However, there is not known general methodology to address such a problem for heterogeneous color channel. In addition, in the present paper we wish to avoid tailoring to a specific detection technique and or a specific dataset.

In the present paper we study a specific aspect of batch steganography to address the problem of spreading payload among color channels that can be considered as three different images.

This problem of batch steganography has already been studied for uncompressed images [4]. The application for color channels seems obvious because uncompressed color images are composed of three color channels (RGB for Red, Green and Blue) that share very similar properties. Color channels vary in terms of local correlation (mostly due to the demosaicing) yet they are strongly correlated and exhibit very similar properties.

However, when it comes to JPEG compressed images, colors are converted from RGB to YCbCr luminance/chrominance color space. By design, information is concentrated into the luminance channel (Y) which, hence, exhibits a much larger variability than the chrominance channels (Cb and Cr). In addition those channels are treated

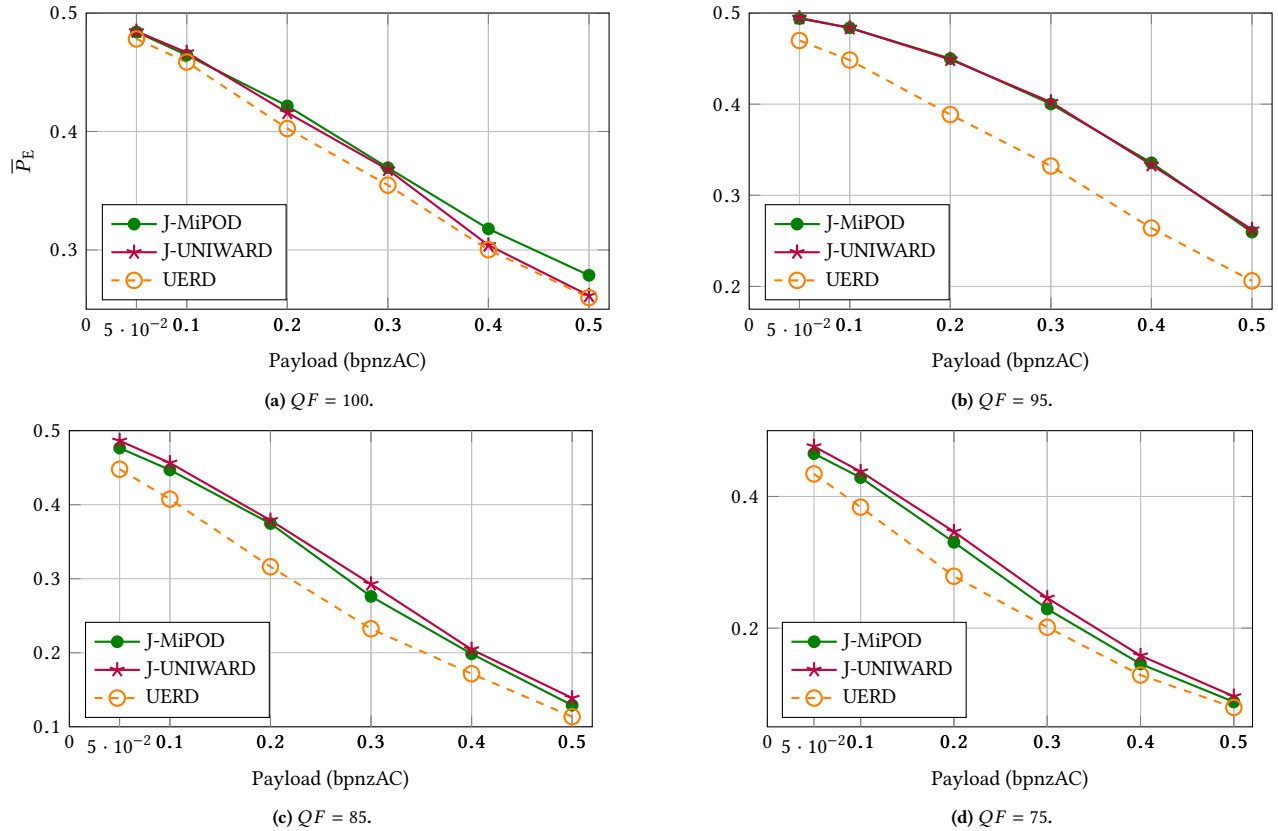


Figure 1: Evolution of P_E as a function of the embedding payload for BOSSbase grayscale images

differently: the quantization step applied onto the coefficients resulting from Discrete Cosine Transform (DCT) is different and the chrominance may be subjected to sub-sampling. In this context, it is easily understandable that methods for batch steganography over uncompressed images are not relevant anymore. In addition, as already discussed, cost-maps in JPEG has always been designed for grayscale images, hence without considering peculiarities of color channels. Most of steganographic schemes were also designed in a heuristic manner, which prevents taking into account the heterogeneity between the different color channels.

Breaking with these approaches, the present paper proposes a statistically-founded method for embedding in JPEG compressed images that takes into account both the quantization and the variance of coefficients. Therefore, there is no reason that this approach could not be applied for color images.

To this end, it is simply proposed to estimate the variance of all DCT coefficients (in all three color channels) as proposed in 2 before embedding a given message length while jointly minimizing the detectability over the color channels.

In order to assess the relevance of the proposed methodology, it is proposed to evaluate also the efficiency of several alternatives approach. For all of them, we extracted the cost maps for each color channel independently denoted $\rho^{(c)}$ with c the index of color channel $c \in \{Y, Cb, Cr\}$, representing luminance Y and chrominance channels Cb and Cr respectively. Using the cost-maps for the three

different color channels, we proposed the four following strategies for payload spreading:

- “Color channels Fixed repartition” (CCFR), referred to as “Arbitrary repartition of the payload between the 3 channels” in [21], consists in setting a fraction γ of the payload R in bits per non-zero AC coefficient (bpnzAC) to be embedded chrominance channels; The payloads in color channels (in bpnzAC) are thus determined by setting:

$$R_Y = \delta(1 - \gamma)R; R_{Cb} = R_{Cr} = \delta\gamma R \quad (9)$$

where δ is a scaling factor that enforces the payload constraint.

- Concatenated, or CCM for “Color Channels Merging” Strategy, consists in minimizing the additive distortion

$$\operatorname{argmin}_{\beta} D = \sum_{c \in \{Y, Cb, Cr\}} \sum_{m,n} \beta_{m,n}^{(c)} \rho_{m,n}^{(c)}. \quad (10)$$

- Detectability Limited Sender (DeLS) uses a Newton-Raphson method to equalize the detectability, measured as the deflection coefficients, for each channel:

$$q^{(c)} = \sum_{m,n} \frac{\beta_{m,n}^{(c)2}}{\sigma_{m,n}^{(c)4}}, \quad (11)$$

| | J-MiPOD | J-UNIWARD | UERD |
|---------------------------|---------|-----------|-------|
| Grayscale images | 6.38 | 11.27 | 1.43 |
| Color CCFR $\gamma = 0.2$ | 9.01 | 9.82 | 1.22 |
| Color CCM | 9.05 | 9.86 | 1.10 |
| Color DeLS | 12.98 | 12.69 | 10.09 |
| Color DiLS | 10.57 | 12.93 | 1.71 |

Table 1: Comparison of average embedding time for the three different embedding schemes and different use cases. Note that DiLS requires the (time-consuming) estimation of variance $\sigma_{m,n}^{(c)2}$ as proposed in [19].

where the probabilities of embedding $\beta_{m,n}^{(c)}$ are computed by adjusting the payload in each channel. Note that the variances are estimated using the proposed J-MiPOD while embedding probabilities are computed for any algorithm using the associated cost map.

- Distortion Limited Sender (DiLS), uses a similar Newton-Raphson method for equalizing the distortion for each channel:

$$d^{(c)} = \sum_{m,n} \beta_{m,n}^{(c)} \rho_{m,n}^{(c)}. \quad (12)$$

The last three methods are inspired from [4]. The first method is directly inspired by [21] which also adopts the Color Channel Merging strategy.

4 NUMERICAL RESULTS

The results presented in this paper have been obtained using the 10 000 images of size 512×512 from BOSS base [1]. We have slightly modified the original development in order to store images into color (original BOSS base is made of grayscale images) and compressed those images in JPEG using `convert` command from `imagemagick`. Two steganographic schemes, among the state of the art, have been compared with the method proposed in the present paper, referred to as J-MiPOD; those are J-UNIWARD [14], UERD [12]. All the detection results were obtained using the concatenation of DCTR features [13] for each color channel and with the low-complexity linear classifier [5] which uses 5 000 images for training (as well as validation) and 5 000 images for testing.

First of all, Figure 1 presents a comparison in terms of security, or more precisely, in terms of detection accuracy measured as the minimal total classification error rate under equal priors P_E . Those results present a comprehensive comparison with the payload ranging from 0.05 to 0.5 and four different JPEG quality factors. As opposed to what was claimed in [6], those results show that J-MiPOD is very competitive, performing almost as well as J-UNIWARD in a vast majority of cases. In addition, one should note that the method behind MiPOD and J-MiPOD was designed to be efficient for grayscale uncompressed images. In the present paper the methodology is extended with the very same variance estimator applied over chrominance channel while applying simple transformation to obtained variance of DCT coefficients. In other words, it is very likely that there is room for improving the estimation of variance of DCT coefficients used in the present paper.

In addition, Table 1, which compares the average embedding time

in grayscale images, shows that J-MiPOD is almost twice as fast as J-UNIWARD, yet more than four time slower than UERD.

Regarding color images, the Figure 2 shows a comparison of the four strategy studied for all the three same embedding schemes. Those results are presented for different quality factors ($QF = \{75, 95, 100\}$) and for different payloads ($R = \{0.2, 0.4\}$).

Those results are more difficult to analyze generally because it highly depends on quality factors. Obviously, for $QF = 75$ J-MiPOD associated with DeLS or DiLS strategies performs much better than all other alternatives.

For $QF = 100$ J-MiPOD also achieves a very high practical security which exceeds the one of J-UNIWARD while roughly equalizing UERD's. More importantly one can note that for those two quality factors, the Detectability Limited Sender strategy for spreading the payload across color channels is the most efficient for both J-MiPOD and UERD and works very efficiently for J-UNIWARD.

However, the situation is quite different for $QF = 95$. Indeed, J-UNIWARD performs significantly better for this payload. In addition, neither DeLS nor DiLS work very well for J-MiPOD while one would expect this statistically-based embedding scheme to work well with detectability-based spreading strategies. For color images, the different in terms of computational time between J-UNIWARD and J-MiPOD becomes smaller. This especially holds true for DeLS which requires computing the variance of each pixel $\sigma_{m,n}^{(c)2}$, or the Fisher-Information, using the method proposed in [19]; this step represents an important overhead cost, especially for UERD.

We tried to understand those results better by analyzing the detection accuracy for each color channels independently. Such results are shown in Figure 3 and clearly shows that some questions remain to be answered. The Figures 3a-3b show what can be observed in a vast majority of the cases: the detection accuracy is mostly given by the channel(s) in which P_E is the lowest. Consequently, one should find the right payload distribution in order not to observe one of the channels being easily detectable.

However, what can be observed in Figures 3c-3d is much more counterintuitive. Indeed in both cases one can observe that, in a few cases, jointly analyzing the color channels brings a significant improvement in detection accuracy: this suggests that some features must be analyzed together. On the opposite, some results show the exact opposite: in some cases the detection accuracy obtained with only one channel is way higher than when merging features from all channels. Even more surprising, those results are from the same QF and the same spreading strategy, only the overall payload changes.

Such phenomenon clearly highlights two things. First, embedding across the three-color channels of JPEG images is far from being straightforward. Depending on the embedding scheme, the payload and the quality factor, a method for spreading the payload may perform well or not. Second, a lot remain to be done along that direction, not only to design a method that would work well in general for color JPEG images, but also to get understand how detectability changes for those channels.

5 CONCLUSION AND FUTURE WORKS

This short paper addresses two problems; the first is related to extension of MiPOD embedding scheme for JPEG images and the second

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | 0.269 | <u>0.301</u> | 0.245 |
| CCM Strategy | 0.265 | 0.297 | 0.230 |
| DeLS Strategy | 0.321 | 0.296 | <u>0.250</u> |
| DiLS Strategy | 0.292 | 0.220 | 0.185 |

(a) $QF = 75$, Payload $R = 0.2$.

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | 0.086 | <u>0.120</u> | 0.092 |
| CCM Strategy | 0.084 | 0.108 | 0.085 |
| DeLS Strategy | 0.218 | 0.108 | <u>0.096</u> |
| DiLS Strategy | 0.180 | 0.047 | 0.051 |

(b) $QF = 75$, Payload $R = 0.4$.

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | <u>0.397</u> | 0.413 | 0.349 |
| CCM Strategy | 0.393 | 0.416 | 0.325 |
| DeLS Strategy | 0.383 | 0.415 | <u>0.359</u> |
| DiLS Strategy | 0.380 | 0.373 | 0.300 |

(c) $QF = 95$, Payload $R = 0.2$.

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | <u>0.205</u> | 0.234 | 0.170 |
| CCM Strategy | 0.197 | 0.240 | 0.140 |
| DeLS Strategy | 0.176 | 0.235 | <u>0.183</u> |
| DiLS Strategy | 0.170 | 0.157 | 0.163 |

(d) $QF = 95$, Payload $R = 0.4$.

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | 0.317 | 0.294 | 0.311 |
| CCM Strategy | 0.330 | <u>0.324</u> | 0.269 |
| DeLS Strategy | 0.348 | 0.315 | <u>0.330</u> |
| DiLS Strategy | 0.322 | 0.274 | 0.330 |

(e) $QF = 100$, Payload $R = 0.2$.

| Spreading strategy | J-MiPOD | J-UNIWARD | UERD |
|---------------------|--------------|--------------|--------------|
| CCFR $\gamma = 0.2$ | 0.180 | 0.157 | 0.171 |
| CCM Strategy | 0.187 | <u>0.179</u> | 0.152 |
| DeLS Strategy | <u>0.199</u> | 0.168 | 0.205 |
| DiLS Strategy | 0.167 | 0.142 | 0.198 |

(f) $QF = 100$, Payload $R = 0.4$.

Figure 2: Comparison of the security of the different method for spreading payload in color images. The security is measured here of the detection accuracy P_E for three embedding schemes, three JPEG quality factors and two payloads.

leverage this novel methodology for embedding into color images. On the one hand, it has been shown that one can successfully extend the method of embedding that seeks to minimize the power of most powerful LRT in JPEG domain using the same Gaussian independent model. On the opposite, when it comes to addressing the problem of spreading the payload across color channels of JPEG images, with the same method or with practical strategies, the results are much less impressive. On the one hand, J-MiPOD and Detectability Limited Sender strategy overall seems to perform very well. On the other hand, the results largely depend on the JPEG quality factor.

This short paper is a first step into the practical direction of embedding into JPEG color images. It opens the door to several questions and improvements that we will be addressing future works. First and most straightforward steps include improving J-MiPOD, in terms of both security and computational complexity, as well as assessing its performance in different cases, especially on the ALASKA dataset [3]. Second and most interesting steps are related to embedding in color images. We wish to understand the relationship between channels and overall detectability. We also would like to verify that the deflection coefficient is indeed related to the detectability measured with practical steganalysis. Eventually, we would like to extend the present work to the more general problem of spreading the payload over several color images (possibly compressed with different quality factors).

REFERENCES

- [1] P. Bas, T. Filler, and T. Pevný. 2011. Break Our Steganographic System — the ins and outs of organizing BOSS. In *13th International Workshop. LNCS vol.6958*, Springer-Verlag, New York, Prague, Czech Republic, 59–70.
- [2] M. Boroumand, M. Chen, and J. Fridrich. 2018. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security* 14, 5 (2018), 1181–1193.

| Stego scheme | $Y + Cb + Cr$ | Y | Cb | Cr |
|--------------|---------------|--------|--------|--------|
| JMIPOD | 0.2181 | 0.2206 | 0.4820 | 0.4820 |
| J-UNIWARD | 0.1076 | 0.0959 | 0.4856 | 0.4864 |
| UERD | 0.0955 | 0.1072 | 0.2878 | 0.2947 |

(a) $QF = 75$, Payload $R = 0.4$, DeLS.

| Stego scheme | $Y + Cb + Cr$ | Y | Cb | Cr |
|--------------|---------------|--------|--------|--------|
| JMIPOD | 0.1795 | 0.1831 | 0.4901 | 0.4897 |
| J-UNIWARD | 0.0474 | 0.1967 | 0.0879 | 0.0735 |
| UERD | 0.0508 | 0.1320 | 0.1101 | 0.1352 |

(b) $QF = 75$, Payload $R = 0.4$, DiLS.

| Stego scheme | $Y + Cb + Cr$ | Y | Cb | Cr |
|--------------|---------------|--------|--------|--------|
| JMIPOD | 0.3952 | 0.2918 | 0.4964 | 0.4960 |
| J-UNIWARD | 0.4136 | 0.3680 | 0.2736 | 0.2566 |
| UERD | 0.3488 | 0.2899 | 0.2413 | 0.2727 |

(c) $QF = 95$, Payload $R = 0.2$, CCFR $\gamma = 0.2$.

| Stego scheme | $Y + Cb + Cr$ | Y | Cb | Cr |
|--------------|---------------|--------|--------|--------|
| JMIPOD | 0.2046 | 0.1710 | 0.4352 | 0.4218 |
| J-UNIWARD | 0.2340 | 0.3563 | 0.2006 | 0.1941 |
| UERD | 0.1695 | 0.2217 | 0.2456 | 0.2959 |

(d) $QF = 95$, Payload $R = 0.4$, CCFR $\gamma = 0.2$.

Figure 3: Comparison of the detection accuracy P_E as obtained when analyzing each color channel separately and merging the features altogether.

- [3] R. Cogranne, Q. Giboulot, and P. Bas. 2019. The ALASKA Steganalysis Challenge: A First Step Towards Steganalysis. In *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security* (Paris, France) ACM, 125–137.
- [4] R. Cogranne, V. Sedighi, and J. Fridrich. 2017. Practical strategies for content-adaptive batch steganography and pooled steganalysis. In *2017 IEEE International*

- Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2122–2126.
- [5] R. Cogranne, V. Sedighi, J. Fridrich, and Tomáš Pevný. 2015. Is Ensemble Classifier Needed for Steganalysis in High-Dimensional Feature Spaces?. In *Information Forensics and Security (WIFS), IEEE 7th International Workshop on (Italy)*.
- [6] T. Denmark and J. Fridrich. 2017. Model based steganography with precover. *Electronic Imaging* 2017, 7 (2017), 56–66.
- [7] T. Filler, J. Judas, and J. Fridrich. 2011. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes. *Information Forensics and Security, IEEE Transactions on* 6, 3 (2011), 920–935.
- [8] J. Fridrich and J. Kodovský. 2012. Rich Models for Steganalysis of Digital Images. *Information Forensics and Security, IEEE Transactions on* 7, 3 (june 2012), 868–882. <https://doi.org/10.1109/TIFS.2012.2190402>
- [9] J. Fridrich and J. Kodovský. 2013. Multivariate Gaussian model for designing additive distortion for steganography. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2949–2953.
- [10] Q. Giboulot, R. Cogranne, and P. Bas. 2020. JPEG Steganography with side Information from the Processing Pipeline. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, IEEE, Barcelone, Spain.
- [11] M. Goljan and J. Fridrich. 2015. CFA-aware features for steganalysis of color images. In *Media Watermarking, Security, and Forensics 2015*, Proc. SPIE, Vol. 9409.
- [12] L. Guo, J. Ni, W. Su, C. Tang, and Y.-Q. Shi. 2015. Using statistical image model for JPEG steganography: uniform embedding revisited. *IEEE Transactions on Information Forensics and Security* 10, 12 (2015), 2669–2680.
- [13] V. Holub and J. Fridrich. 2015. Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT. *Information Forensics and Security, IEEE Transactions on* 10, 2 (Feb 2015), 219–228. <https://doi.org/10.1109/TIFS.2014.2364918>
- [14] V. Holub, J. Fridrich, and T. Denmark. 2014. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security* 2014, 1 (2014), 1–13. <https://doi.org/10.1186/1687-417X-2014-1>
- [15] A. D. Ker, & al. 2013. Moving steganography and steganalysis from the laboratory into the real world. In *Proceedings of the first ACM workshop on Information hiding and multimedia security* (Montpellier, France). ACM, 45–58.
- [16] M. Kirchner and R. Böhme. 2014. "Steganalysis in Technicolor" Boosting WS detection of stego images from CFA-interpolated covers. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3982–3986.
- [17] B. Li, M. Wang, J. Huang, and X. Li. 2014. A new cost function for spatial image steganography. In *International Conference on Image Processing*. IEEE, 4206–4210.
- [18] T. Pevný, P. Bas, and J. Fridrich. 2010. Steganalysis by Subtractive Pixel Adjacency Matrix. *IEEE Trans. Inform. Forensics and Security* 5, 2 (2010), 215–224.
- [19] V. Sedighi, R. Cogranne, and J. Fridrich. 2016. Content-Adaptive Steganography by Minimizing Statistical Detectability. *IEEE Transactions on Information Forensics and Security* 11, 2 (Feb 2016), 221–234. <https://doi.org/10.1109/TIFS.2015.2486744>
- [20] G. Simmons. 1983. The prisoners problem and the subliminal channel. *CRYPTO* (1983), 51–67.
- [21] T. Taburet, Louis Filstroff, Patrick Bas, and Wadih Sawaya. 2018. An empirical study of steganography and steganalysis of color images in the JPEG domain. In *International Workshop on Digital Watermarking*. Springer, 290–303.
- [22] Than Han Thai, Rémi Cogranne, and Florent Retraint. 2014. Camera Model Identification Based on the Heteroscedastic Noise Model. *Image Processing, IEEE Transactions on* 23, 1 (2014), 250–263. <https://doi.org/10.1109/TIP.2013.2290596>
- [23] T. Taburet, P. Bas, W. Sawaya, and J. Fridrich. 2020. Natural Steganography in JPEG Domain with a Linear Development Pipeline. *arXiv preprint arXiv:2001.02653* (2020).
- [24] S. Vahid, J. Fridrich, and R. Cogranne. 2015. Content-Adaptive Pentary Steganography Using the Multivariate Generalized Gaussian Cover Model. *Media Watermarking, Security, and Forensics 2015*, Proc. SPIE 9409.
- [25] M. Yedroudj, F. Comby, and M. Chaumont. 2018. Yedroudj-net: An efficient CNN for spatial steganalysis. In *Intl Conference on Acoustics, Speech and Signal Processing*. IEEE.