Statistical Correlation as a Forensic Feature to Mitigate the Cover-Source Mismatch

Antoine Mallet, Patrick Bas, Rémi Cogranne

▶ **To cite this version:**

Antoine Mallet, Patrick Bas, Rémi Cogranne. Statistical Correlation as a Forensic Feature to Mitigate the Cover-Source Mismatch. 12th ACM Workshop on Information Hiding and Multimedia Security (ACM IH&MMSEC'24), Jun 2024, Baiona, Spain. 10.1145/3658664.3659638 . hal-04571878

HAL Id: hal-04571878
https://hal.science/hal-04571878

Submitted on 9 May 2024

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.

Statistical Correlation as a Forensic Feature to Mitigate the Cover-Source Mismatch

Antoine Mallet*
antoine.mallet@utt.fr
LIST3N - Univ. of Technology of Troyes
Troyes, Grand Est, France

Patrick Bas
Patrick.Bas@centralelille.fr
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL
Villeneuve d'Ascq, Hauts de France
France

Remi Cogranne
remi.cogranne@utt.fr
Univ. of Technology of Troyes
Troyes, Grand Est, France

ABSTRACT

The present paper deals with the cover-source mismatch (CSM) problem in operational steganalysis. It first investigates the distribution of the noise in natural images, and shows how this property can be used to build a fingerprint of the cover-source, to address the issue of source identification from a single image. In particular, fingerprints from different noise extraction techniques are studied. Results show that these fingerprints can be complementary. The method proposed in the present paper aggregates them in a unique forensic feature to build a more accurate source identification algorithm than when using steganalysis features, such as the discrete cosine transform residual (DCTR). Last, the paper exploits the proposed forensic tool to mitigate CSM via "atomistic steganalysis". Used together with steganalysis methods, experimental results highlight the superiority of our approach, as compared to other atomistic mitigation strategies. The relevancy of these results is further studied on out-of-camera images coming from Flickr and the ALASKA dataset. We show that for some devices, our approach gives results superior to the omniscient scenario.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; • **Applied computing** → **Investigation techniques**; • **Security and privacy**;

KEYWORDS

forensics, steganalysis, cover-source mismatch, image processing pipeline, fingerprint

ACM Reference Format:
Antoine Mallet, Patrick Bas, and Remi Cogranne. 2024. Statistical Correlation as a Forensic Feature to Mitigate the Cover-Source Mismatch. In *Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security (IHMMSec '24), June 24–26, 2024, Baiona, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3658664.3659638

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
IH&MMSEC'24, June 24–26, 2024, Baiona, Spain
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0637-0/24/06
https://doi.org/10.1145/3658664.3659638

1 INTRODUCTION

Steganography is the art of hidden communication. It uses innocuous media, such as videos or text files as cover objects to embed a secret communication in. Steganalysis, on the opposite, tries to detect cover from stego objects. It designs detectors, most commonly Machine Learning (ML) models, which use a training set, containing cover and stego objects, to confront a testing set. In the most optimistic scenario, both training and testing samples come from the same origin, called a cover-source; it also assumes that the same steganographic scheme was used to generate the stego objects in both sets. This scenario might seem naive, but it can be considered a conservative interpretation of Kerckoff's principle, summarized under Shannon's maxim: "the enemy knows the system" [23].

In a more realistic approach, however, it is very difficult to know the cover-source of the testing set or, even worse, from a single object to inspect. A steganalysis detector can still be trained, but it would most likely be carried out over a cover-source that differs from the one used to generate the test set. This mismatch between the two distributions, called the cover-source mismatch (CSM) in operational steganalysis, might lead the steganalyst to dramatically low accuracy on the test set.

Many attempts at mitigating this drop of accuracy have been proposed ; let us recall the three major ones (although there exists more [16]):

- the *holistic* strategy aims at building the decision rule with the best possible generalization ability; it relies on a training set containing cover-sources as diverse as possible. Designing a training set was recently addressed [1]. Otherwise, holistic steganalysis trades accuracy on specific sources for its generalization ability [28].
- the *atomistic* strategy introduces a cover-source identification step that suppresses the CSM when performing steganalysis; This approach performs really well when dealing with CSM with a fixed number of cover sources [12, 26].
- the *domain adaptation* framework tries to learn an invariant representation to convert unknown cover-sources to. It is relevant, contrary to the atomistic approach, when dealing with unseen cover-sources. This approach is promising but can suffer from the double impact of steganography and CSM on current state-of-the-art features [25].

The holistic and domain adaptation ideas try to cope with CSM while performing steganalysis. On the other hand, the purpose of the atomistic approach is to eradicate CSM preemptively. As a result, it has been shown that it can actually perform just as well as the

clairvoyant scenario. To do so, however, it assumes the availability of a cover-source identification tool and knowledge of the existing cover-sources.

The current state of the art defines the CSM as the discrepancy between the distributions of training and testing samples, but the term is also used to designate the drop in accuracy of the detector when facing the CSM. We evaluate that using the same term for both the cause and the consequence is confusing. Rather, we suggest talking about *CSM* for the former and *CSM problem* (in steganalysis) for the latter.

Identifying the potential causes and the consequences helps us excavate the difference between the two problems of the steganalyst. We suggest that each problem should be dealt with separately. In this scenario, the atomistic approach appears to be the most appropriate way to free steganalysis from the CSM problem.

In section 2, we summarize the existing literature on source identification and atomistic steganalysis. In section 3 we give a detailed explanation of the rationale and methodology behind the proposed Correlation Feature ($C_{\mathcal{F}}$). In order to be comprehensive, we specifically highlight its strong & weak points. Section 4 details the experimental setup used in this paper, explaining the choices in designing our training and testing sets, as well as our choices of model. Results are presented and discussed in section 5. Section 6 concludes this paper and draws future work ideas.

2 RELATED WORK

The atomistic scheme is almost as old as the CSM problem; early studies reporting drops of accuracies when using different datasets in training and testing models already mentioned training one detector per cover-source as well as a multiclassifier to detect the cover source (or the embedding scheme) [20].

The atomistic mitigation strategy relies on a two-step process, as shown in Fig. 3. The idea is to get rid of the CSM before performing steganalysis. To do that, it relies on a multiclassifier, trained to recognize the cover-source of tested images, followed by a set of steganalysis detectors: one for each of the considered cover sources. This source identification step is crucial as, if reliable, it allows steganalysis to perform as well as in a scenario without CSM, called the clairvoyant scenario.

Over the years, multiple atomistic approaches have been proposed. To detect the JPEG quality factor, [2] suggested using a tool "out-of-the-shelf" from the forensic literature. Others suggested unsupervised clustering [11, 19]. Steganalysis features are also popular in building the source identification step, e.g. cc-PEV [14], cc-JRM [3], DCTR [7]. Indeed, while tailored for detecting steganography, they are all highly sensitive to the type of cover-source. But despite their competitive results, steganalysis features bear one major flow : they act as a fingerprint of both the steganography – which it is designed to, and the cover-source – which it suffers from. In the context of CSM, a good steganalysis feature would be resistant to the impact of the cover-source on images. On the other hand, in the context of steganography, a good forensic feature (i.e. to perform source identification) should be equally resistant to the impact of embedding, and should only capture the impact of the cover-source. This paper proposes such a forensic feature.

Extracting the noise in an image is an open problem. Under the assumption that the image processing pipeline is both linear and stationary, the noise can be modeled with an heteroscedastic Gaussian distribution. This distribution can then be estimated [6]. But this assumption is strict, especially when considering that, e.g., the gamma correction is a non- linear operation. It is therefore hard to leverage the properties of the Gaussian model of the noise in a practical scenario, where the steganalyst does not have access to side information.

3 CORRELATION FINGERPRINT OF THE IMAGE PROCESSING PIPELINE

3.1 Rationale

Pixel noise is inherent to any natural image. It comes from (1) the quantic measure of the number of photons hitting the photosensor when taking a picture and (2) the components of the camera itself (see [15] for references). Fig. 1 is a schematical illustration of the reasons inducing a multivariate distribution of the noise. The three correlation matrices show the relationship between pixels in an 8×4 neighborhood. A red cell indicates a positive correlation between two pixels, and a blue one indicates a negative correlation. This noise can be considered independent in the RAW domain, just after the acquisition [5]. However, processings such as demosaicking, denoising, and JPEG compression will introduce correlations between pixels, and their associated noise [24]. Naturally, different processings should introduce different correlations, as obtained after applying a sharpening and a denoising operation to the raw decorrelated noise. If one can capture the correlation of the noise in the image, he should be able to characterize the cover-source.

In an image, pixels are highly correlated together w.r.t. the content. To compute the correlation induced by the image processing pipeline (IPP), one should first remove the correlation caused by the content. In the present study, we try multiple methods to extract the noise in the image. But, to limit the impact of the content, we also provide a method to only keep the noise coming from the regions of the images bearing the least content.

There exist many solutions to estimate noise in an image, from trivial high-pass filtering to state-of-the-art noise extraction or denoising algorithms. In section 3.2, we describe the 6 methods that we considered in this experiment.

3.2 Noise estimation

In this section, we provide a concise introduction to the different noise estimation techniques, noted $\mathcal{F}$, that we explored. The goal here is to extract a multivariate noise coming from the acquisition of an image I, and fingerprinted by the IPP.

High-pass filtering. is the first type of filter we looked at. The straightforward convolutive Laplacian kernels are tested, both the 4-neighbor and 8-neighbor filters. The Sobel filter was also considered as an "off-the-shelf" tool; all 3 were implemented in the *opencv*

[1]Content icon comes from https://flaticon.com. Camera icon comes from https://iconarchive.com, and Acquisition parameter icon from https://icon-icons.com.

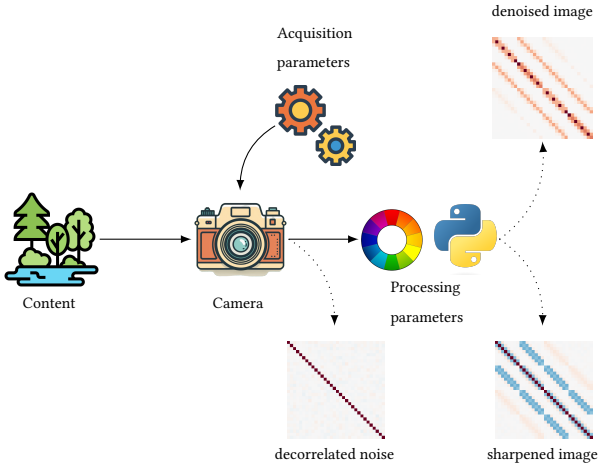


Figure 1: Schematic reasoning for the multivariate distribution of the noise in natural images.[1]

library.

$$\begin{cases} \mathcal{F}_1 = \mathrm{I} \circledast \mathcal{L}_4, \\ \mathcal{F}_2 = \mathrm{I} \circledast \mathcal{L}_8, \\ \mathcal{F}_3 = \mathrm{I} \circledast (\mathcal{S}_v^2 + \mathcal{S}_h^2). \end{cases} \tag{1}$$

where $\circledast$ is the convolution operator and

$$\begin{gathered} \mathcal{L}_4 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \mathcal{L}_8 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \\ \mathcal{S}_v^2 = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -4 & -2 \\ 1 & 2 & 1 \end{bmatrix} \text{ and } \mathcal{S}_h^2 = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & -2 & 1 \end{bmatrix}. \end{gathered} \tag{2}$$

Wavelet Filtering. is another famous family of algorithms in image processing. The Daubechies wavelet was considered here, based on the work of [5].

$$\mathcal{F}_4 = \mathrm{I} \circledast \mathcal{W} \circledast \mathcal{W}^T, \tag{3}$$

where $\mathcal{W}^T$ is the transpose matrix of $\mathcal{W}$ and

$$\mathcal{W} = \begin{bmatrix} .035 & .085 & -.135 & -.460 & .807 & -.333 \end{bmatrix}. \tag{4}$$

NoisePrint [4]. is a camera fingerprint extraction method. It is based on siamese convolutional neural networks, trained to extract similar noise estimations out of images from identical camera models. Its initial use was to detect image forgeries; two kinds of fingerprints would then be detected in a single image. We use it in our case to extract a single fingerprint that – hopefully, contains information on the IPP:

$$\mathcal{F}_5 = \mathrm{NPR}(\mathrm{I}). \tag{5}$$

DRU-Net [27]. is a deep learning denoiser. It is designed based on integrating the residual blocks of ResNet [8] in the architecture of the well-known U-net [22], and trained using images manually degraded with Gaussian noise. By computing the noise residual between the input and the output of the network, we can also hope to obtain a decent fingerprint of the processing pipeline:

$$\mathcal{F}_6 = \mathrm{I} - \mathrm{DRU}(\mathrm{I}). \tag{6}$$

3.3 Building the $C_{\mathcal{F}}$ feature

In an ideal case, one would want to compute the correlation of the noise between pixels in the whole image, noted I. For an image of size 512×512, however, that would result in a correlation matrix $\Sigma \in \mathcal{M}^{2^{18} \times 2^{18}}$, containing around 6.9^{10} coefficients. For this reason, and to take into account the JPEG compression pattern, we opt for correlations in an 8×8 neighborhood, that will be synchronized on the JPEG grid all throughout the paper. We therefore define neighbourhoods $b_i^I \in B^I$, where B^I is the set of 8×8 matrices scanned from the noise residuals.

Furthermore, as stated above, we want to limit the presence of the content in the estimation of the correlation of the noise. To that end, we sample blocks b_i^I for which the intra-block mean and variance are simultaneously low. This ensures that the selected blocks will be the smoothest of the image and centered around 0, i.e. they will bear the minimal amount of content that can persist in the filtered domain, such as edges and gradients, and, to a certain extent, textures.

We exploit the fact that, in the filtered domain, the mean and the variance of the samples are highly positively correlated. This ensures that a block with a low residual mean will very likely also have low variance. The sampling strategy is defined as such:

$$B_s^{\mathcal{F}_k} = \left\{ b_i^{\mathcal{F}_k} \;/\; b_i^{\mathcal{F}_k} \in \min_i^N \mathbb{V}[b_i^{\mathcal{F}_k}] \;\cup\; \min_i^N \mathbb{E}[b_i^{\mathcal{F}_k}] \right\}, \tag{7}$$

where $\mathbb{V}[.]$ is the variance, and $\mathbb{E}[.]$ the mean value, and $\min_i^N S$ is the set of the N lowest elements of set S. The correlation matrix is then computed between the samples in the set $B_s^{(\mathcal{F}_k)}$:

$$\Sigma^{(\mathcal{F}_k)} = (\sigma^{(\mathcal{F}_k)})_{u,v} = \frac{\mathrm{Cov}(b_u^{(\mathcal{F}_k)}, b_v^{(\mathcal{F}_k)})}{\sqrt{\mathbb{V}(b_u^{(\mathcal{F}_k)})\mathbb{V}(b_v^{(\mathcal{F}_k)})}}. \tag{8}$$

Once the correlation matrices are computed, the correlation features can be formed as an aggregation of the correlation coefficients of all of the correlation matrices. Since a correlation matrix is symetrical, and its diagonal is equal to 1, we get:

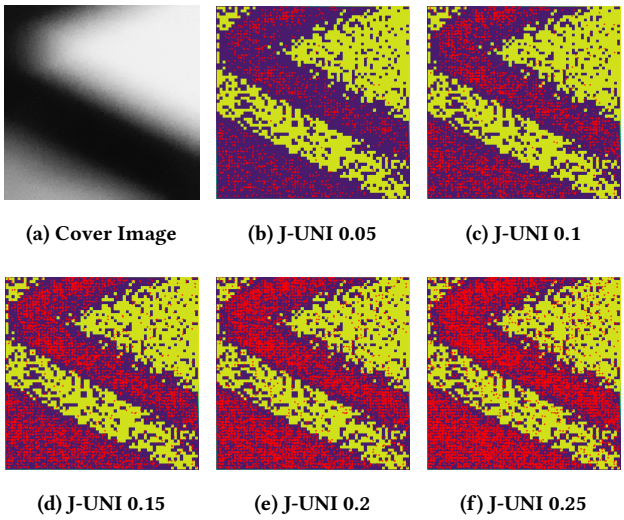
$$C_{\mathcal{F}} = \left\{ \mathrm{tri}(\Sigma^{\mathcal{F}_k}), \forall k \in 0 \dots 6 \right\}, \tag{9}$$

where $\mathrm{tri}(A)$ is the vectorization of the lower triangular coefficients of matrix A.

For 8×8 patches, we get 64×64 correlation matrices. Thus, there is $\frac{64 \times 64}{2} - 64 = 1984$ coefficients per noise estimation.

3.4 Resistance to embedding

As advocated in Sec. 2, one of the purposes of designing a new forensic feature is to ensure high resistance to steganographic embedding. The sampling strategy makes it naturally resistant to adaptative schemes, as highlighted in Fig. 2. For a given cover image (2a), the selected and rejected blocks, chosen according to Eq. 7, are shown in yellow and purple respectively. Then, the location of the embedding changes are shown in red. As expected, the selected (yellow) image blocks contain a vast minority of the embedding



(a) Cover Image (b) J-UNI 0.05 (c) J-UNI 0.1

(d) J-UNI 0.15 (e) J-UNI 0.2 (f) J-UNI 0.25

Figure 2: Sampling strategy against adaptative embedding strategy (J-uniward): (a) cover image, (b)-(f) embeddings at 0.05, 0.1, 0.15, 0.2, 0.25 bpnzAC respectively. Yellow and purple areas indicate the selected and rejected samples for the correlation estimation, respectively. The red dots indicate the DCT coefficients modified with the embedding.

changes, even as the embedding rate increases from 0.05 bpnzAC to 0.25 bpnzAC (2b-2f).

The reason is simple: adaptative schemes embed in the regions of the image where the changes are the least detectable, i.e. where the image is the most textured. Therefore, the sampling strategy naturally avoids the blocks where the changes are made. From Fig. 2b -2f though, one can see that as the payload increases, the number of changes included in the sampled blocks increases.

4 EXPERIMENTAL SETUP

In this section, we describe the experimental conditions to measure the performance of our new atomistic scheme compared to other atomistic schemes. We give particular emphasis on the synthetic and out-of-camera image datasets used. We also motivate our choices of ML models for both forensic and steganalysis, as well as the baseline used to show the relevance of our method.

4.1 Proposed atomistic scheme

Atomistic steganalysis is composed of two steps, as illustrated in Fig. 3. First, source identification is done using a multiclassifier based on the proposed $C_{\mathcal{F}}$ feature. Then, for each cover-source, a binary classifier is trained on steganalysis features extracted from images of the corresponding cover-source, to *detect* whether the samples are cover or stego objects. We used the DCTR and the Gabor filter residual (GFR) features as comparative baselines for the source identification step, as they were shown to perform well for this task [7].

For the forensic analysis, we chose the linear ensemble classifier proposed in [13], that has shown good performances with DCTR [9] in [7]. For the steganalysis, we chose a simple logistic regression

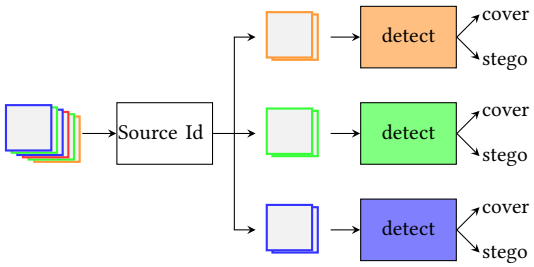


Figure 3: Illustration of the general workflow of the atomistic scheme. A multiclassifier performs source identification on images of unknown sources, which are then given to the steganalyzers trained on their predicted cover-sources.

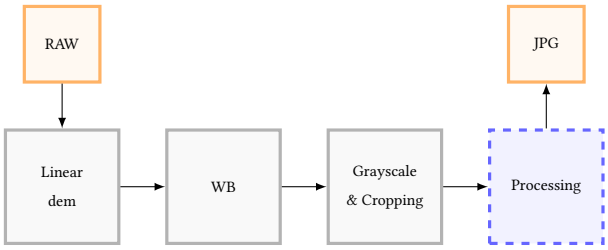


Figure 4: General flow of the processing pipelines used to generate our sources. RAW images come from the ALASKA Dataset. The processing part highlighted in blue changes between the sources, and can be seen in Table 1, while the other remain unchanged. The JPEG compression is done at QF 100.

(LR) model, as in [1], used with the DCTR features for all three source identifiers.

4.2 Datasets

4.2.1 Synthetic sources. Hand-crafted processing pipelines are first used to provide a controlled experimental setup to test our features on. The main advantage is that we are able to control the introduction of CSM through changes in the IPP. Images are developed using the open source software *Rawtherapee*. Two kinds of post-processings are used to generate synthetic sources: the directional pyramid denoising (DPD) [18] and the unsharp masking (USM) [21]. For both, 4 values of intensities are used. Eight additional sources are created by combining a strong DPD (resp. USM) followed by different intensities of USM (resp. DPD).

The whole processing pipeline is shown in Fig. 4. RAW images are first processed by linear demosaicking and a white balance. The specific post-processing is then applied. Images are finally converted to grayscale and center-cropped to size 512×512. Conversion to JPEG at quality factor 100 is then performed.

4.2.2 Out-of-camera sources. Additionally, we tested the source identification on several "out-of-camera" (OOC) JPEG images. The goal of this experiment is to measure the loss of accuracy when using our atomistic scheme, comprising 16 detectors trained on

synthetic sources, as compared to a clairvoyant detector, trained on the wilder OOC sources directly.

OOC sources all contain 1000 images. They are of two origins. Two are extracted from Flickr, and are formed by images taken by 2 users, equipped with a Sony SLT-A37 and a Canon Powershot SX30 IS, respectively. When looking for traces of post-processings in the metadata of the images, mentions of *adobe* were found for the Canon images. However, they consistently appear in all of the images. Therefore, we concede a slight abuse of language by calling this cover-source OOC. The other four are original ALASKA OOC images. They consist of an iPhone 11 Pro, a Xiaomi Mi 10T Pro, a Huawei P40 Pro, and a Nikon D810. These images will be used to measure the CSM problem between our atomistic scheme and dedicated OOC detectors. Since we only have 1,000 images per source, we apply a 5-split K-Fold procedure when training and testing the steganalysis detectors. The reported results of the steganalyzers in the clairvoyant scenario in Sec. 5.3 are the averaged results on the 5 splits. All the sources, both synthetic and OOC, are summarized in table 1.

Table 1: Processings and camera models used to define the synthetic and OOC cover-sources used in our experiments.

Synthetic sources				OOC sources	
USM	DPD	DPD-USM	USM-DPD	Flickr	ALASKA
50	30	90-50	350-30	Sony	iPhone 11
150	50	90-150	350-50		Xiaomi
250	70	90-250	350-70	Canon	Huawei
350	90	90-350	350-90		Nikon

5 EXPERIMENTAL RESULTS

We first investigate the results of the source identification step in section 5.1. We then report the performances of the atomistic scheme on the synthetic setup in section 5.2. The case of OOC sources is studied in section 5.3.

The adaptive scheme J-UNIWARD [10] was used to generate stegos, at a payload $\rho = 0.5$ bpnzac (bit per non-zero AC coefficients).

5.1 Source identification

To highlight the complementarity of the different filters $\mathcal{F}_k, k \in \{1 \dots 6\}$, we show the performance of the source identification model with the correlation coefficients of each filter separately in table 2. As we can see, noiseprint has the best informativeness on the source by far, further confirming the relevance of this model. Then, we see that the laplacians and the Sobel filters give better performances than the wavelet filter. Our explanation is that the impact of the latter on the pixels is more important than the formers, erasing the fingerprint of the processing pipeline. Finally, as shown in [17] with non-local means, out-of-the-shelf denoising-based residual estimation, such as DRU-net, can struggle, especially when dealing with already denoised images.

To further emphasize the resistance to embedding, we also report the classification results with only covers and balanced covers and stegos in the training and testing sets. Results are almost equal for all filters. We then compared the detection accuracy of the source identifier when using the full $C_{\mathcal{F}}$ feature, the DCTR, and the GFR features. Identically, we tested the performances for both the cover-only and the balanced cover and stego setups. Results are reported in table 3. Again, both setups bear similar results. Although our proposed scheme outperforms DCTR, it is still vastly inferior to GFR in this setup.

Table 2: Detection accuracy in % of all the features: the 6 filters. The accuracies between cover only and 50% cover and stegos are very close, highlighting the low amount of degradation due to embedding (at 0.5 bpnzac).

	$\mathcal{L}_4$	$\mathcal{L}_8$	$\mathcal{S}$	$\mathcal{W}$	Npr	Dru
Cover only	50.5	52.4	52.8	41.2	66.2	46.6
Balanced	50.4	52.3	52.6	40.9	66.3	46.5

Table 3: Source identification accuracy for the aggregated features $C_{\mathcal{F}}$, DCTR and GFR.

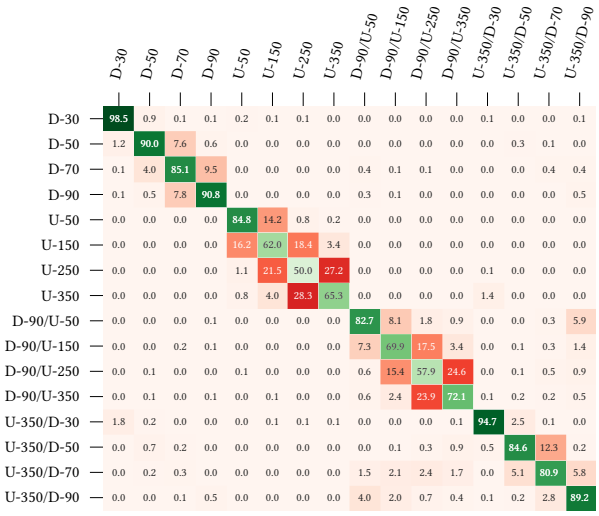
	$C_{\mathcal{F}}$	DCTR	GFR
Cover only	78.6	75.5	90.5
Balanced	78.7	75.2	90.3

On the other hand, in table 4, we look at the accuracy of the 3 detectors in the worst-case scenario, i.e. when training on only one class (cover or stego) and testing on the other. This time, we can see that the detector based on the $C_{\mathcal{F}}$ feature is not impacted, while the ones based on the steganalysis features become completely blind. While not very realistic in an experimental setup, these striking results should raise some awareness of this weakness of the steganalysis features.

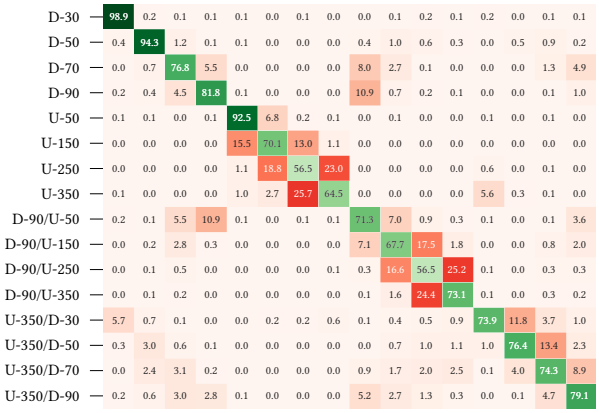
Another way to visualize the CSM (but not its impact on steganalysis) is to look at the confusion matrices of the source identification models, shown in Fig. 5. Fig. 5a highlights that that when using the $C_{\mathcal{F}}$ feature, errors mostly occur between cover-sources sharing the same set of processings. On the other hand, Fig. 5b DCTR tends to misclassify images from the "simple" cover-sources, defined by only one processing, as coming from "complex" ones, containing both sharpening and denoising. Fig. 5c further illustrates the superiority of the GFR features for the source identification task with the given set of cover-sources.

5.2 Atomistic steganalysis

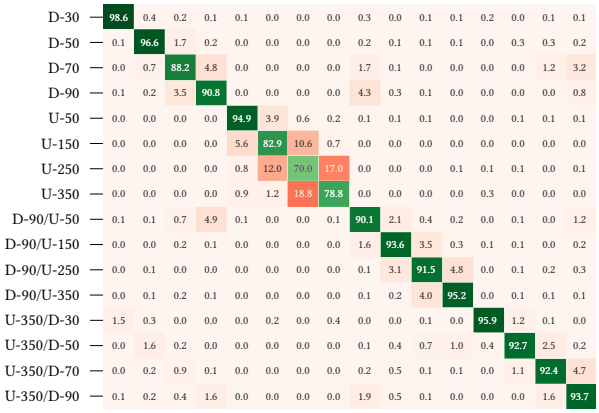
Atomistic steganalysis is then performed using the proposed source identification methods. For each cover-source, we train a specific steganalysis detector, using the images used in the training of the source identification step. Similarly, the test images are the tested images from the first step. The results of the steganalysis step, when using DCTR, are reported in table 5. We see that, despite the difference in accuracies in the source identification, the steganalysis schemes end up with similar results. Note that, despite the discrepancies reported for the source identification in table 3, we end up with average accuracies very close to the clairvoyant



(a) $C_{\mathcal{F}}$



(b) DCTR



(c) GFR

Figure 5: Confusion matrices obtained with the different source identification methods.

Table 4: Accuracy (in %) of the source identification step for the considered features in a worst-case scenario, where training on either covers or stegos, and testing on the other class.

Feature	$C_{\mathcal{F}}$		DCTR		GFR	
Train \ Test	Cover	Stego	Cover	Stego	Cover	Stego
Cover	78.6	79.0	75.5	6.2	90.5	6.3
Stego	78.4	79.1	5.8	75.3	6.1	90.5

scenario. This can be explained by the fact that with a relatively high payload of $\rho = 0.5$ bpnzAC, steganalysis models trained on similar sources will probably output identical predictions for an image coming from one of these sources. Note again that when using GFR instead of DCTR to perform steganalysis, the obtained accuracies are still very close to the clairvoyant scenario.

Table 5: Results of atomistic steganalysis on synthetic sources, using the DCTR features in the steganalysis step.

IPP \ Source Id		Clairvoyant	$C_{\mathcal{F}}$	DCTR	GFR
USM	50	*58.4*	59.4	**59.5**	58.1
	150	*54.2*	50.6	**55.7**	51.8
	250	*51.3*	50.1	**53.3**	51.0
	350	*53.2*	52.3	49.0	**53.6**
DPD	30	*67.3*	67.2	**69.1**	67.6
	50	*80.6*	79.6	**82.5**	80.2
	70	*89.9*	90.6	**91.3**	89.4
	90	*94.0*	94.7	94.5	**95.7**
DPD-USM	90-50	*90.3*	**91.4**	90.8	**91.4**
	90-150	*83.3*	82.7	**85.1**	83.6
	90-250	*77.0*	76.6	**78.1**	76.5
	90-350	*72.7*	69.1	70.9	**72.4**
USM-DPD	350-30	*51.3*	**52.4**	52.2	51.7
	350-50	*53.6*	55.7	52.8	**56.7**
	350-70	*63.8*	64.4	63.5	**65.0**
	350-90	*72.2*	70.6	72.2	**72.3**
	Average	*69.6*	69.2	**70.0**	69.8

5.3 Out-of-camera images

In this last experiment, we use the complete atomistic scheme trained on synthetic sources, and observe its overall detection accuracy compared to clairvoyant detectors trained and tested on each specific OOC cover-sources.

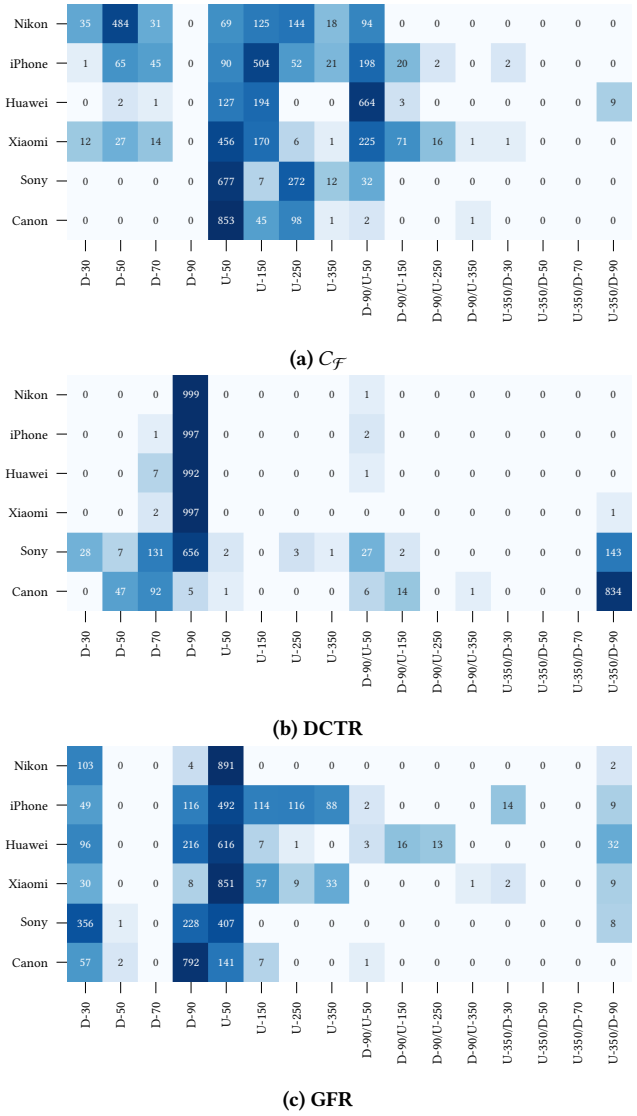
First, we can observe the differences in identifying the source of the OOC images, shown in Fig. 6. Whereas the DCTR detector mostly identifies the images as strongly denoised (see Fig. 6b) and the GFR one as being mostly slightly sharpened (see Fig. 6c), the $C_{\mathcal{F}}$ model is more nuanced (see Fig. 5a). This can be a consequence of the sampling strategy, which is adaptive to the amount of textures

Table 6: Detection accuracy (in %) of atomistic steganalysis on OOC sources.

	Clairvoyant	$C_{\mathcal{F}}$	DCTR	GFR
Sony	*90.8*	65.8	67.5	**81.7**
Canon	*58.8*	60.1	59.0	**62.0**
iPhone	*87.2*	**76.3**	49.2	75.1
Xiaomi	*64.8*	**78.4**	54.5	61.0
Huawei	*66.8*	68.4	51.9	**68.8**
Nikon	*72.8*	**75.9**	50.0	55.2
Average	*73.5*	**70.8**	55.3	67.3

in the image. Indeed, very textured images will necessarily produce "noisy" correlation coefficients, which can correspond to correlation matrices produced by a cover-source that has a stronger sharpening operation. A similar reasoning can be made for low-textured images.

In the second step of the atomistic scheme, we perform steganalysis similarly to Sec. 5.2. This time, for the three atomistic models, we observe detection accuracies substantially lower than the ones obtained with the clairvoyant detector (table 6). However, this time our scheme also clearly outperforms both GFR and DCTR features. This is probably a consequence of the more "accurate" predictions of the $C_{\mathcal{F}}$ model in the source identification step. One plausible explanation is that modern smartphones embed adaptive image correction algorithms. This can result in larger variety within a single cover-source. Consequently, the relatively low amount of samples during training might lack the ability to predict the unseen images of the same cover-source. On the other hand, our atomistic approach can grasp the actual distribution of the noise in every image, and correctly distribute them to the correct specific steganalyzer. In particular, this could explain the seemingly outstanding results obtained with the $C_{\mathcal{F}}$ model on the Xiaomi images. Experiments of a larger scale would be a good follow-up to these very promising results, which already validate the relevance of our work.



(a) $C_{\mathcal{F}}$

(b) DCTR

(c) GFR

Figure 6: Predicted IPP by the three source identification models for the considered OOC images.

6 CONCLUSION

In this work, we paid attention to define carefully the CSM as the cause of the drop in accuracy, and not being both the cause and the consequence. Having two problems to deal with at once, CSM first and steganography then, we proposed an atomistic scheme based on a newly proposed forensic feature designed to be a fingerprint of the IPP, the main cause of the creation of cover-sources.

We showed that the source identification using correlation coefficients of the noise in images is a promising alternative to the most common steganalysis features. Its interpretability is much greater, as it is based on a rigorous statistical model of the noise in developed natural images. It is also robust to the embedding, even at relatively high payloads. We also showed that our approach has better generalization capabilities to cover-sources unseen in the training of neither the source identifier nor the steganalysis models. Experimental results are very promising, and highlight the relevance of our approach.

With regard to the atomistic framework, our paper raised a number of questions. First, we opened the atomistic steganalysis to the problem of generalizing to unseen cover-sources. Second, the joint impact of the cover-source and the stego-scheme on the source identification can also be a point of concern in practical scenarios.

On another note, although the promising ideas behind the $C_{\mathcal{F}}$ feature have been validated in this paper, we can draw the following leads. First, the perspective of it becoming a rich model will naturally raise the question of the redundancy of the information contained in each filter. But, while it is a practical approach, its efficiency is to compare to the one of a single filter, such as noiseprint, which can already bear very good results on its own.

Designing our own noise extraction method, tailored for extracting a fingerprint of the cover-source, might also be a valid approach.

REFERENCES

[1] Rony Abecidan, Vincent Itier, Jérémie Boulanger, Patrick Bas, and Tomáš Pevný. 2023. Leveraging Data Geometry to Mitigate CSM in Steganalysis. In *2023 IEEE International Workshop on Information Forensics and Security (WIFS)*. 1–6. https://doi.org/10.1109/WIFS58808.2023.10374944

[2] Mauro Barni, Giacomo Cancelli, and Annalisa Esposito. 2010. Forensics-Aided Steganalysis of Heterogeneous Images. In *ICASSP*. IEEE, 1690–1693.

[3] Dirk Borghys, Patrick Bas, and Helena Bruyninckx. 2018. Facing the Cover-Source Mismatch on JPHide using Training-Set Design. In *IH&MMSec*. ACM, 17–22.

[4] Davide Cozzolino and Luisa Verdoliva. 2019. Noiseprint: A CNN-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security* 15 (2019), 144–159.

[5] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing* 17, 10 (2008), 1737–1754.

[6] Quentin Giboulot. 2022. *Statistical Steganography based on a Sensor Noise Model using the Processing Pipeline.* Ph. D. Dissertation. Université de Technologie Troyes.

[7] Quentin Giboulot, Rémi Cogranne, Dirk Borghys, and Patrick Bas. 2020. Effects and solutions of cover-source mismatch in image steganalysis. *Signal Processing: Image Communication* 86 (2020), 115888.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[9] Vojtěch Holub and Jessica Fridrich. 2014. Low-complexity features for JPEG steganalysis using undecimated DCT. *IEEE Transactions on Information forensics and security* 10, 2 (2014), 219–228.

[10] Vojtěch Holub, Jessica Fridrich, and Tomáš Denemark. 2014. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security* 2014 (2014), 1–13.

[11] Xiaodan Hou, Tao Zhang, Gang Xiong, Zhibo Lu, and Kai Xie. 2014. A Novel Steganalysis Framework of Heterogeneous Images Based on GMM Clustering. *SPIC* 29, 3 (2014), 385–399.

[12] Xiaodan Hou, Tao Zhang, Gang Xiong, and Baoji Wan. 2012. Forensics-aided Steganalysis of Heterogeneous Bitmap Images with Different Compression History. In *MINES*. IEEE, 874–877.

[13] Jan Kodovsky, Jessica Fridrich, and Vojtěch Holub. 2011. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on information forensics and security* 7, 2 (2011), 432–444.

[14] Jan Kodovský, Vahid Sedighi, and Jessica Fridrich. 2014. Study of Cover Source Mismatch in Steganalysis and Ways to Mitigate its Impact, In MWSF. *EI* 9028, 204–215.

[15] Takao Kuroda. 2017. *Essential principles of image sensors.* CRC press.

[16] Antoine Mallet, Martin Beneš, and Rémi Cogranne. 2024. Cover-source Mismatch in Steganalysis: Systematic Review. In *JIS*. eurasip.

[17] Antoine Mallet, Rémi Cogranne, Patrick Bas, and Quentin Giboulot. 2023. Identification de Développements d'Images par Matrices de Corrélations. In *XXIXème Colloque Francophone de Traitement du Signal et des Images (GRETSI'23)*. Université de Grenoble and Association Gretsi, Grenoble, France.

[18] Truong T Nguyen and Soontorn Oraintara. 2008. The shiftable complex directional pyramid—Part II: Implementation and applications. *IEEE Transactions on Signal Processing* 56, 10 (2008), 4661–4672.

[19] Jérôme Pasquet, Sandra Bringay, and Marc Chaumont. 2014. Steganalysis with Cover-Source Mismatch and a Small Learning Database. In *EUSIPCO*. IEEE, EURASIP, 2425–2429.

[20] Tomáš Pevný and Jessica Fridrich. 2008. Multiclass Detector of Current Steganographic Methods for JPEG Format. *TIFS* 3, 4 (2008), 635–650.

[21] Andrea Polesel, Giovanni Ramponi, and V John Mathews. 2000. Image enhancement via adaptive unsharp masking. *IEEE transactions on image processing* 9, 3 (2000), 505–510.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 234–241.

[23] Claude E Shannon. 1949. Communication theory of secrecy systems. *The Bell system technical journal* 28, 4 (1949), 656–715.

[24] Théo Taburet, Patrick Bas, Wadih Sawaya, and Jessica Fridrich. 2020. Natural steganography in JPEG domain with a linear development pipeline. *IEEE Transactions on Information Forensics and Security* 16 (2020), 173–186.

[25] Liran Yang, Min Men, Yiming Xue, Juan Wen, and Ping Zhong. 2021. Transfer subspace learning based on structure preservation for JPEG image mismatched steganalysis. *Signal Processing: Image Communication* 90 (2021), 116052.

[26] Likai Zeng, Xiangwei Kong, Ming Li, and Yanqing Guo. 2015. JPEG Quantization Table Mismatched Steganalysis via Robust Discriminative Feature Transformation. In *MWSF*, Vol. 9409. SPIE, 270–278.

[27] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. 2021. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 10 (2021), 6360–6376.

[28] Dominik Šepák, Lukáš Adam, and Tomáš Pevný. 2022. Formalizing Cover-Source Mismatch as a Robust Optimization. In *EUSIPCO*. IEEE, EURASIP.